FUNDAMENTALS OF ENGINEERING ELECTRONICS

BY

WILLIAM G. DOW

Assistant Professor of Electrical Engineering
University of Michigan

Printed in U. S. A.

PREFACE

It has for some time been a conviction of the author and his associates that, no matter how facile an engineer may be in the manipulation of electronic circuits, his effectiveness is distinctly limited unless he has a satisfactory understanding of the operating principles of the electronic circuit *elements* that he uses. The ability to grasp readily the reasons for the behavior of new devices as they appear is especially important. The relative ease with which electronic devices can be made to order to accomplish specific purposes is an added reason for placing initial emphasis in electronics instruction on internal operating principles. These considerations have been the basis for the method of instruction in electronics at the University of Michigan ever since courses in the subject were introduced about eight years ago.

In the preparation of this book, which is an outgrowth of the author's teaching experience, an attempt has been made to maintain a proper balance between two underlying objectives:

(1) To give the reader a realistic and quantitatively usable conception of the principles that govern the internal behavior of electronic devices (this is the primary objective); and

(2) To familiarize the reader with methods of circuit analysis customarily employed in connection with the most common engineering applications of electronic devices.

Accordingly, the chief emphasis in this book is placed on internal operating principles. A large part of the text is devoted to a study of the effects of the use of various geometries and materials in electronic devices, and a relatively small part to circuit studies.

The point of view is that of an engineer: Principles of importance in engineering work are selected for study; illustrations of these principles are drawn from engineering practice; and physical concepts are so treated as to permit ready determinations of *magnitudes*. Familiarity with relative magnitudes is of course essential to a satisfactory engineering understanding of any scientific subject-matter.

Reasoning from purely physical concepts has been used rather than mathematical formulation, wherever the latter could be avoided without loss of definiteness. However, a large part of the subject-matter requires mathematical analysis for the establishment of proper quantitative concepts, and wherever that is true, mathematical methods have been used freely.

It is perhaps unfortunate that one of the most difficult subjects treated falls naturally very early in the text, in Chapter II. However, in his work with undergraduates the author has tried teaching triode electrostatic field analysis later in the course, and has also tried a qualitative treatment, involving a statement of the general results with little attention to the mathematical formulation. Such methods have not proved satisfactory. The conclusions reached by means of field analyses are rather striking, and class enthusiasm is dampened if denied satisfaction of the curiosity that naturally arises as to how these conclusions are arrived at.

This book has been especially designed for use in full-year courses for undergraduate or graduate students; however, the content is so arranged that it is readily adaptable, with certain omissions, for one-semester courses. It is hoped that the book will also find a place as a reference work for engineers in industry.

The author wishes to acknowledge gratefully the encouragement and active assistance rendered during the entire period of development of the electronics work by Professors A. D. Moore, S. S. Attwood, and L. N. Holland, all of the Electrical Engineering Department of the University of Michigan; also to express appreciation of the care and thoughtfulness with which portions of the manuscript were reviewed and ably criticized by Professor Attwood, and by Professor Samuel Goudsmit of the Physics Department of the University of Michigan. The author is indebted to all other members of the staff of the Electrical Engineering Department, and to many of those in the Physics]Department, especially to Professor O. S. Duffendack, for ever-ready advice and counsel; also to Messrs. Ralph Bodine and John Lopus for the thoughtful and careful draftmanship exercised during the preparation of the original drawings for the figures.

W. G. Dow

March 20, 1937

CONTENTS

PART I

ELECTRONS

INTRODUCTION

HIGH-VACUUM THERMIONIC TRIODES

CHAPTER I

POTENTIAL DISTRIBUTION DIAGRAMS

CHAPTER II

THE ELECTROSTATIC FIELD OF A TRIODE

v

CHAPTER III

ELECTRON BALLISTICS

CHAPTER IV

CATHODE RAYS

CHAPTER V

SPACE-CHARGE FLOW

CHAPTER VI

CHAPTER VII

THERMIONIC CATHODES

CHAPTER VIII

WORK FUNCTIONS OF HOMOGENEOUS SURFACES

CHAPTER IX

CHAPTER X

DISTRIBUTIONS OF RANDOM VELOCITIES OF GAS PARTICLES

CHAPTER XI

ELECTRICAL EFFECTS OF RANDOM MOTIONS

CHAPTER XII

AMPLIFIER CIRCUIT PRINCIPLES

CHAPTER XIII

HARMONICS; CLASS B AND PUSH-PULL AMPLIFIERS

CHAPTER XIV

PART II

ELECTRONS, ATOMS, AND RADIATION

CHAPTER XV

ATOMIC ENERGIES

CHAPTER XVI

ENERGY LEVELS FOR PARTICULAR ELEMENTS

CHAPTER XIX

CHAPTER XX

CHAPTER XXI

LIST OF TABLES

FUNDAMENTALS OF
ENGINEERING ELECTRONICS

PART I

ELECTRONS

INTRODUCTION

HIGH-VACUUM THERMIONIC TRIODES

1. A Triode in a Simple Amplifier Circuit. The first part of this book is devoted to a study of the properties and uses of high-vacuum thermionic grid-controlled electron tubes. The simplest of these, called *triodes*, each contain three electrodes: grid, anode or *plate*, and cathode or *filament*.

Fig. 1 is a diagram of a simple circuit which employs a triode like that of Fig. 2 for amplifying the variations of more or less irregularly alter-
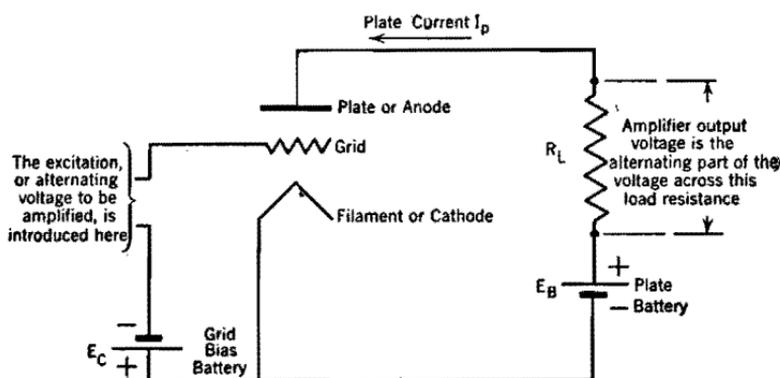
Fig. 1. Triode connections in an amplifier circuit.

nating voltages. A voltage whose alternating variations are to be amplified, called the input or *excitation* voltage, is introduced in series with the grid bias battery. The potential difference between grid and cathode therefore consists of a steady (direct-current) voltage plus an alternating one. The plate circuit carries (1) a steady current, corre-

1

sponding to zero excitation voltage, plus (2) an alternating current that follows the pattern of the excitation voltage, because of the electro-static control over electron passage exerted by the grid. The voltage drop in the load resistance R_L has cor-responding direct-current and alter-nating-current parts. The latter is the useful *output voltage* of the ampli-fier; it has the same pattern as the excitation voltage, but on an enlarged scale.
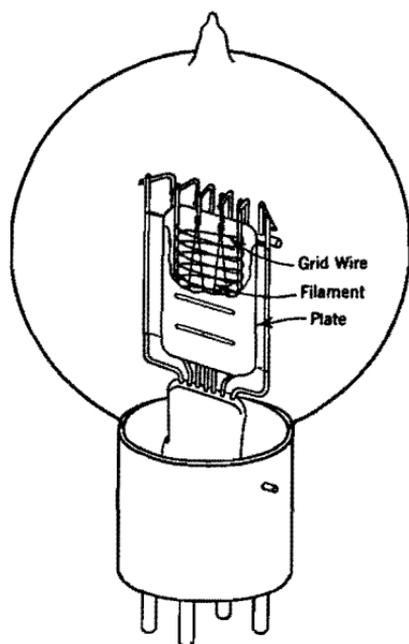
The *voltage gain* of the amplifier is the ratio of the output voltage, that is, the alternating part of the voltage across the load resistance, to the excitation voltage. Numerical prob-lems relative to circuits like that of Fig. 1 are solved by the use of *tube characteristic curves*, of the general type illustrated in Fig. 3, and with the aid of imaginary alternating-current equivalent circuits that are representative of tube properties.

Fig. 2. Western Electric type 101-F triode.

(a) Mutual characteristics   (b) Plate characteristics
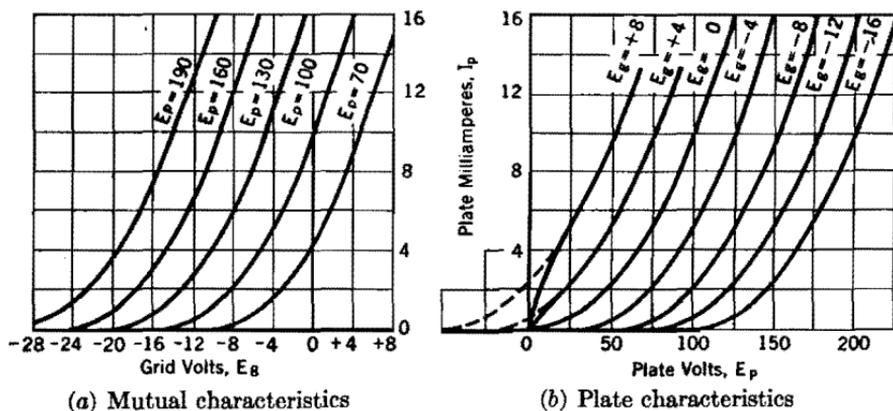
Fig. 3. Average characteristic curves for a triode (Western Electric type 101-F).

2. Electrostatic Control of Plate Current by Grid and Plate Voltages.

The region inside the glass envelope of a tube of the type illustrated in Fig. 2 is made as nearly perfect a vacuum as is commercially feasible. The cathode (filament) surface is made of some substance that is a thermionic emitter, that is, releases electrons when at a red or white

heat.   An auxiliary filament circuit, not shown in Fig. 1, is used to main-
tain the required filament temperature.   The number of electrons
released per second is not important as long as there are *enough* of them,
because the actual current flow is electrostatically controlled by the
values of plate and grid potentials, not in general by filament tem-
perature.   The filament must be hot enough to provide the maximum
number of electrons likely to be demanded per second by the elec-
trostatic control under normal conditions, yet not so hot as to destroy
the filament or the electron-emitting properties of the filament surface.
In general the normal filament temperature sets an upper limit to the
tube current.

Attention will first be directed to conditions within a triode that has
parallel plane cathode and plate surfaces, as illustrated in the top por-
tions of Figs. 4a and 4b.   The solid lines in the voltage diagrams in Fig. 4
describe typical point-to-point variations in potential (the *potential
distributions*) within such a triode when the cathode is cold, so that no
electrons are emitted and the tube interior is entirely free from space
charge.   Of course no current flows under such conditions.   The dotted
lines in Figs. 4a and 4b similarly describe potential distributions after the
cathode is heated, so that electrons are emitted, and negative space
charge does therefore exist in the interior.   Under such conditions cur-
rent does flow, because of the movement of the electrons that constitute
the space charge.

Thus the *solid* lines in the four diagrams of Fig. 4 all represent *space-
charge-free* potential distributions that correspond to four different
values of grid voltage.   The upper solid line in each diagram describes
the space-charge-free potential variation along a path midway between
adjacent grid wires; the lower solid line, that along a path through the
center of a grid wire.

The general " topography " of space-charge-free potential variation
in the tube interior can be simulated by a sheet of rubber in tension.
The rubber must initially be tightly stretched between the edges of two
supports in such a way as to give its surface an inclination, the support
at one end being higher than that at the other, just as the plate potential
is higher than that of the cathode.   Then a miniature inverted picket
fence is lowered, in a vertical plane intermediate between and parallel to
the supports, until the ends of the pickets depress regularly spaced
points of the sheet to a level slightly below that of the lower support
(" cathode ").   The level reached by the ends of the pickets corre-
sponds to grid potential.

The upper solid line of any one of the diagrams of Fig. 4 represents
the profile of such a stretched sheet, in a plane placed perpendicular to

the supports and midway between adjacent pickets. Each lower line represents similarly a profile for a plane similarly located except that
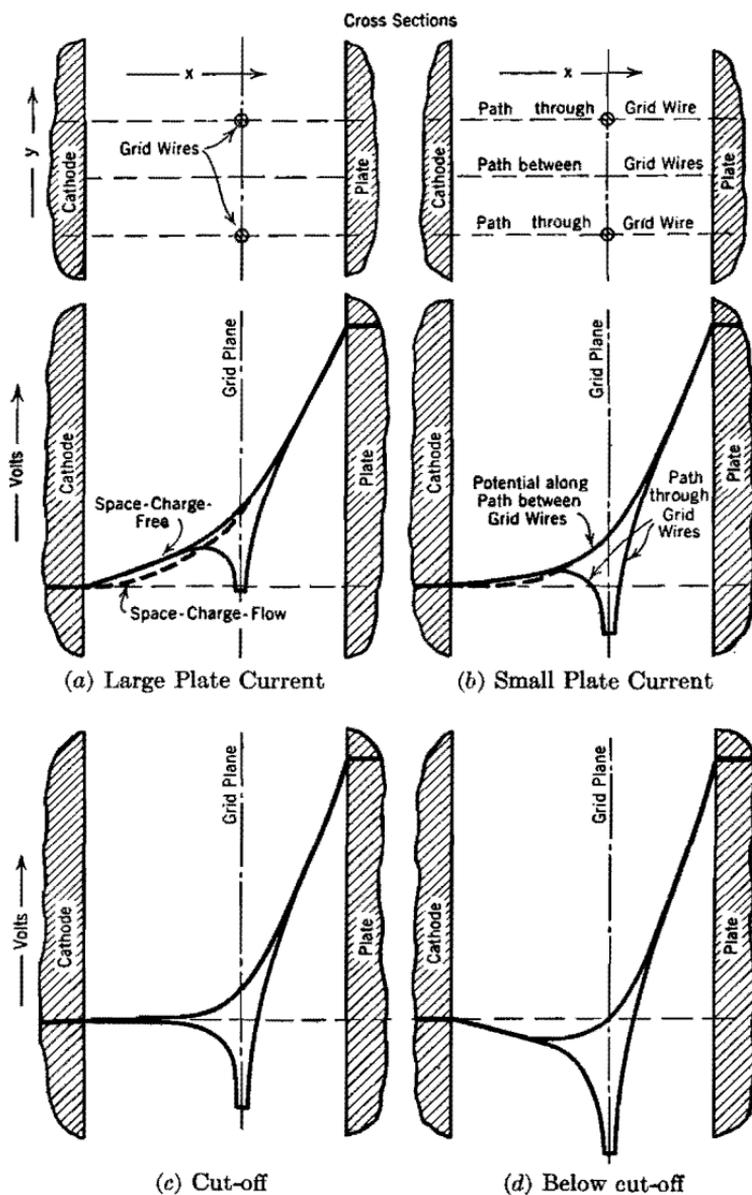
FIG. 4.   Potential distribution diagrams for a parallel-plane triode at various grid voltages.

it passes through a picket. In order to make the analogy satisfactory the rubber must be stretched tightly enough so that after lowering

the pickets the portions of the sheet near the supports are flat inclined surfaces free from wrinkles. This flatness corresponds to the fact that the space-charge-free electric field in a triode is practically uniform near the cathode and plate surfaces, although of greater strength near the plate than near the cathode.

A slant in the potential line adjacent to the cathode indicates the existence there of a potential gradient. In the study of electric fields potential gradient is found to be identical with the negative of electric field intensity, and with the negative of the flux density if the dielectric constant is unity, as it is within a vacuum tube. The slope of the potential line just off the cathode must therefore be a direct measure of the number of flux lines that terminate on each square centimeter of cathode area, and so of the electric charge density on the cathode.

The slope of the potential line near the cathode is dependent on grid potential, in the manner indicated by the contrasts between the various diagrams in Fig. 4. Therefore *the magnitude of the negative charge on the cathode changes whenever the grid potential is changed*, that is, whenever the picket fence is raised or lowered.

As soon as the cathode is heated the escape of electrons moves the negative surface charge into the region near to but outside the cathode, where it becomes space charge. The surface charge before electron release is represented in Figs. 4a and 4b by the sharp turns of the solid potential lines at cathode surfaces. The space charge after release similarly corresponds to the gradual bends in the dotted potential lines. The total amount of turn, and so the total charge, is about the same whether on the surface (solid lines) or just outside of it (dotted lines). Therefore *the total amount of space charge after electron release is proportional to the steepness of the space-charge-free off-cathode gradient, and is for that reason dependent on grid potential.*

The flow of space charge constitutes the plate current; the less the space charge, the smaller the current. Therefore the lowering of the grid potential from its value in Fig. 4a to a more negative one in Fig. 4b results in a decrease in plate current. Thus, by controlling the electrostatic field in the tube, the grid voltage controls the magnitude of the plate current. However, the dependence of current on space-charge-free gradient, and so on grid potential, is more than linear, because a decrease in this gradient (1) calls for a smaller space charge, also (2) makes the space charge flow more slowly.

If the grid potential is low enough to make the off-cathode gradient zero, as in Fig. 4c, or negative, as in Fig. 4d, no plate current flows after electron release. The zero off-cathode gradient condition, Fig. 4c, corresponds to " cut-off," that is, to the termination at the zero-plate-

current axis of a *mutual characteristic curve* of the type illustrated in Fig. 3a.

3. Organization of Text Material. The first two chapters of the book prepare the reader for and present an analysis of the electrostatic field within a triode before any electrons are released. The primary purpose of this analysis is to establish the nature of the dependence of the space-charge-free off-cathode gradient on plate and grid voltages, especially the latter. The dependence of tube current after electron release on the off-cathode gradient before release is established in later chapters. The results of the two studies are subsequently combined to show why the current-voltage relationship must be of the general type illustrated in Fig. 3.

Studies of the principles that govern electron movements in the presence of electric and magnetic fields, of space-charge flow in four- and five-electrode tubes, and of the principles underlying thermionic emission, are introduced at appropriate points. The last three chapters of Part I are devoted to a study of high-vacuum thermionic tube applications, chiefly in amplifier circuits.

Part II is devoted to a study of the types of electronic behavior that involve interactions between electrons, atoms, and electromagnetic radiation. This requires an introductory study of the mechanisms of ionization and light radiation; photoelectric and gaseous-discharge theory and applications are treated in the later chapters.

CHAPTER I

POTENTIAL DISTRIBUTION DIAGRAMS

4. Units and Conversions. The fundamental relationships used in the analysis of electric fields within electronic devices are most easily developed and expressed in the electrostatic system of units[4 I†] (statvolts, statamperes, statcoulombs), and those used in the analysis of magnetic fields in magnetic units[4 VIII] (abvolts, abamperes, abcoulombs), yet it is important to be able to express results in practical units (volts, amperes, coulombs). The important conversion relations[4 XIV] are:

$$1 \text{ volt} \quad = 10^8 \text{ abvolts} \quad = \tfrac{1}{300} \quad \text{statvolt} \tag{1}$$

$$1 \text{ ampere} = 0.1 \text{ abampere} = 3 \times 10^9 \text{ statamperes} \tag{2}$$

$$1 \text{ coulomb} = 0.1 \text{ abcoulomb} = 3 \times 10^9 \text{ statcoulombs} \tag{3}$$

For the equations in the text it is designated at the extreme right of the page whether the electrical units are in the electrostatic system (esu), the magnetic system (emu) or the practical system (p). Equations not so designated are valid for all three systems.

A convenient and easily checked method of making conversions between systems of units can be illustrated by application to a space-charge-limited-current equation that is derived in a later section. The conversion steps are as follows:

(1) Write the equation in its known form:

$$J = \frac{\sqrt{2}}{9\pi} \sqrt{\frac{e}{m}} \frac{E^{\frac{3}{2}}}{s^2} \tag{4 esu}$$

Here J = current density, e = electronic charge, E = potential.

(2) Rewrite it, substituting empty brackets for the symbols which are to be converted, labeling the units of each bracket by subscripts:

$$\left[\quad \right]_{sa} = \frac{\sqrt{2}\left[\quad \right]_{sc}^{\frac{1}{2}}\left[\quad \right]_{sv}^{\frac{3}{2}}}{9\pi \sqrt{ms^2}} \tag{5 esu}$$

(3) Insert in each empty bracket the symbol for the type of unit into which conversion is to be made, together with the proper multiplier. It is important to recognize that a symbol is a substitute for the *number* that measures, in proper units, a *definite physical condition*. When converting statvolts into volts, one should place inside the bracket the symbol for volts multiplied by $\tfrac{1}{300}$, for that is the number by which one must multiply the number 600, which measures for example

† For letter and number references, see Bibliography.

7

the peak of an a-c 424-volt wave in volts, to secure 2, its measure in statvolts. In the present example this is done as follows:

$$\left[J_a \times 3 \times 10^9 \right]_{sa} = \frac{\sqrt{2}\left[e_c \times 3 \times 10^9 \right]_{sc}^{\frac{1}{2}} \left[E_v \times \frac{1}{300} \right]_{sv}^{\frac{3}{2}}}{9\pi \sqrt{ms^2}} \tag{6 p}$$

(4) Simplify the expression after removing the brackets:

$$J_a = \frac{\sqrt{2}}{9\pi} \sqrt{\frac{e_c}{m}} \frac{E^{\frac{3}{2}}}{s^2} \left[\frac{1}{(300)^{\frac{3}{2}} (3 \times 10^9)^{\frac{1}{2}}} \right]$$

or

$$J_a = \frac{10^{-7}}{81\pi \sqrt{5}} \sqrt{\frac{e_c}{m}} \frac{E^{\frac{3}{2}}}{s^2} \tag{7 p}$$

5. Electric Intensity and Potential; Force on an Electron. The field relationship between electric intensity and potential[4] underlies the entire study of electronic devices. The simplest mathematical expressions of its two important aspects are

$$F = -\frac{dE}{ds} \tag{8 esu}$$

and

$$E_2 - E_1 = -\int_1^2 F \, ds \tag{9 esu}$$

The first form describes mathematically the fact that electric intensity F is the negative space derivative of electric potential E. The second form indicates that, conversely, the difference of potential between two points in the field is the negative line integral of the intensity along a path (any path) between the two points.

In Equations (8) and (9), F is the electric field intensity in dynes per statcoulomb, exactly the same as the statvolts per centimeter; E is the electric potential in statvolts, and ds is incremental distance in centimeters, measured parallel to the direction of the electric force. If the line integral is taken over a path that is not everywhere parallel to the direction of the electric flux, ds represents only the component, in the direction of the field, of each increment of the path.

The electric intensity is a vector; in rectangular coordinates its components are

$$F_x = -\frac{\partial E}{\partial x} \qquad F_y = -\frac{\partial E}{\partial y} \qquad F_z = -\frac{\partial E}{\partial z} \qquad (10 \text{ esu})$$

The application of Equations (8) and (9) to the field intermediate between two parallel metal plates (no grid), as for example those of an air condenser, is illustrated graphically in Fig. 5a. Potential is measured vertically, distance horizontally. Therefore the *potential gradient,* which is the slope of the potential line, is numerically equal to the

electric intensity, though of opposite sign.   Since the force on an electron
or ion of charge e in the intermediate region is eF dynes, the slope of
the potential line is proportional to the force on, and the acceleration
experienced by, such a particle.

Diagrams similar to Fig. 5a are peculiarly helpful because the force
relationships suggested by them correspond roughly to everyday
experience.   The potential line of Fig. 5a may be likened to the side of
a hill (Fig. 5b), the top level of the hill corresponding to the potential
of the more positive plate, the bottom level to that of the more negative
plate.   It is apparent that the force tending to roll a ball down the hill
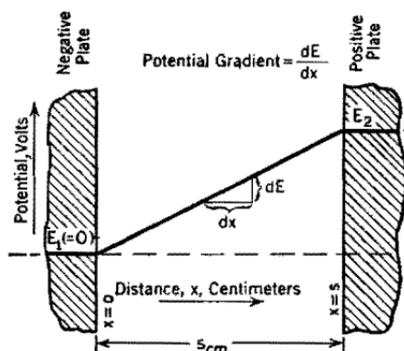
FIG. 5a.   Potential distribution be-   FIG. 5b.   Force tending to roll a ball
tween two parallel metal surfaces.   down a hill is dependent on the gradient.
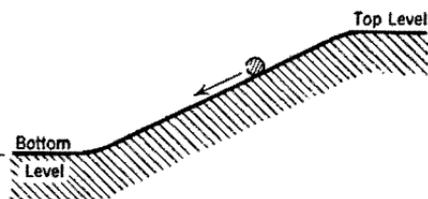
is dependent on its slope, which is merely another name for the gradient
or space derivative of the elevation.   Similarly the force on an electron
is dependent on the slope of the potential line.

It is common practice to make positive potentials correspond to
upper levels in potential diagrams, and negative potentials to lower
levels.   Unfortunately, this practice makes electrons, the more mobile
of the charged particles dealt with, have the property of rolling uphill
rather than down.

The force on an electron is proportional to the slope of the potential
line, that is, to the tangent of the angle it makes with the horizontal,
while in the gravitational analogy the force on the ball is proportional
to the sine of the angle between hillside and horizontal.   For this reason
the analogy is satisfactory only as long as comparison is confined to
gently sloping hillsides for which the sine and the tangent are nearly
alike.

**6. Poisson's and the Laplace Equations;   Potential Distribution
Diagrams in One and Two Dimensions.**   The utility of diagrams which,
like Fig. 5a, describe *potential distribution,* is not limited to simple

geometries. No matter what the nature of the potential variation, such diagrams picture usefully the proportionality between potential gradient and the force tending to accelerate an electron or ion. If three space coordinates, x, y, and z, are necessary to describe the configuration of an electric field, the force tending to accelerate an electron in any one of the three reference directions is proportional to the negative space derivative of the potential in that direction. For that reason potential sections or profiles, similar to the elevation profiles that surveyors use in mapping highway or railroad rights-of-way, are very useful.

In many electric circuit elements, for example, in inductances, capacitances, electric machines, wires, incandescent lamps, and the like, it is generally satisfactory to assume that electric charge exists only on the surfaces of the conductors. Surface charge can be satisfactorily measured in surface density units, that is, in coulombs or statcoulombs per square centimeter of conductor surface. In electronic apparatus electric charge is frequently distributed throughout a volume, and must be measured in units of volume density or space-charge density, by stating the number of coulombs or statcoulombs per cubic centimeter.

If the volume charge results from the presence of electrons, the space-charge density is negative; if from the presence of positively charged ions, the space-charge density is positive. If both kinds of charged particles are present, the space-charge density corresponds to the algebraic sum of their opposite effects. It is often negligibly small in regions having very high concentrations of both kinds of particles.

The general partial differential equation relating potential E, rectangular space coordinates x, y, and z, and space-charge density ρ, the last in statcoulombs per cubic centimeter, is [B 40]

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = -4\pi\rho \qquad (11 \text{ esu})$$

This expression is universally known in electrical work as *Poisson's equation*. Many engineering problems deal primarily with fields in which there is no space charge; in such cases the following more simple expression applies:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = 0 \qquad (12 \text{ esu})$$

Equation (12) is known as *the Laplace equation.*[B 40]   It applies in any interelectrode region from which space charge is absent, and in which the dielectric constant is everywhere uniform. The general problem of

electrostatic field analysis is that of finding solutions of Equation (12) which satisfy boundary conditions imposed by electrode geometry and potentials.   For most simple and many complicated geometries special methods are used which accomplish this purpose without mention of the underlying equation,[4] but in all cases the correct solutions will be found to satisfy it.

The electric field in the region between the parallel plates of Fig. 5a is said to be *one-dimensional*.   The plates are considered to be of infinite extent.   It is obvious that the potential has a constant value at all points in any intermediate plane parallel to the two plates, that is, the equipotentials are all perfectly flat surfaces perpendicular to the x coordinate direction.   Evidently explorations confined to travel in the y and z directions will discover no change in potential, so that the last two terms of the left-hand sides of Equations (11) and (12) drop out; Poisson's and the Laplace equations in one dimension are simply

$$\frac{d^2E}{dx^2} = -4\pi\rho \qquad \text{(13 esu)}$$

and

$$\frac{d^2E}{dx^2} = 0 \qquad \text{(14 esu)}$$

These expressions represent mathematically the fact that potential varies only with changes in the one coordinate dimension x, hence the use of the term " one-dimensional."

It has been pointed out in connection with Fig. 5a that the slope of the space-potential line is the potential gradient, dE/dx.   The expression d^2E/dx^2 is merely a simplified way of writing

$$\frac{d}{dx}\left(\frac{dE}{dx}\right)$$

which is of course the space rate of change of the slope, and is called the *flexion* of the potential curve.   Equation (14) therefore requires that the potential distribution curve for a one-dimensional field in a region free from space charge must be a straight line, for only a line that is straight has zero flexion.   If in any region the space-potential line is in fact curved, either space charge is present, or the field is not one-dimensional, or both.

The plate and cathode surfaces of Fig. 4 are considered infinite in extent and the grid wires infinitely long.   Hence an exploration confined to travel along a line parallel to a grid wire will discover no variation in potential.   Only the x and y gradients $\partial E/\partial x$ and $\partial E/\partial y$ have values, so the fields in Fig. 4 are said to be *two-dimensional*.   Poisson's

and the Laplace equations in rectangular coordinates for a two-dimensional field are:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} = -4\pi\rho \qquad \text{(15 esu)}$$

and

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} = 0 \qquad \text{(16 esu)}$$

It is fortunately true that a great many electric fields whose configuration is important in engineering work are two-dimensional, for a variety of mathematical methods of no great difficulty are available for the determination of the properties of fields that satisfy Equation (16). The electric fields around transmission lines, within the sheaths of power and communication cables, around conductors in the armature slots of electrical machines, and in the interiors of many electronic devices are essentially of this nature. The equipotentials in such regions have flexion in one plane only. They may be called cylinders or cylindrical surfaces, although only in certain special cases are they circular cylinders.

The second x-derivative of the potential, $\partial^2 E/\partial x^2$, may in general have a value, so the potential profiles in the x direction have flexion, either (A) in a one-dimensional field due to the presence of space charge, as along the dotted lines near the cathodes of Figs. 4a and 4b, or (B) in a two- or three-dimensional field with or without space charge, as along the curved parts of the solid lines in the same figures. In cases of the (A) type

$$(A) \qquad\qquad \frac{d^2 E}{dx^2} = -4\pi\rho \qquad \text{(17 esu)}$$

In cases of the (B) type, without space charge

$$(B) \qquad\qquad \frac{\partial^2 E}{\partial x^2} = -\frac{\partial^2 E}{\partial y^2} \quad \text{(two-dimensional)} \qquad \text{(18 esu)}$$

or

$$(B) \qquad\qquad \frac{\partial^2 E}{\partial x^2} = -\left(\frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2}\right) \quad \text{(three-dimensional)} \qquad \text{(19 esu)}$$

Near plate and near cathode of a parallel-plane triode (Fig. 4) the electric field is essentially uniform. The statement that the field is uniform is merely another way of saying that it is one-dimensional. Correspondingly, *before* the entrance of electronic space charge the space-potential line near the cathode is *straight*, as described by the *solid* lines. The *dotted* line in Fig. 4a is *curved* near the cathode because

it describes the potential distribution *after* electrons are permitted to enter and establish a strong negative space charge just outside the cathode.

In the neighborhood of the grid, Fig. 4, all the solid (space-charge-free condition) lines are curved. Their curvature indicates that in that region the field is not one-dimensional. And of course it cannot be, for the potential varies along a path from one grid wire over an intermediate hill to the next wire. The potential of each grid wire is negative, that of points midway between grid wires positive, relative to the cathode.

Equation (18) is interesting in that it indicates that if, in a space-charge-free two-dimensional field, the flexion of the x profile is positive (convex downward) that of the y profile must be negative (convex upward). This is suggestive of the "topography" of a saddle. Models of potential topography corresponding to the various diagrams in Fig. 4 would contain many saddle-like portions.

7. Surface and Space–Charge Density in Potential Diagrams. Fig. 5a can be used to explain how the presence of surface-charge density is represented in potential profile diagrams. The electric field of Fig. 5a terminates in surface charges at the two plates; at each surface the abrupt change in the potential gradient, that is, in the slope of the potential line, is a direct measure of the surface-charge density.

The lines of electric flux between the plates terminate at the metal surfaces. But flux lines terminate only on electric charges; mathematically, 4π flux lines end on each statcoulomb of charge. Hence the charge density on the terminal surface must be proportional to the flux density, so to the electric field intensity and the potential gradient that exist just outside the surface. Mathematically, at the negative plate,

$$F = +4\pi\sigma, \quad \text{or,} \quad \frac{dE}{dx} = -4\pi\sigma \qquad \text{(20 esu)}$$

Here σ stands for surface-charge density in statcoulombs per square centimeter, and the dielectric constant is unity.

A little confusion in sign results from an attempt to apply Equation (20) at the positive plate. It can be adapted to express correctly in a single equation the relation between field properties and surface charges at both plates by using the *change* in field intensity and gradient experienced in passing through the surface from left to right, that is, in the direction of an increase in x, as follows:

$$\Delta F = +4\pi\sigma; \quad \text{or,} \quad \Delta\left(\frac{dE}{dx}\right) = -4\pi\sigma \qquad \text{(21 esu)}$$

At the left-hand plate, the gradient increases from zero inside to a positive value outside, so that its change is positive; the surface charge is in fact negative as called for by Equation (21). At the right-hand plate, the gradient changes from a positive value outside to zero inside, hence the change in crossing the surface in the plus x direction is negative; the surface charge is in fact positive as demanded by the equation.

The preceding discussion may be summarized by the statement that *the change in potential gradient in passing through a surface is a direct measure of the charge density on the surface.* This statement apparently calls for a layer of surface charge along the boundary between two substances having different dielectric constants. In a detailed study of the properties of dielectrics[B V] it is found that the existence of dielectric constants greater than unity is due to electric polarization of the medium. This polarization results in the appearance at the boundaries of dielectrics of " bound " surface charges of exactly the kind called for above.

Any solid sphere may be thought of as made up of a very large number of hollow spherical shells, any solid cylinder as composed of many hollow cylindrical shells, and any rectangular block as composed of a stack of thin sheets. Similarly the space charge in a one-dimensional field may be thought of as consisting of many very thin layers of space charge; the space-charge density in successive layers may or may not be the same. If the layers are thin enough they resemble surface charges, so that the reasoning associated with Equation (21) may be adapted to give a graphical picture of the meaning of Equation (17).

Fig. 6 illustrates the nature of the potential change to be expected in passing through one such thin layer. If the charge content of the layer is described as σ statcoulombs per square centimeter of area, Equation (21) can be applied to relate the change in potential gradient in passing through the layer to the charge content, as follows:
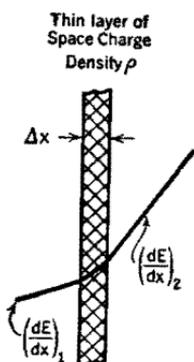
FIG 6. Change in potential gradient in passing through a thin layer of space charge.

$\left(\dfrac{dE}{dx}\right)_1$ and $\left(\dfrac{dE}{dx}\right)_2$ signify the slopes of the potential line.

$$\Delta\left(\frac{dE}{dx}\right) = \left(\frac{dE}{dx}\right)_2 - \left(\frac{dE}{dx}\right)_1 = -4\pi\sigma \qquad (22 \text{ esu})$$

But the charge content of this thin layer can also be described by stating the value of the space-charge density ρ within it, in statcoulombs per cubic centimeter. In order to use this description the symbol σ must be replaced by $\rho \, \Delta x$, which is of course the charge

within one square centimeter of area of the layer whose thickness is now specified as Δx. With this change, Equation (22) becomes

$$\left(\frac{dE}{dx}\right)_2 - \left(\frac{dE}{dx}\right)_1 = \Delta\left(\frac{dE}{dx}\right)_1 = -4\pi\rho\,\Delta x \qquad \text{(23 esu)}$$

If now the thickness of the layer of charge is made to approach zero as a limit, the potential change $\Delta(dE/dx)$ and the thickness Δx are written $d(dE/dx)$ and dx, and Equation (17) results.

Equation (20) may be thought of as the special case of Equation (17) that arises when space-charge density increases to infinity and the volume it occupies vanishes, their product remaining constant at the value of σ.

8. Potential ǃ Diagrams for One-Dimensional Fields in Regions Containing Space Charge. Fig. 7 illustrates a shape that the space-potential or potential distribution curve in the one-dimensional region between two parallel plates may have when the intermediate region contains *uniform* space-charge density. For reasons stated later the situation so represented can never persist for more than a minute fraction of a second, but it is nevertheless serviceable for illustrative purposes. Two successive integrations of Equation (17), using a constant value for ρ
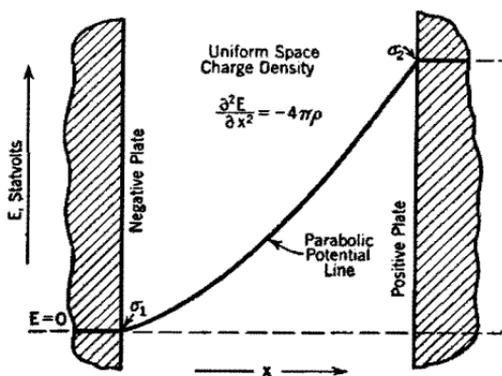
FIG. 7. Typical potential distribution in a one-dimensional region containing uniform space charge density.

and taking the origin at the negative plate, show that the potential line in Fig. 7 must be a parabola (second degree equation) of the general form

$$E = -2\pi\rho x^2 + C_1 x + C_2 \qquad \text{(24 esu)}$$

But when $\qquad x = 0,\ E = 0\ $ and $\ \dfrac{dE}{dx} = C_1$

so that $\qquad C_2 = 0\ $ and $\ C_1 = \left(\dfrac{dE}{dx}\right)_{x=0}$

But $\left(\dfrac{dE}{dx}\right)_{x=0} = -4\pi\sigma_1$, so that

$$E = -2\pi\rho x^2 - 4\pi\sigma_1 x \qquad \text{(25 esu)}$$

where σ_1 is the surface-charge density at the more negative plate.

The *convex-downward* flexion of the potential line in Fig. 7 corresponds to a *negative* space charge, such as would result from a uniform concentration of electrons. A *positive* space charge would result in a parabola *convex upward*. The sharp bend at the left-hand end can also be described as convex downward; correspondingly the surface charge σ_1 is negative. At the right-hand surface there is a sharp convex-upward bend, indicating a positive surface charge.

If interelectrode space charge is due to electrons, they must move in response to the force exerted on them by the electric field whose form they help to determine. For that reason *a uniform electron concentration can exist only for an extremely short time or under highly artificial experimental conditions.*

No matter whether space-charge density is uniform or not, the flexion of the potential line is at every point, for one-dimensional geometry, proportional to the negative of the space-charge density, and the
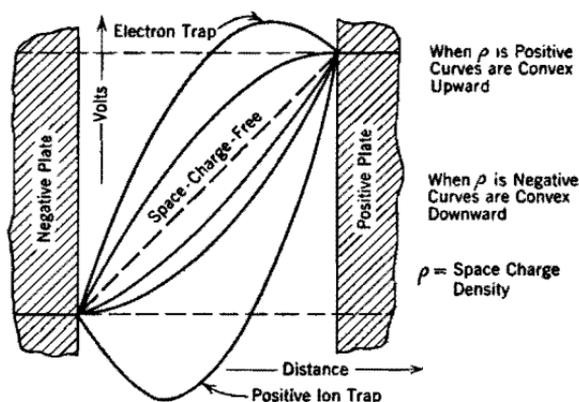
FIG. 8.   Potential distributions for various values of uniform space charge density, parallel plane electrodes.

abrupt change of slope at a surface is a direct measure of the surface-charge density.

Fig. 8 illustrates a variety of uniform space-charge-density conditions, all having the same overall potential difference between the plates, but with varying amounts and kinds of space charge. Of particular interest are the two curves, one convex downward, the other convex upward, in which the potential line is horizontal at one of the surfaces. Such horizontal entrance indicates zero surface charge at that point, the total space charge then being equal to the surface charge on the other electrode.

Of interest too is the uppermost curve, corresponding to a rather large *positive* space-charge density. If a single electron were set free

in such a field, it would be accelerated toward the highest point of this curve, the vertex of the parabola. An electron introduced into any part of the region whose potential is higher than that of the more positive plate could not escape to either plate. Because of its inertia, such an electron would oscillate near the vertex, like a ball at the bottom of a parabolic trough. Of course a sufficient number of electrons so placed would tend to neutralize the positive space charge and reduce the height of the vertex.

9. Spherical and Cylindrical Coordinates. Occasionally the following general forms taken by Poisson's equation when referred to spherical and cylindrical coordinate systems are useful:

Spherical coordinates, as in Fig. 9:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial E}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial E}{\partial\theta}\right)$$

$$+ \frac{1}{r^2\sin^2\theta}\frac{\partial^2 E}{\partial\phi^2} = -4\pi\rho \qquad (26 \text{ esu})$$

Cylindrical coordinates, as in Fig. 10:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E}{\partial\phi^2} + \frac{\partial^2 E}{\partial z^2} = -4\pi\rho \qquad (27 \text{ esu})$$

An important special type of electric field is that in which field properties depend solely on the radial distance from an axis. Since
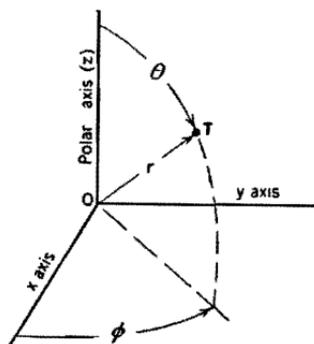
Fig. 9. Spherical coordinates of the point T.
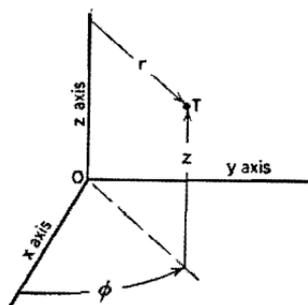
Fig. 10. Cylindrical coordinates of the point T.

potential and space charge in this type of field do not vary from point to point *around* the axis of symmetry or with position along the axis, the cylindrical-coordinate form of Poisson's equation reduces to:

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dE}{dr}\right) = -4\pi\rho \qquad (28 \text{ esu})$$

It is sometimes useful for graphical purposes to convert Equation (28) into the form

$$\frac{d^2E}{(d \log r)^2} = -2\rho'r = -2\rho'\epsilon^{\log r} \qquad (29 \text{ esu})$$

where $\rho' = 2\pi\rho r$, representing the space-charge density integrated around the axis. If a potential E that depends on radius only is plotted against $\log r$ the resulting curve is a straight line wherever space charge is absent; it is convex downward in regions of negative space charge and convex upward in regions of positive space charge.

The flexion of such a semi-logarithmic curve is proportional to the *moment* of the space-charge density integrated around the axis, not just to space-charge density, so that flexion of given magnitude near to the inner cylinder corresponds to a very much greater space-charge density than does the same flexion at a point near the surface of the outer cylinder. Similarly the slope or first derivative of such a semi-logarithmic curve is the moment of the electric field, for

$$\frac{dE}{d \log r} = r\frac{dE}{dr} = -rF \qquad (30 \text{ esu})$$

These relations recall the familiar engineering fact that moments of quantities are significant in any mathematical analysis of force actions in mechanical parts that have cylindrical symmetry.

PROBLEMS

CHAPTER I

1. The potential difference between the infinite parallel plates of Fig. 8 is 3 statvolts, and the spacing between the plates 1.5 cm. A uniform space charge density that is sufficient to make the potential gradient just outside the negative plate zero exists in the intermediate region.

(a) Find the magnitude of the uniform space charge density, in statcoulombs per cubic centimeter.

(b) Find the corresponding electron concentration (number of electrons per cubic centimeter). See Table XVII for the value of an electron's charge.

(c) Find the potential gradient, in statvolts per centimeter, just outside the positive plate.

(d) Find the surface charge density on the positive plate, in statcoulombs per square centimeter.

(e) Find the space charge, in statcoulombs, within a volume that extends from one plate to the other, and has a cross-sectional area of 1 sq cm.

(f) What acceleration, in centimeters per second per second, is experienced by an electron when at a point midway between the two plates? (See Table XVII for values of electronic charge and mass.)

2. Suppose that the potential between the plates of Fig. 8 is as described by the

lowest potential distribution line there shown. What is the direction of the force on an electron one-quarter of the way from the negative toward the positive plate? On one three-quarters of the way? What is happening to the space-charge density at the vertex of the potential line?

3. Suppose that the potential distribution between the pair of infinite parallel plates spaced 1.5 cm apart, as in Problem 1, is described by the relation $E = 2.25x^{\frac{1}{4}}$ (esu). Find the new values of the quantities asked for in parts (d) and (e) of Problem 1; also evaluate the quantity $\rho \sqrt{E}$ at $x = 0$, at $x = 0.75$, and at $x = 1.5$.

4. Suppose that at a point midway between grid-wire centers in Fig. 4a the *space-charge-free* value of the flexion, d^2E/dx^2, of the x-profile potential distribution curve is 20 statvolts per centimeter per centimeter. According to Poisson's equation the introduction of flowing negative space charge (electrons) makes the algebraic sum of the x-profile and y-profile flexions become more positive. If the total change due to the introduction of space charge is equally divided between the two flexions, what is the value of each at the specified point when ρ has the value -0.1 statcoulomb per sq cm?

5. The radii of two concentric cylinders are respectively 0.04 and 0.60 cm. The potential of the inner cylinder is zero, and that of the outer is $+1.2$ statvolts.

(a) Find τ_c, the charge on the inner cylinder, in statcoulombs per centimeter of axial length.

(b) Find the surface charge densities on the two electrodes.

(c) Find the potential gradient just outside the inner cylinder and that just inside the outer one.

(d) Plot a curve of E vs. r (similar to Fig. 33b).

CHAPTER II

THE ELECTROSTATIC FIELD OF A TRIODE

10. Conformal Transformations. The analysis of apparently complicated two-dimensional electric and magnetic fields is often facilitated by *conformal transformation*[B 264 YY] of the actual coordinate system to an equivalent one which makes the proporties of the field easily determinable by comparison with familiar geometries. Such treatment greatly simplifies the study of the electric field within a triode in the absence of space charge; the effect of space charge can be introduced subsequently. The subject-matter of this section is the conformal transformation of coordinates as applied to any electric or magnetic field problem; later sections make the triode application.

If x and y are the coordinates of the original field, and x' and y' those of the equivalent field, the useful type of transformation is such that the underlying two-dimensional differential equation

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} = 0 \qquad (31 \text{ esu})$$

transforms into the similar one

$$\frac{\partial^2 E}{\partial x'^2} + \frac{\partial^2 E}{\partial y'^2} = 0 \qquad (32 \text{ esu})$$

That is, the Laplace equation must be satisfied in both the actual and equivalent systems. All transformations that can be described by an equation of the following general type satisfy this requirement, and are called *conformal transformations:*

$$x + jy = f(x' + jy') \qquad (33)$$

Here j is the complex-number operator, that is, $j = \sqrt{-1}$, and $f(x' + jy')$ may be any function at all of $(x' + jy')$. It might, for example, be convenient in a particular case to use the transformation

$$x + jy = (x' + jy')^{\frac{2}{3}} \qquad (34)$$

It is not difficult to prove that any transformation of the type indicated by Equation (33) makes the potential E satisfy the Laplace equation in both systems if it does in either of them. [B 261 YY] The proof will not be given here, however, because the immediate concern is with the *use* of the conformal-transformation method of field analysis.

The quantity $x + jy$ in Equations (33) and (34) describes a vector whose horizontal and vertical components in the original coordinate system are x, y, while $x' + jy'$ describes similarly a vector with x', y', horizontal and vertical components in the transformed system. Every possible vector in the original coordinate system has its counterpart in the equivalent system.

A graphical illustration of the nature of a conformal transformation appears in Fig. 11. The left-hand diagram in this figure represents, not at all an electric field, but simply a sheet of graph paper that is ruled rectangularly in the ordinary way, coordinates being x and y relative to a central origin O. Now let the sheet of paper be cut along the heavy horizontal line from the origin to the extreme left, and progressively opened up along the cut by a systematic bending of the coordinate lines, of the nature indicated by Figs. 11b and 11c. Each of the coordinate lines becomes a curve, as illustrated by those for $x = a$ and $x = b$.

Any such systematic warping of the original regular pattern is the equivalent of a transformation of coordinates. The particular kinds of warping that lead to Figs. 11b and 11c correspond to *conformal* transformations. It is characteristic of the conformal treatment that 90° intersections of the original coordinate lines remain 90° intersections after transformation, with the exception of unique points (" singular points ") such as the origin in Fig. 11a. Initially square blocks become " curvilinear squares " of the type used in freehand mapping of electric and magnetic fields.[A VI, C] The smaller the square block considered, the more nearly it retains the original shape.

The point T in Fig. 11a has the coordinates $x = a$ and $y = b$. The point T' to which T transforms in Fig. 11b can also be identified as that for which $x = a$ and $y = b$, or it can be described in the new coordinates as the point for which $x' = a'$ and $y' = b'$. The point T is the terminus of the vector $a + jb$, horizontal and vertical components being a and b; the point T' is similarly the terminus of the vector $a' + jb'$, with horizontal and vertical components a' and b' in the new system. The relation between Figs. 11b and 11a is expressed by Equation (34); as applied to the particular points T and T' it requires that the two vectors be so related that

$$a + jb = (a' + jb')^{\frac{3}{2}} \tag{35}$$

An identical relation exists between *any* two corresponding points or vectors; that is the meaning of Equation (34).

Figs. 11a and 11c are related by the equation

$$x + jy = (x'' + jy'')^2 \tag{36}$$

The mathematical equations of the families of curved lines on Fig. 11c, into which the original straight coordinate lines are transformed, can be obtained by expanding Equation (36) into the form

$$x + jy = x''^2 - y''^2 + 2x''y'' \tag{37}$$
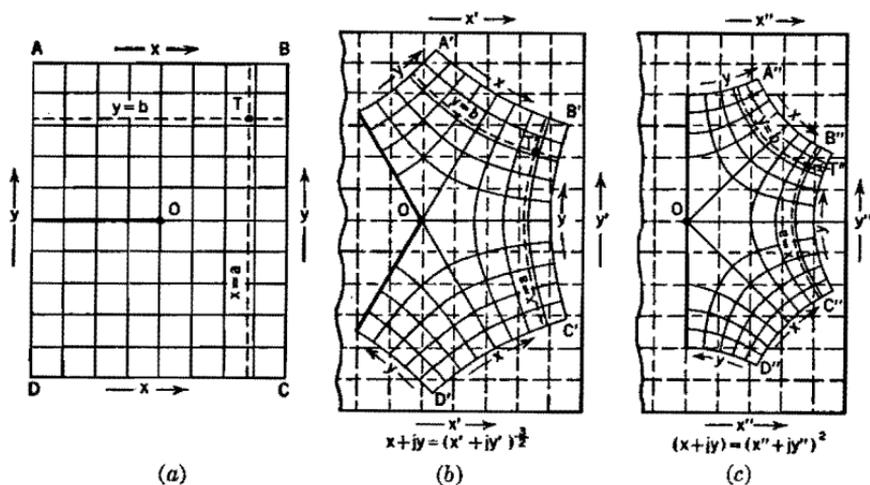
FIG. 11.   Conformal transformation of a rectangular sheet marked with rectangular coordinates.
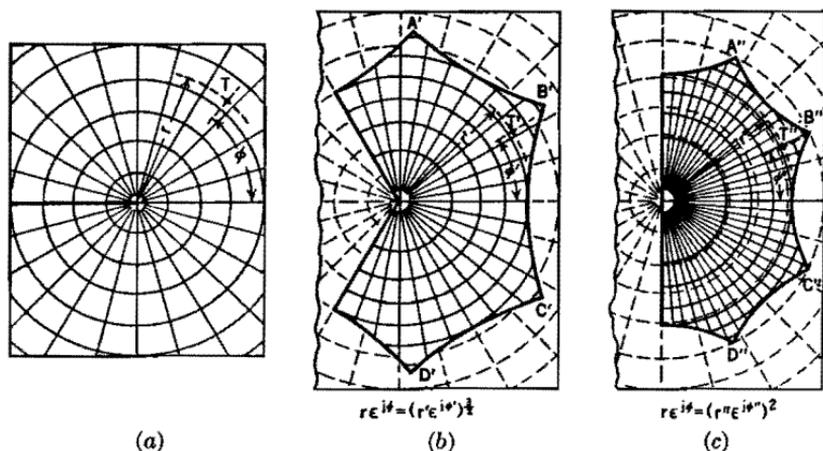
FIG. 12.   Conformal transformation of a rectangular sheet marked with polar coordinates.

Real and imaginary parts can then be separately equated, as in any expression involving imaginary quantities, so that the equations

$$x = x''^2 - y''^2 \tag{38}$$

$$y = 2x''y'' \tag{39}$$

describe the positions taken on the x'', y'' coordinates by the original x, y lines.

It is often desirable to describe conformal transformations in polar coordinates. The vector terminating at any point such as T, Fig. 11a and Fig. 12a, is equally well described by the expressions $x + jy$ and $r\epsilon^{j\phi}$, r being the radius vector and ϕ the angle with the x axis, that is

$$r = \sqrt{x^2 + y^2}$$
$$\phi = \text{arc cos}\frac{x}{r} = \text{arc sin }\frac{y}{r} \tag{40}$$
$$x = r \cos \phi$$
$$y = r \sin \phi$$

Similarly the vector terminating at T' in the transformed diagram can be described as $r'\epsilon^{j\phi'}$, related to $x' + jy'$ in exactly the way that $r\epsilon^{j\phi}$ is to $x + jy$. It is sometimes helpful to recall that the product of two polar vectors $r_1\epsilon^{j\phi_1}$ and $r_2\epsilon^{j\phi_2}$ is the vector $r_1r_2\epsilon^{j(\phi_1+\phi_2)}$. Figs. 12a, b, c represent a sheet of graph paper ruled with polar coordinates and given exactly the treatment illustrated in Fig. 11, that is,

$$r\epsilon^{j\phi} = (r'\epsilon^{j\phi'})^{\frac{3}{2}} \tag{41}$$
$$r\epsilon^{j\phi} = (r''\epsilon^{j\phi''})^2 \tag{42}$$

Figs. 11 and 12 merely contrast two different ways of marking a flexible sheet; the cutting and warping is exactly the same in the two illustrations.
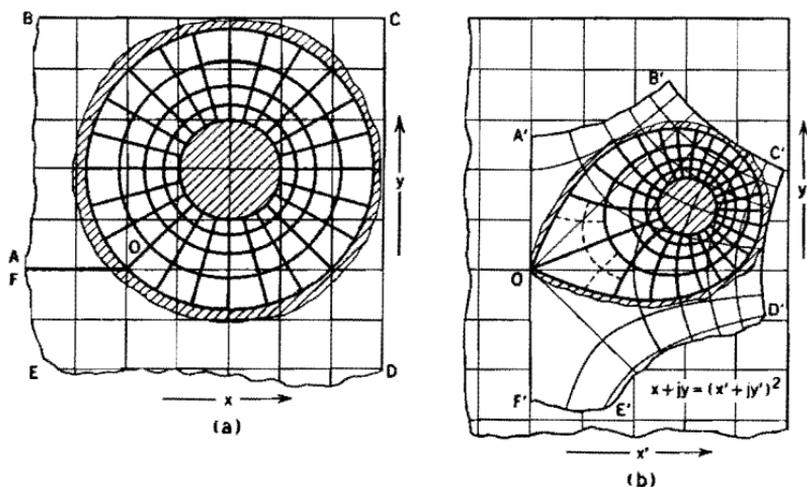
Fig. 13. Conformal transformation of an electrostatic field map, originally cylindrical.

The original x, y, graph may have drawn on it the equipotential and flux-line pattern of a two-dimensional electrostatic field. Such a

pattern must, in the absence of space charge, satisfy Equation (31) in the interelectrode region, and possess at the electrode surfaces whatever potentials the problem requires. If the transformation is a conformal one, the transformed equipotentials and flux lines will satisfy Equation (32), and will fit the warped locations of the original boundaries.

Figs. 13a and b illustrate the transformation of the flux pattern between concentric circular cylinders on the x, y, graph into that between corresponding oval-sectioned cylinders at the same potential difference on the x', y', set. If, as in this figure, the warping of the coordinate-set on which the field is mapped corresponds to a *conformal transformation*, the total charge on each electrode, the potentials between them, all capacitances, including any mutual capacitance coefficients, total energy storage, and all other overall properties of the electric field carry over without change from one representation to the other. The potential gradients, surface-charge densities, energy storages per unit volume, and other detail quantities are very greatly modified. Either field can be completely understood, however, either as to detail or overall properties, from analysis of the other.

11. Conformal Transformation of a Parallel-Plane Triode.[1,2,3,4,5,6,7,D]
Fig. 14 represents the three electrodes of a parallel-plane triode, drawn on the x, y, r, φ graph, which will sometimes be called the "W" plane in accordance with common usage. Grid-wire radius is R centimeters, and the grid center plane is a centimeters from the cathode, b centimeters from the plate. Fig. 15 represents a restricted portion of the same set of electrodes transferred to an x', y', r', ϕ', coordinate set, located on the "Z" plane. The conformal transformation used is[1]

$$2\pi nx + 2\pi njy = \log (x' + jy') \tag{43}$$

The x, y, origin is taken at the center of a grid wire; n is the number of intervals between grid wires in each centimeter, $1/n$ the center distance between grid wires. The mathematical transfer of a given point on the W plane to its proper Z-plane location is best accomplished, for this particular transformation, by mixing the rectangular and polar notations, thus:

$$2\pi nx + 2\pi njy = \log (r'\epsilon^{j\phi'})$$
$$= \log r' + j\phi' \tag{44}$$

Separation of the real and imaginary parts in Equation (44) shows that

$$r' = \epsilon^{2\pi nx} \tag{45}$$

$$\phi' = 2\pi ny \tag{46}$$

For example, the W-plane origin transfers to the point

$$r' = \epsilon^0 = 1 \tag{47}$$

$$\phi' = 2\pi n \times 0 = 0 \tag{48}$$

$2\pi n$ is chosen as the multiplier for x and y in order to make $\phi' = \pi$ when $y = \frac{1}{2}n$, thus limiting the Z-plane representation to a portion of the W plane that extends half-way from the origin to each adjacent grid-wire center. The cathode surface becomes, on the Z plane, a small circle of radius $s_c = \epsilon^{-2\pi n a}$, and the plate becomes a very large circle of radius $s_p = \epsilon^{+2\pi n b}$.

The grid wire becomes a figure that is essentially circular if its actual W-plane diameter is not more than about one-sixth of the center-distance between grid wires.[1] The Z-plane grid-wire figure crosses the horizontal axis ($\phi' = 0$) at extreme inner and outer points; the radii to the crossing points are:

$$\text{Innermost: } s_1 = \epsilon^{-2\pi n R} \tag{49}$$

$$\text{Outermost: } s_2 = \epsilon^{+2\pi n R} \tag{50}$$

The radius of the transformed grid-wire figure, for those cases in which it can be considered circular, is:

$$R' = \tfrac{1}{2}(s_2 - s_1) = \tfrac{1}{2}(\epsilon^{2\pi n R} - \epsilon^{-2\pi n R}) = \sinh 2\pi n R \tag{51}$$

The distance between its center and the cathode center is:

$$s_g = \tfrac{1}{2}(s_2 + s_1) = \tfrac{1}{2}(\epsilon^{2\pi n R} + \epsilon^{-2\pi n R}) = \cosh 2\pi n R \tag{52}$$

The dashed circle on Fig. 14 is the outline of a grid wire for which $2nR = \frac{1}{2}$. Such a grid wire would be too large to be treated as a circle in the transformed position, as indicated by the corresponding dotted Z-plane figure.

The quantity $2nR$ will sometimes be called the *screening fraction*, as it describes the extent to which the grid plane is blockaded by grid wires.

12. Placement of Charges to Satisfy Triode Boundary Conditions. The analysis of the space-charge-free electric field within a triode like that diagrammed in Fig. 14 will be developed in terms of the Z-plane coordinates of Fig. 15, and subsequently expressed in terms of the real coordinate dimensions. At normal operating voltages the plate potential of a triode is much higher, and the grid potential a little lower, than that of the cathode, so there must be negative charges on the cathode and grid and a positive charge on the plate.

Of course the actual charges within the tube reside on the surfaces of the electrodes (space charge is absent). However, if a set of *line* charges are imagined to be located at appropriate points in Fig. 15, the Z-plane cathode, grid, and plate circles, which describe electrode surfaces, can be identified with equipotentials of the field due to these line charges. Equations can then be written for the potentials everywhere in terms of the charges; the resulting expressions can be used to describe the

differences between the potentials of the various electrodes, and also to describe the potential distribution in the regions between the electrodes. The magnitudes of the charges may be selected to satisfy any given set of interelectrode potential differences.
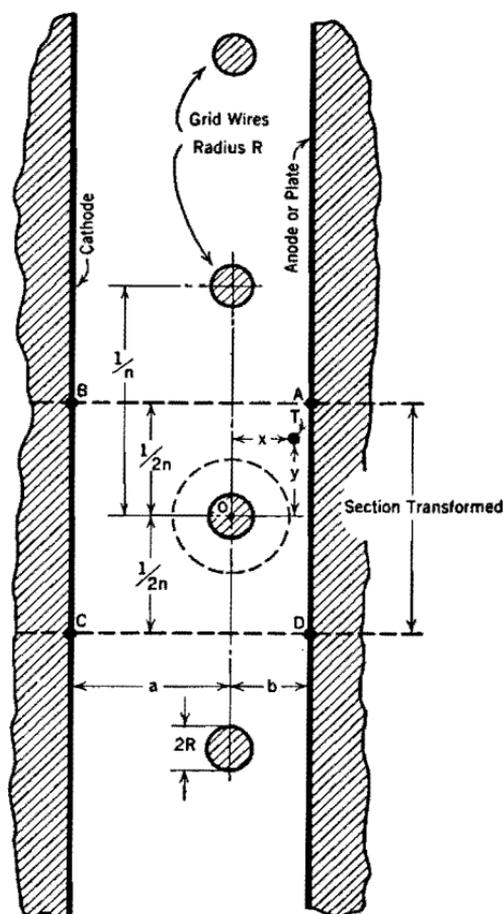
Fig. 14.   Parallel plane triode. x and y are coordinates of any point T relative to origin at grid-wire center. n grid wires per cm. Points A, B, C, D on this " W "-plane representation transform to points A', B', C', D' on the " Z "-plane representation, Fig. 15.

In any electrostatic field 4π flux lines attach to each statcoulomb of charge.   Symmetry requires all the flux lines around an infinite line charge like that of Fig. 16 to be radial; therefore $4\pi\tau$ flux lines must radiate from each centimeter length of a line charge of τ statcoulombs per centimeter length.   A cylinder having, like that in Fig. 16, radius r and length one centimeter, has a circumferential area of $2\pi r$ square centimeters.   This circumferential area is penetrated by $4\pi\tau$ flux lines;

therefore the flux density at radius r is $4\pi\tau/2\pi r = 2\tau/r$. If, as in all problems here dealt with, the dielectric constant is unity, the electric field strength F is the same as the flux density, so that

$$F = \frac{2\tau}{r} \qquad \text{(53 esu)}$$
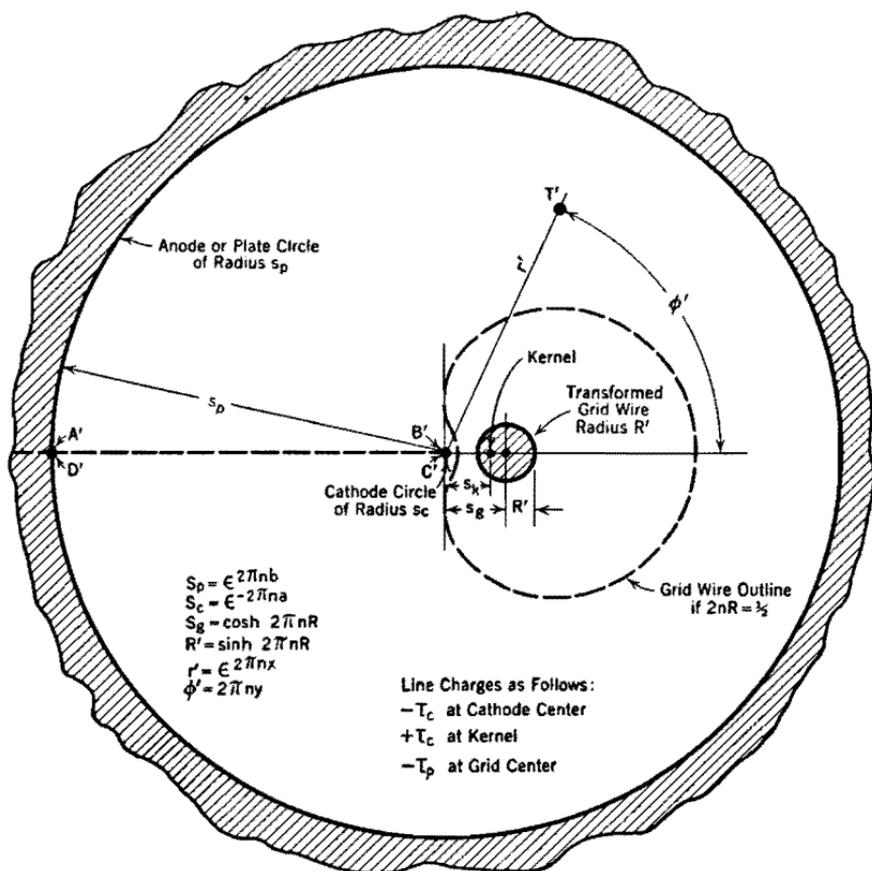
FIG. 15. Transformed or "Z-plane" representation of Section $ABCD$ of the triode of Fig. 14.

Electrostatic potential is the negative line integral of intensity, so that the potential difference between two points at radii r_1 and r_2 in the field of this line charge is expressed as follows:

$$E_2 - E_1 = -\int_{r_1}^{r_2} \frac{2\tau}{r}\, dr \qquad \text{(54 esu)}$$

or

$$E_2 - E_1 = -2\tau \log\frac{r_2}{r_1} \qquad \text{(55 esu)}$$

Equation (55) is a very general expression that will be used repeatedly in the following discussions of potential distributions.

The dimensions of all ordinary tubes are such that the Z-plane cathode circle (more properly *cylinder*) is very small, so that it is an equipotential,

$4\pi T$ flux lines must emerge through circumferential surface, of area $2\pi r$ square cm.

Hence $\begin{cases} D = \dfrac{4\pi T}{2\pi r} = \dfrac{2T}{r} \\ F = \dfrac{2T}{kr} \end{cases}$

radius r cm.

Line Charge T statcoulombs per cm. length

1 cm.
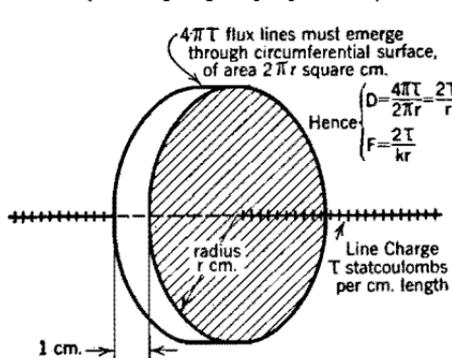
FIG. 16. One-centimeter cylinder around a line charge. D = electric flux density, F = electric field intensity, k = dielectric constant.

regardless of the effect of grid and plate charges, if a negative line charge $-\tau_c$ statcoulombs per centimeter of length is located along its axis. This line charge may then be *imaged* in the grid cylinder in order to keep the latter an equipotential. To this end there must be located at the *kernel* of the Z-plane grid circle a line charge $+\tau_c$, equal and opposite to that at cathode center. The position of the kernel is determined by the general image relation, applicable either in imaging a point in a sphere[A 120] or a line in a cylinder, that the product of the distances from the inside and outside charge locations to the center of the sphere or cylinder is equal to the square of the latter's radius. The proof of the correctness of this relation in the case of a line charge and a cylinder is as follows:
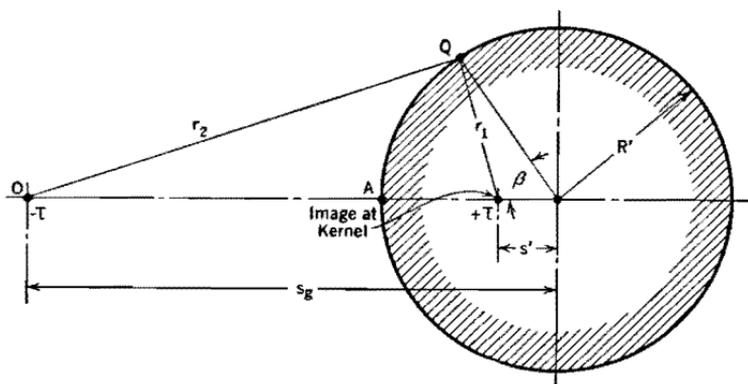
FIG. 17.   Image of a line charge $-\tau$ in a cylinder.

The circle of Fig. 17 represents a cylindrical surface of radius R'. It is to be demonstrated that it is an equipotential surface of the electric field produced by two line charges, $-\tau$ and $+\tau$, that are parallel to and in a plane with the cylinder axis, provided that $s_g s' = R'^2$, in accordance with the image relation stated above. Q is any point on the cylinder,

A a point in the plane of the charges. The potential difference between Q and A *owing to* $-\tau$ is

$$E_Q - E_A = -2\ (-\tau) \log \frac{r_2}{s_g - R'} \qquad (56\ \text{esu})$$

Owing to $+\tau$ at the kernel, it is:

$$E_Q - E_A = -2\ (+\tau) \log \frac{r_1}{R' - s'} \qquad (57\ \text{esu})$$

The cylinder is to be an equipotential. Therefore the total difference of potential between Q and A, which is the algebraic sum of Equations (56) and (57), must vanish. This requirement is satisfied only if

$$\frac{r_2}{s_g - R'} = \frac{r_1}{R' - s'} \qquad (58)$$

Equation (58) can also be written

$$\frac{\sqrt{R'^2 + s_g{}^2 - 2R's_g \cos \beta}}{s_g - R'} = \frac{\sqrt{R'^2 + s'^2 - 2R's' \cos \beta}}{R' - s'} \qquad (59)$$

The solution of Equation (59) for s' is straightforward, but a little tedious; the result is

$$s' = \frac{R'^2}{s_g}, \quad \text{or,} \quad s's_g = R'^2 \qquad (60)$$

The correctness of this solution can be quickly shown by substituting R'^2/s_g for s', then factoring R'/s_g from right-hand numerator and denominator; an identity results.

In the present application to Fig. 15, the charge outside the grid cylinder, $-\tau_c$ statcoulombs per centimeter of length of axis, lies along the cathode center, distant s_g from the axis of the grid cylinder. The radius of the grid cylinder is R'. If s_k is used to symbolize the distance from the cathode to the kernel or image-charge location,

$$s_g(s_g - s_k) = R'^2 \qquad (61)$$

which may also be written

$$s_k = \frac{s_g{}^2 - R'^2}{s_g} \qquad (62)$$

In the present application

$$s_g{}^2 - R'^2 = \cosh^2 2\pi nR - \sinh^2 2\pi nR = 1 \qquad (63)$$

so that

$$s_k = \frac{1}{s_g} \qquad (64)$$

With just the two line charges $-\tau_c$ and $+\tau_c$ present the electric field near cathode and grid is as illustrated in Fig. 18. The cathode charge is negative, as under normal conditions; but the plate has as yet no charge and the grid a positive charge. By locating a third line charge $-\tau_p$ along the grid-cylinder axis these disagreements with normality can be corrected *without disturbing the equipotential nature of the grid figure.* The quantity $-\tau_c$ usually has a negative numerical value, and $+\tau_p$ a positive one. However, when the grid voltage is below the cut-off value, as in Fig. 4d, $-\tau_c$ has a positive numerical value; similarly $+\tau_p$ may have a negative value if the plate is at a lower potential than the grid.

FIG. 18.   Z-plane triode flux pattern with equal and opposite charges on cathode and grid; none on plate. The plate circle is remote with center at cathode.

Tube dimensions are usually such that the Z-plane plate circle is very remote from the grid circle, so much so that both grid and cathode figures, so also *all three* line charges, are for all practical purpose at the axis of the plate cylinder.
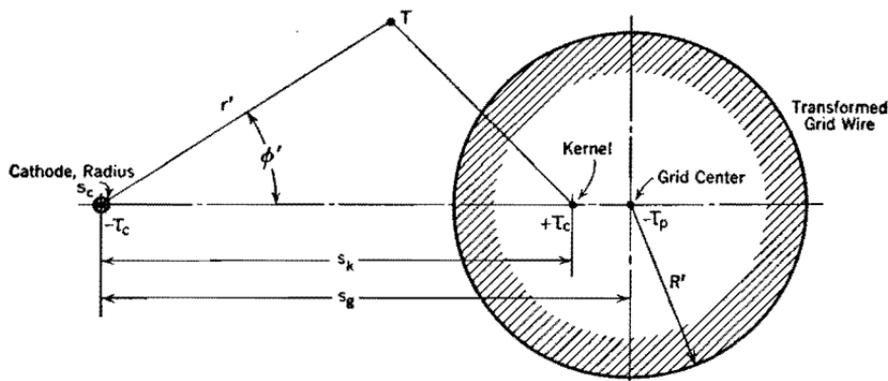
FIG. 19.   Detail of placement of Z-plane charges in Fig. 15.   Plate circle remote with center at cathode.

Hence the electric field in the neighborhood of the plate may be treated as though due to the net grid-and-cathode charge $-\tau_p$ located along the plate cylinder axis.   Such a field terminates in a charge $+\tau_p$ stat-

coulombs per centimeter of cylinder length distributed uniformly over the interior of the plate cylinder, and makes the latter an equipotential surface.

The array of three line charges, $-\tau_c$, $+\tau_c$, and $-\tau_p$, detailed in Fig. 19, makes all the Z-plane electrode boundaries equipotentials for any and all values of $-\tau_c$ and $+\tau_p$; that is, the *geometrical* boundary conditions are satisfied. The plate-to-cathode and grid-to-cathode voltages can be given any desired values by proper choices of values for $-\tau_c$ and $+\tau_p$, in this way satisfying the *potential* boundary conditions. The charges on the electrodes, in statcoulombs per centimeter length of the Z-plane figure, are as follows:

$$\text{Cathode} \quad -\tau_c$$
$$\text{Plate} \qquad +\tau_p \qquad\qquad \text{(65 esu)}$$
$$\text{Grid} \qquad +\tau_c - \tau_p$$

Since the grid has ordinarily a negative charge, it is to be expected that τ_p will usually be numerically greater than τ_c.

13. Equations for Space-Charge-Free Potential Distribution. The point T in Figs. 15 and 19 might be located anywhere in the region between the electrodes. It has coordinates r' and ϕ'. Equation (55) can be used to express, first, the potential difference between T and the cathode due to the cathode-center charge, second, that due to the image charge, third, that due to the grid-center charge. Any point on the cathode surface may be used in setting up these three potential expressions, but it is necessary to use the same point for all of them, as is done in the following equations:

Due to $-\tau_c$ at the origin:

$$E_T - E_c = -2(-\tau_c) \log \frac{r'}{s_c} \qquad \text{(66 esu)}$$

Due to $+\tau_c$ at grid kernel:

$$E_T - E_c = -2(+\tau_c) \log \frac{\sqrt{r'^2 + s_k{}^2 - 2r's_k \cos \phi'}}{s_k - s_c} \qquad \text{(67 esu)}$$

Due to $-\tau_p$ at grid center:

$$E_T - E_c = -2(-\tau_p) \log \frac{\sqrt{r'^2 + s_g{}^2 - 2r's_g \cos \phi'}}{s_g - s_c} \qquad \text{(68 esu)}$$

The correct value of $E_T - E_c$, that due to all three charges, is the algebraic sum of these three expressions.

It is chiefly interesting to know the nature of the potential variation (a) along a line through the center of a grid wire, and (b) along a line midway between grid wires. The former is obtained by letting $y = 0$,

that is, $\phi' = 0$; it corresponds to the potential variation along a line extending horizontally to the right from the Z-plane origin. For the latter $y = \frac{1}{2}n$, that is, $\phi' = \pi$; it corresponds to the potential variation along a line horizontally to the left from the Z-plane origin. In one case $\cos \phi' = +1$, in the other $\cos \phi' = -1$, so that in both cases the radicals are the square roots of perfect squares, being respectively the sum and difference of the distances involved. Using $\phi' = 0$, $\cos \phi' = +1$,

Due to all charges:

$$E_T - E_c = 2(-\tau_c) \log \frac{|r' - s_k|}{s_k - s_c} \frac{s_c}{r'} + 2\tau_p \log \frac{|r' - s_g|}{s_g - s_c} \quad \text{(69 esu)}$$

a = 0.6 cm.
b = 0.4 cm.
n = 2.22 grid wires/cm.
R = 0.01875 cm.
E_p = 127.8
E_g = −7.84
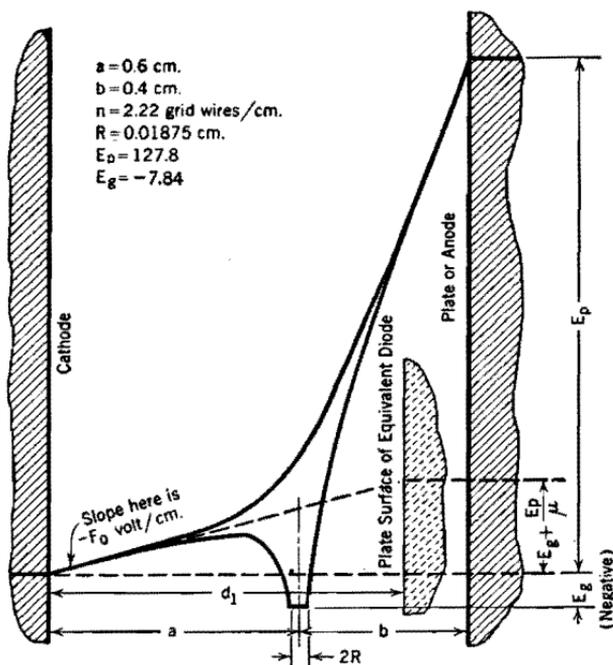
FIG. 20. Mathematically-determined potential distribution in a parallel-plane triode, showing relations between space-charge-free off-cathode gradient and tube dimensions. Compare with Fig. 4.

The upper solid curve describes the potential distribution along a path between grid wires, the lower curves, that along a path through grid-wire centers.

Absolute values of $r' - s_k$ and $r' - s_g$ must be used because only the magnitudes of the respective Z-plane distances are involved in the potential derivation; similarly the radicals in Equations (66, 67, 68) are all considered positive.

Equation (69) can be considerably simplified by making use of the

fact that s_c, which is much smaller than s_g and s_k, can be omitted from both denominators, and by using the relation $s_k = 1/s_g$. The simplified expression for potential variation along a path midway between grid wires is

$$E_T - E_c = 2(-\tau_c) \log (s_c s_g) \left| 1 - \frac{1}{r' s_g} \right| + 2(+\tau_p) \log \left| \frac{r'}{s_g} - 1 \right| \quad \text{(70 esu)}$$

For a path midway between grid wires $\phi' = \pi$, $\cos \phi' = -1$, so that the two negative signs become positive, giving

$$E_T - E_c = 2(-\tau_c) \log (s_c s_g) \left| 1 + \frac{1}{r' s_g} \right| + 2(+\tau_p) \log \left| \frac{r'}{s_g} + 1 \right| \quad \text{(71 esu)}$$

Equations (70) and (71) permit calculation of the space-charge-free potential for any value of x along the respective paths from cathode to plate, provided $-\tau_c$ and $+\tau_p$ are known. Fig. 20 contains a typical pair of potential distribution curves calculated in this way.

Points describing the potential distribution along paths between those for $y = 0$ and $y = \frac{1}{2}n$ can be determined by adding Equations (66), (67), and (68), then using properly selected values of ϕ' in calculating $E_T - E_c$. Curves through points so located lie between the two extreme curves shown in Fig. 20.

14. Charge Magnitudes. Suppose that the plate-to-cathode and grid-to-cathode potential differences are specified, and it is required to find the magnitudes of the charges $-\tau_c$ and $+\tau_p$. The path through the grid wire intersects the grid surface at two points; for the one nearest the origin $r' = s_g - R'$, and for the more remote one $r' = s_g + R'$. Substitution of either value into Equation (70) gives the equation for grid-to-cathode potential. Using the near point

$$E_g - E_c = 2(-\tau_c) \log (s_c s_g) \left| \frac{s_g^2 - s_g R' - 1}{s_g(s_g - R')} \right| + 2(+\tau_p) \log \left| \frac{s_g - R'}{s_g} - 1 \right|$$
$$\text{(72 esu)}$$

The second term in Equation (72) reduces to $2\tau_p \log (R'/s_g)$. In the first term

$$s_g^2 - 1 = \cosh^2 2\pi n R - 1 = \sinh^2 2\pi n R = R'^2 \quad \text{(73)}$$

so that $s_g^2 - s_g R' - 1 = R'(R' - s_g)$. The whole expression for grid-to-cathode potential thus reduces to

$$E_g - E_c = 2(-\tau_c) \log s_c R' + 2(+\tau_p) \log \frac{R'}{s_g} \quad \text{(74 esu)}$$

A similar expression for plate-to-cathode potential results if s_p is sub-

stituted for r' in Equation (70), and it is recognized that $s_p/s_g \gg 1$; the form taken is

$$E_p - E_c = 2(-\tau_c) \log s_c s_g + 2(+\tau_p) \log \frac{s_p}{s_g} \qquad \text{(75 esu)}$$

If the electrode geometry and voltages are given, Equations (74) and (75) can be solved simultaneously for the charges $-\tau_c$ and $+\tau_p$ which can then be used in Equations (70) and (71).

In solving for the charges it is convenient to make use of four *potential coefficients*, A_1, A_2, B_1, and B_2, defined in such a way that Equations (74) and (75) can be written as follows:

$$E_g - E_c = A_1(-\tau_c) + B_1(+\tau_p) \qquad \text{(76 esu)}$$
$$E_p - E_c = A_2(-\tau_c) + B_2(+\tau_p) \qquad \text{(77 esu)}$$

Evidently

$$A_1 = 2 \log s_c R' \qquad\qquad B_1 = 2 \log \frac{R'}{s_g}$$

$$A_2 = 2 \log s_c s_g \qquad\qquad B_2 = 2 \log \frac{s_p}{s_g} \qquad (78)$$

A_1, A_2, B_1 are normally negative; B_2 is normally positive (primarily because of the limitation of $2nR$ to values less than $\frac{1}{6}$).

15. Space-Charge-Free Off-Cathode Field Intensity. For reasons suggested in the introduction, and explained in detail in a later chapter, the normal tube current *when space charge is flowing* is primarily dependent on, though not proportional to, the value F_0 of the electric field strength that exists just off the cathode surface in the absence of space charge, that is, in the *space-charge-free* condition.   Since the space-charge-free potential gradient dE/dx is normally positive at the off-cathode location, F_0 is normally negative.   The sequence of potential distribution diagrams discussed in the introduction, Figs. 4a, 4b, 4c, 4d, illustrates the effect on F_0 of progressively reducing the grid voltage, the plate voltage remaining constant.

In Figure 4a the off-cathode gradient $(-F_0)$ is large, favoring the passage of a relatively large plate current as soon as the cathode is heated enough to convert the surface charge on the plate into space charge outside of it.   In Figure 4b $(-F_0)$ is small, favoring a small plate current.   If, as in Figure 4d, the off-cathode gradient is negative, of course no electrons can leave the cathode, no matter how hot it may be, for the electric field promptly pushes back any that may start to emerge.

Each of the mutual characteristic curves of Figure 3a is the result of a series of observations taken at successively more negative grid po-

tentials, the plate voltage remaining constant. As the grid potential is decreased the plate current becomes less and less, finally vanishing at the point called " cut-off." Figure 4c illustrates the potential distribution at cut-off; F_0 is just zero, and there is no charge on the cathode. If potentials are such as to make the off-cathode gradient even slightly positive, there will be current flow when the cathode is heated; if such as to make the off-cathode gradient negative, there will be none. Cutoff marks the division point between current and no-current conditions.

16. Dependence of F_0 on an Equivalent Voltage. The mathematical statement of the dependence of F_0 on tube geometry and potentials is arrived at by recognizing that it is proportional to $-\tau_c$, and that $-\tau_c$ can be expressed in terms of an *equivalent voltage* which is a composite of grid-to-cathode and plate-to-cathode potentials.

Let σ_c symbolize the space-charge-free surface charge density on the cathode, in statcoulombs per square centimeter. Then

$$\sigma_c = n(-\tau_c) \tag{79}$$

This relation grows out of the fact that each centimeter of axial length of the Z-plane figure corresponds to one centimeter length, perpendicular to the paper, of the actual parallel-plane triode, Fig. 14. Since there are n such sections per centimeter of height of the actual structure, the actual surface charge density must be n times the cathode charge per section, as stated by Equation (79).

As stated in the discussion preceding Equation (53), invariably 4π electric flux lines attach to each statcoulomb of charge, so that $4\pi\sigma_c$ flux lines must terminate on each square centimeter of cathode area. Since the dielectric constant is unity, the electric field intensity just off the cathode surface must have the same value as the flux density there, that is,

$$F_0 = 4\pi\sigma_c = 4\pi n(-\tau_c) \tag{80 esu}$$

In order to facilitate expressing F_0 in terms of an equivalent voltage, as in Equation (82) on the next page, the solution for $-\tau_c$ obtained from Equations (76) and (77) can be written as follows:

$$-\tau_c = \frac{E_g + \dfrac{E_p}{(-B_2/B_1)}}{A_1 + \dfrac{A_2}{(-B_2/B_1)}} \ . \tag{81 esu}$$

Here E_g and E_p stand for grid-to-cathode and plate-to-cathode *potential differences*, which were written as $E_g - E_c$ and $E_p - E_c$ in earlier equations, for example, in Equations (76) and (77).

The following is the most useful form of the *equivalent voltage* equation for F_0, in form similar to Equation (81):

$$-F_0 = \frac{E_g + \dfrac{E_p}{\mu}}{d_1} \tag{82}$$

Here $E_g + \dfrac{E_p}{\mu}$ is the equivalent voltage; the quantity μ, called the *amplification factor* of the triode, describes the relative effectiveness of grid and plate potentials in the determination of F_0 and so of plate current. μ is related to the potential coefficients as follows:

$$\mu = -\frac{B_2}{B_1} \tag{83}$$

μ is always positive, and is ordinarily greater than unity, that is, the closeness of the grid to the cathode usually more than offsets the fact that it is a grid rather than a solid plate. See also Equation (97).

The quantity d_1 is of course a distance in centimeters, as F_0 is always measured either in volts or statvolts per centimeter. d_1 will be referred to as the spacing of an *equivalent space-charge-free diode*, for it is the distance that must exist between two parallel metal plates (no grid present) in order that the space-charge-free potential gradient between them shall be $-F_0$ when their potential difference is the equivalent voltage. This is illustrated in Fig. 20. Comparison of Equation (82) with (80) and (81) combined shows that

$$d_1 = -\frac{1}{4\pi n}\left(A_1 + \frac{A_2}{\mu}\right) \tag{84}$$

Thus both μ and d_1 are properties of the potential coefficients, therefore dependent on tube geometry only, and not at all dependent on electrode potentials. The coefficients A_1 and A_2 in Equation (84) always have negative numerical values. See also Equation (98).

17. Electrostatic Coefficients; Cathode Charge Always Proportional to an Equivalent Voltage. The significance of Equations (76) and (77) and of the solutions obtained from them for $-\tau_c$ and $+\tau_p$ can be somewhat clarified by a brief study of potential and capacitance coefficients.[74, B 93, B 96] If three conductors are insulated from but near to one another, the placing of a charge on any one of them results in changes in the potentials of all three, the magnitude of the potential *change* on each being proportional to the amount of charge placed on the one. This same idea is expressed by saying that the potential of each conductor is a linear function of (not proportional to) each of the three charges. The mathematical expression of such behavior in

terms of the cathode, grid, and plate of a triode as a whole, not just the transformed section, is

$$E_c = Q_c P_{cc} + Q_g P_{gc} + Q_p P_{pc}$$
$$E_g = Q_c P_{cg} + Q_g P_{gg} + Q_p P_{pg} \qquad (85)$$
$$E_p = Q_c P_{cp} + Q_g P_{gp} + Q_p P_{pp}$$

E_c, E_g, and E_p are the respective absolute potentials of cathode, grid, and plate. Q_c is the entire cathode charge, and Q_p and Q_g are similarly the entire plate and grid charges.   The P's are potential coefficients of a more general type than the A's and B's of earlier sections, but like them have values dependent only on interelectrode geometry.   There are in reality only six P's, for invariably:

$$P_{gc} = P_{cg}; \quad P_{pc} = P_{cp}; \quad P_{gp} = P_{pg} \qquad (86)$$

In any real electronic device the algebraic sum of the three charges is zero, and interest centers around differences of potential rather than absolute potentials.   By subtracting the first equation from the other two, and eliminating Q_g by the substitution

$$Q_g = -(Q_c + Q_p) \qquad (87)$$

a pair of equations of the following type is obtained:

$$E_g - E_c = Q_c a_1 + Q_p b_1 \qquad (88)$$
$$E_p - E_c = Q_c a_2 + Q_p b_2 \qquad (89)$$

(Compare these with Equations (76) and (77) which apply to the transformed section only.)   The a's and b's, which are potential coefficients similar to the earlier A's and B's, are various additive and subtractive combinations of the P's.   Equations (88) and (89) can be solved for the Q's, thus

$$Q_c = \frac{(E_g - E_c)b_2 - (E_p - E_c)b_1}{a_1 b_2 - a_2 b_1} \qquad (90)$$

$$Q_p = \frac{(E_g - E_c)a_2 - (E_p - E_c)a_1}{a_1 b_2 - a_2 b_1} \qquad (91)$$

These expressions have the general form

$$Q_c = (E_g - E_c)C_{gc} + (E_p - E_c)C_{pc} \qquad (92)$$
$$Q_p = (E_g - E_c)C_{gp} + (E_p - E_c)C_{pp} \qquad (93)$$

The C's are called *capacitance coefficients*, and are of course dependent entirely upon internal tube geometry.

Equation (92) is very useful in that it shows that the charge on the

cathode is *always* a linear function of each of the two tube potentials that are commonly measured.   It can be rewritten as follows:

$$Q_c = C_{gc}\left[(E_g - E_c) + \frac{(E_p - E_c)}{C_{gc}/C_{pc}} \right] \tag{94}$$

So it appears that the cathode charge must be in general proportional to the *equivalent voltage:*

$$(E_g - E_c) + \frac{(E_p - E_c)}{C_{gc}/C_{pc}} \tag{95}$$

Since C_{gc} is almost invariably larger than C_{pc}, this equivalent voltage is the grid-to-cathode voltage plus a definite fraction of the plate-to-cathode voltage.   The magnitude of the fraction is entirely dependent on tube geometry and is, therefore, the same for all combinations of plate and grid potentials.

The reasoning just outlined is a generalization of that presented in the preceding analysis of a single transformation section.   Equations (76) and (77) are in fact the exact counterparts of Equations (88) and (89).   Consequently the solution for $-\tau_c$ can be put into a form similar to Equation (94).   Abbreviating $E_g - E_c$ and $E_p - E_c$ to E_g and E_p,

$$-\tau_c = C'_{gc}\left[E_g + \frac{E_p}{C'_{gc}/C'_{pc}} \right] \tag{96 esu}$$

The primes on the C's indicate that they are capacitance coefficients for a single transformation section only.   Thus Equation (82) can be multiplied through by $4\pi n$ to make it resemble Equation (96); then after rearranging it is apparent that

$$\mu = C'_{gc}/C'_{pc} \tag{97}$$

$$C'_{gc} = \frac{1}{4\pi n d_1} \tag{98 esu}$$

$1/n$ is the *area* of one centimeter depth of plate, for the grid section before transformation, and C'_{gc} just the capacitance of the corresponding area of the equivalent space-charge-free diode.

The C's of an entire tube are of course closely related to those for individual sections.   There are n transformation sections per square centimeter of cathode and plate area, so that if there are S square centimeters of area, it might be expected that C_{gc} and C_{pc} would be simply SnC'_{gc} and SnC'_{pc}.   Actually the capacitances between lead-in wires and electrode-supporting structures make the true values of C_{gc} and C_{pc} exceed this expectation.   The values of C_{gc} and C_{pc} are important in radio-frequency tube applications.[E, F, G, H, I]

The true overall capacitance coefficients can be measured without

great difficulty.   If, for example, one terminal of an alternating voltage supply is connected to the plate, the other to the grid and cathode (which must be cold), an alternating-current ammeter in the cathode lead wire measures the rate of change of Q_c.   For this arrangement $E_g - E_c$ is zero, so that Equation (92) reduces to

$$Q_c = (E_p - E_c)C_{pc}$$

C_{pc} can therefore be calculated if $E_p - E_c$ and the cathode current are measured.   The procedure just described is not the most convenient and practical method of measuring the overall capacitance coefficients, but it illustrates their nature, and description of it illustrates the possibility of direct measurement.

All of the C's for which both subscripts are the same, such as C_{pp} in Equation (93), have positive numerical values.   For if the grid and cathode are tied together, so that $E_g - E_c$ is zero, a positive value of the plate charge Q_p must correspond to a positive value of $E_p - E_c$. All the capacitance coefficients having unlike subscripts, such as C_{gc} and C_{gp}, are negative.   For if the plate and cathode are tied together, making $E_p - E_c$ zero, and a positive charge is located on the plate, all the flux lines whose plus ends attach to the plate must have their minus ends on the grid, making the latter negative in potential with respect to both plate and cathode.   Thus if $E_p - E_c$ is zero and Q_p positive, $E_g - E_c$ must be negative, so that C_{gp} must be negative.   Similar reasoning can be applied to all of the C's whose two subscripts are unlike.

18. Amplication Factor μ in Terms of Dimensions.   The amplification factor μ of a triode is the most important tube constant.   Equation (83) relates μ to interelectrode geometry by way of the potential coefficients, and Equation (97) relates it to the capacitance coefficients for a single transformation-section.   By using the values of the B's from Equations (78) in (83), the direct dependence on Z-plane quantities is found to be

$$\mu = \frac{\log \dfrac{s_p}{s_g}}{\log \dfrac{s_g}{R'}} \tag{99}$$

This equation for μ can be derived more directly[1] by making use of the fact, mentioned in Section 15, that F_0 is zero at cut-off; then, according to Equation (82), the equivalent voltage $E_g + \dfrac{E_p}{\mu}$ must vanish, that is,

$$at\ cut\text{-}off,\ \mu = -\frac{E_p}{E_g} \tag{100}$$

Since F_0 is zero at cut-off, the cathode charge must also be zero at cut-off; and it is possible to express E_g and E_p in terms of Z-plane

dimensions by assuming equal and opposite charges on grid and plate, none on the cathode. To maintain the equipotential nature of the Z-plane grid figure, the grid charge must be located along the grid-wire
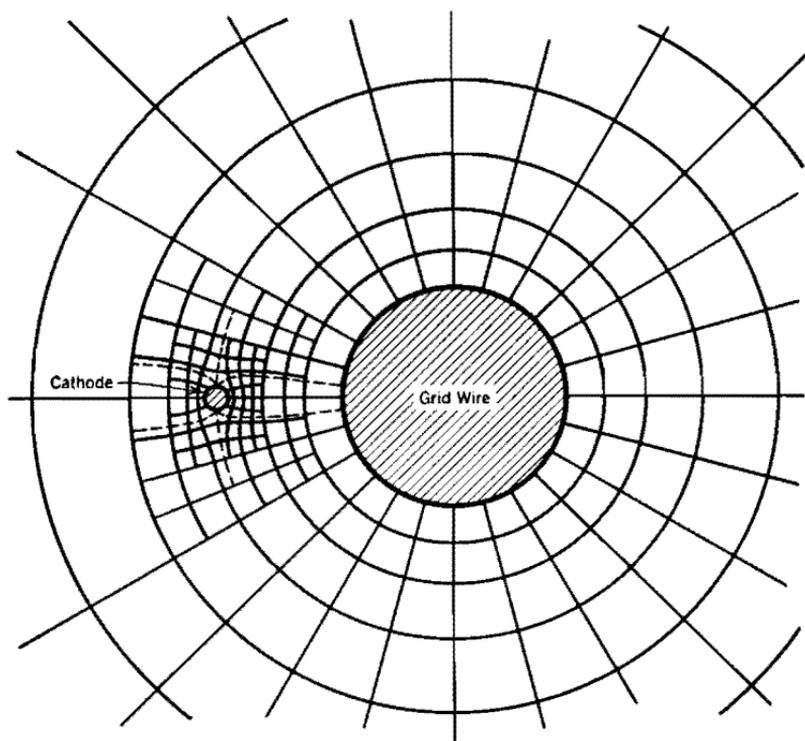
FIG. 21. Z-plane representation of the electric field map, in the neighborhood of grid and cathode, for a triode at cut-off.   No charge on cathode (see Fig. 4c).

center.   The resulting Z-plane field is illustrated in Fig. 21; it is essentially the field around a line charge on the transformed grid wire.   With $-\tau_p$ statcoulombs per centimeter along the grid center [compare Equations (66-67-68)]

$$E_g = +2\tau_p \log \frac{R'}{s_g} \qquad \text{(101 esu)}$$

and

$$E_p = +2\tau_p \log \frac{s_p}{s_g} \qquad \text{(102 esu)}$$

so that at cut-off

$$\mu = -\frac{E_p}{E_g} = -\frac{\log \dfrac{s_p}{s_g}}{\log \dfrac{R'}{s_g}} = +\frac{\log \dfrac{s_p}{s_g}}{\log \dfrac{s_g}{R'}} \qquad \text{(103)}$$

Equation (103) is identical with Equation (99).

Dependence of μ on actual triode dimensions is obtained by inserting into Equation (99) the proper expressions for R' and the various s's, from Section 11. Noting that $\log \epsilon^{2\pi nb} = 2\pi nb$, there results[18][E 178,]

$$\mu = \frac{2\pi nb - \log \cosh 2\pi nR}{\log \coth 2\pi nR} \qquad (104)$$

If $2\pi nR$ is very small, $\sinh 2\pi nR$ is indistinguishable from $2\pi nR$, and $\cosh 2\pi nR$ approaches unity; $\log \cosh 2\pi nR$ vanishes. Hence if the screening fraction $2nR$ is very small, the expression for μ simplifies to the following:

$$\mu = \frac{2\pi nb}{\log \dfrac{1}{2\pi nR}} \qquad (105)$$

For many commercial devices Equation (104) is appreciably more accurate than Equation (105).

These various expressions for μ describe a dependence only on geometry of the grid and plate structure, and not at all on the location of the cathode. The reason for the lack of dependence of μ on cathode position in these expressions is that the whole analysis is based on the assumption that the Z-plane cathode circle is very small. μ is the same, regardless of distance from grid to cathode, as long as that distance is large enough to make $s_c \ll s_g$.

The potential line passing through grid-wire center merges with that midway between grid wires as plate and cathode are approached, and, as illustrated in Fig. 20, the joint potential line is straight for some distance out from both plane electrodes. Since $F = -dE/dx$, and dE/dx is uniform over all parts of a straight line, the *electric field F and the flux density are practically uniform* for a considerable distance out from both plate and cathode. This is illustrated in the typical space-charge-free field map of Fig. 22.

As long as there is a region between grid and cathode in which the electric field approaches uniformity, Equation (99) for μ is valid regardless of the *form* as well as the location of the cathode. Fig. 23 is a field map for a triode in which the cathode is a set of filament wires parallel to the grid wires. The value of μ for this electrode arrangement, or for a similar one with the filament wires parallel instead of perpendicular to the plane of the paper, is exactly the same as it would be if the cathode were a plane surface at some such location as the equipotentials AA'.

19. Spacing of Equivalent Space-Charge-Free Diode. It was pointed out in the discussion following Equation (82) that the quantity d_1 appearing there is the spacing, between parallel plane electrodes, that

would be necessary *without a grid* to enable the equivalent voltage $E_g + \dfrac{E_p}{\mu}$ to produce the actual off-cathode space-charge-free gradient $-F_0$. This is illustrated by the dotted "equivalent diode plate" in
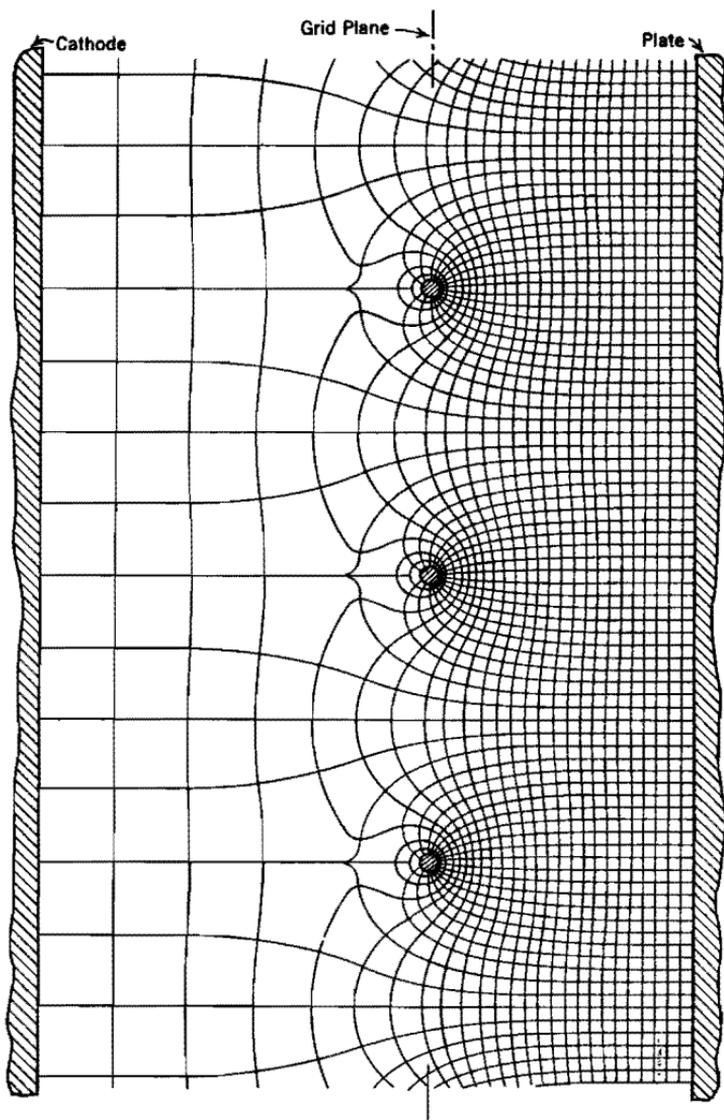
FIG. 22.   Typical space-charge-free electric field map for a parallel-plane triode, normal voltages.

Fig. 20. A line having the slope $-F_0$ reaches the potential $E_g + \dfrac{E_p}{\mu}$ at just the distance d_1 from the cathode; this distance is the same for any set of values of plate and grid voltages.

The equivalent space-charge-free diode spacing is expressed in terms of Z-plane dimensions by substituting values from Equations (78) for the A's of Equation (84). The result is:

$$d_1 = \frac{.1}{2\pi n} \log \frac{1}{s_c R'} + \frac{1}{2\pi n \mu} \log \frac{1}{s_c s_g} \qquad (106)$$
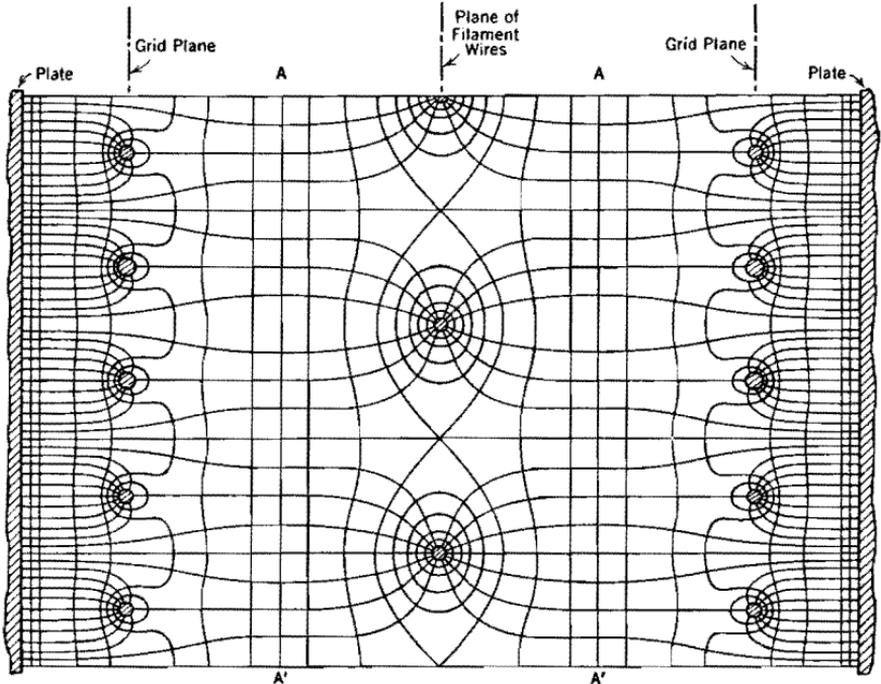
FIG. 23. Electric field map for a triode with filament wires parallel to grid wires, showing the intermediate equipotential planes AA'.

An alternate form of this can be obtained by eliminating R'. From the expression for μ, Equation (99),

$$\log R' = \log s_g - \frac{1}{\mu} \log \frac{s_p}{s_g} \qquad (107)$$

Equation (106) for d_1 can be written as follows:

$$d_1 = \frac{1}{2\pi n} \left(-\log s_c - \log R' + \frac{1}{\mu} \log \frac{1}{s_c s_g} \right)$$

and the value of $\log R'$ from Equation (107) inserted. The following expression for d_1 results:

$$d_1 = \frac{1}{2\pi n} \left[\log \frac{1}{s_c s_g} + \frac{1}{\mu} \log \frac{s_p}{s_c s_g{}^2} \right] \qquad (108)$$

The complete expression for F_0 with this new denominator is as follows:

$$-F_0 = \frac{E_g + \dfrac{E_p}{\mu}}{\dfrac{1}{2\pi n} \log \dfrac{1}{s_c s_g} + \dfrac{1}{2\pi n \mu} \cdot \log \dfrac{s_p}{s_c s_g{}^2}} \tag{109}$$

An interesting expression for d_1 in terms of actual triode dimensions is obtainable by using in Equation (108) proper values of the various s's, as follows:

$$d_1 = a + \frac{a + b}{\mu} - \frac{\mu + 2}{2\pi n \mu} \log \cosh 2\pi n R$$

or

$$d_1 = a \left[1 + \frac{1}{\mu} \frac{a + b}{a} - \frac{\mu + 2}{2\pi n \mu a} \log \cosh 2\pi n R \right] \tag{110}$$

In many cases the term containing $\log \cosh 2\pi n R$ is small relative to the other two; a good first approximation, obtained by neglecting it, is*

$$d_1 = a \left[1 + \frac{1}{\mu} \frac{a + b}{a} \right] \tag{111}$$

The grid-cathode spacing a is evidently the most important single constituent of this expression for d_1, for b is usually less than a, and μ is rarely less and often much greater than 4. So the actual value of F_0, being inversely proportional to d_1, is very greatly dependent on the grid-cathode spacing a.

The *lack* of dependence of μ on a indicates that the *relative* effectiveness of E_g and E_p in influencing the value of F_0 is quite independent of cathode placement.

20. Parallel Plane Grid and Plate Structure with Filamentary Cathode. Many commercial tubes have plate and grid structures similar to those of the parallel-plane triodes that have been analyzed, but have cathodes consisting of filament wires or ribbons which make an angle of nearly 90° with the grid wires, and lie between two parallel grid planes. This type of construction is illustrated in Figs. 2 and 24. Since μ is not dependent either on the local geometry or placement of the cathode, the equivalent voltage $E_g + \dfrac{E_p}{\mu}$ for such devices is calculated from actual dimensions exactly as for a parallel-plane triode. However, the relation between the off-cathode gradient and the equivalent voltage is

---

* R. W. King's expression[2] for the quantity here called d_1 corresponds to the following form of Equation (111):

$$d_1 = \frac{1}{\mu} [b + a(\mu + 1)]$$

very greatly dependent on the placement and local geometry of the cathode. For example, it is apparent that F_0 must be considerably greater for the arrangement of parts shown in Fig. 23 than in Fig. 22, although μ is the same.

The electric field analysis that leads to a determination of the dependence of F_0 on the equivalent voltage in tubes having a grid-like array of filament wires, as in Fig. 24, can be satisfactorily carried out by using an x, y transformation for the grid and plate region and an x, z transformation for the cathode region. The results of these separate treatments must then be merged.

Merging can be satifactorily accomplished by an adaptation of the three-dimensional Laplace relation, Equation (12), to a set of two-dimensional x, y boundaries at grid and plate, and of two-dimensional y, z boundaries at cathode wires. The transformed representation of the two boundary structures is essential to this method, but the Z-plane point of view must be abandoned early in the merging process.

In estimating the merits of these various analyses it should be borne in mind that the chief utility of all of this field analysis lies in the improved understanding it gives of internal field relations, rather than in direct aids to the design of specific apparatus.

21. Conformal Transformation of a Cylindrical Triode. Triodes in which the cathode and plate are concentric cylinders are often used.

While it is difficult to analyze mathematically the electric field that results from such an arrangement if the grid wires are helically wound, it is easy to do so if the grid wires are straight and parallel to the axis, being spaced at equal intervals around a cylindrical surface concentric with cathode and plate. Common sense suggests, and the suggestion has been verified by experiment,[1, 8] that these two kinds of grids have identical effects on the electric field at the cathode, providing the screening fraction, size of grid wire, and spacings relative to the other electrodes are identical.

Figure 25 represents a cross-section, normal to the axis, of a cylindrical triode in which the grid wires are parallel to the axis. r_c, r_g, and r_p are the radii of cathode, grid, and plate respectively, and R of the grid wires.
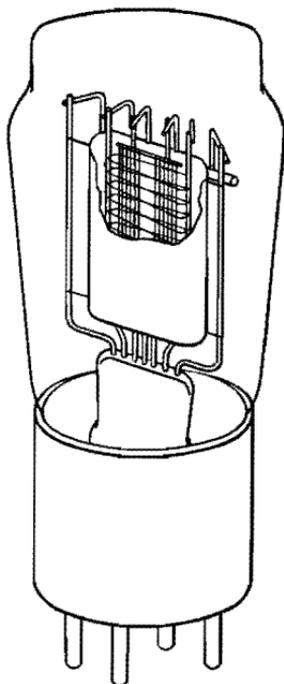
There are n grid wires per centimeter of grid-circle circumference, so that the total number of grid wires is $2\pi r_g n$. The screening fraction is $2nR$ as before; it is the fraction of the grid cylinder that is blockaded by grid wires. The field must be conformally transformed into that of Fig. 15 in such a way that the entire circumference of the Z-plane grid circle corresponds, as before, to a section of the grid arrangement on the actual or W plane that includes just one grid wire, located at the middle of the section transformed. The total angle of the actual grid circle is of course 2π radians, and it contains $2\pi n r_g$ grid wires, so that the angle corresponding to one grid-wire section is $1/n r_g$ radians. This is the value
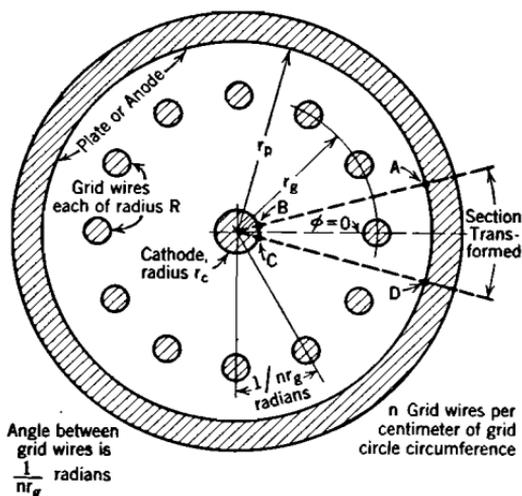
Fig. 25. Cylindrical triode with grid wires parallel to axis. Region $ABCD$ transforms to $A'B'C'D'$ of Fig. 15.

that the W-plane angle ϕ must have when the Z-plane angle ϕ' is 2π radians; so that in general:

$$\phi = \frac{1}{2\pi n r_g}\phi' \qquad (112)$$

The most satisfactory transformation equation is[1]

$$\log\left(\frac{x + jy}{r_g}\right) = \frac{1}{2\pi n r_g}\log(x' + jy') \qquad (113)$$

Equation (113) can also be written

$$\log\left(\frac{r}{r_g}\,\epsilon^{j\phi}\right) = \frac{1}{2\pi n r_g}\log r'\epsilon^{j\phi'} \qquad (114)$$

This expands into

$$\log\frac{r}{r_g} + j\phi = \frac{1}{2\pi n r_g}\log r' + \frac{j\phi'}{2\pi n r_g} \qquad (115)$$

Equation (115) is in proper form to provide the desired relation between the angles, as specified by Equation (112). The r_g factor that appears in the denominator of the *left-hand side* logarithms of Equations (113), (114), and (115) is introduced for convenience; it makes the grid-

cylinder radius the unit of measurement for the true tube dimensions, and leads to the equation

$$\log r' = 2\pi n r_g \log \frac{r}{r_g} \tag{116}$$

for radius conversion between the two figures.

The meanings given by this transformation to the Z-plane dimensions of Fig. 15 are as follows:

$$\text{Cathode radius:} \quad \log s_c = 2\pi n r_g \log \frac{r_c}{r_g} \tag{117}$$

$$\text{Plate radius:} \quad \log s_p = 2\pi n r_g \log \frac{r_p}{r_g} \tag{118}$$

Inner and outer extremes of the transformed grid wire occur at s_1 and s_2, where

$$\log s_1 = 2\pi n r_g \log \left(1 - \frac{R}{r_g}\right) \tag{119}$$

$$\log s_2 = 2\pi n r_g \log \left(1 + \frac{R}{r_g}\right) \tag{120}$$

In all ordinary cases $R \ll r_g$, so that $\frac{R}{r_g} \ll 1$.

A general series expansion for $\log(1 + z)$, where z is any variable, is as follows:[J 116, K 91]

$$\log(1 + z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 \tag{121}$$

If $z \ll 1$ this reduces simply to

$$\log(1 + z) = z \tag{122}$$

By the use of Equation (122), Equations (119) and (120) simplify to:

$$\log s_1 = -2\pi nR, \quad \text{or,} \quad s_1 = \epsilon^{-2\pi nR} \tag{123}$$
$$\log s_2 = +2\pi nR, \quad \text{or,} \quad s_2 = \epsilon^{+2\pi nR} \tag{124}$$

These are the same as Equations (49) and (50), so that, for the usual cases in which $R/r_g \ll 1$

$$s_g = \cosh 2\pi nR \tag{125}$$
$$R' = \sinh 2\pi nR \tag{126}$$

just as in the parallel-plane triode transformation.

These various relations permit adaptation to a cylindrical triode of the potential distribution equations, the solutions for cathode and plate charges, and the expressions for amplification factor and equivalent diode spacing that have been derived for parallel plane electrodes.

Figure 26 illustrates the potential variation along paths through and between grid wires of such a cylindrical triode.

There are $2\pi n r_g$ grid wires altogether, and each grid-wire section has a cathode charge $-\tau_c$ so that the true cathode charge, τ_C statcoulombs per centimeter of tube axis, has the value

$$\tau_C = 2\pi n r_g(-\tau_c) \tag{127}$$

and is negative if $-\tau_c$ is negative, as is usually the case. As in the case of a parallel-plane triode, the current flow when the cathode is
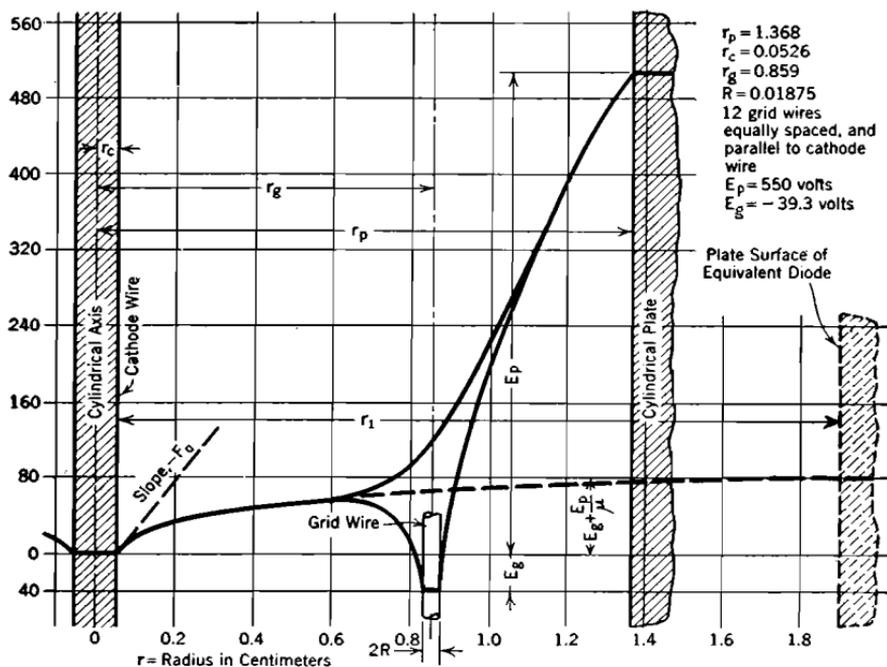
FIG. 26. Space-charge-free potential distribution in cylindrical triode like that of Fig. 25.

heated is dependent primarily on the off-cathode space-charge-free gradient, which will be called $-F_0$ as heretofore. In the near neighborhood of the cathode the true field is like that due to a line charge of strength τ_C. In accordance with Equation (53) the field strength just off the cathode surface must be

$$F_0 = \frac{2\tau_C}{r_c} \tag{128 esu}$$

or

$$F_0 = 4\pi n \frac{r_g}{r_c}(-\tau_c) \tag{129 esu}$$

The value of $-\tau_c$ is in no way changed from that given by Equation (81), which can be written

$$-\tau_c = \frac{E_g + \dfrac{E_p}{\mu}}{2 \log s_c R' + \dfrac{2}{\mu} \log s_c s_g} \qquad (130 \text{ esu})$$

μ has exactly the same significance as heretofore; that is,

$$\mu = \frac{\log \dfrac{s_p}{s_g}}{\log \dfrac{s_g}{R'}} \qquad (131)$$

The denominator of Equation (130) can be rearranged in just the way d_1 was in Equations (106–108), giving for $-\tau_c$ the expression

$$-\tau_c = \frac{E_g + \dfrac{E_p}{\mu}}{2 \log s_c s_g + \dfrac{2}{\mu} \log \dfrac{s_c}{s_p} s_g{}^2} \qquad (132 \text{ esu})$$

Using Equation (129),

$$F_0 = -\frac{E_g + \dfrac{E_p}{\mu}}{\dfrac{r_c}{2\pi n r_g}\left(\log \dfrac{1}{s_c s_g} + \dfrac{1}{\mu} \log \dfrac{s_p}{s_c s_g{}^2}\right)} \qquad (133)$$

Equation (133) should be compared with Equation (109). It is evident that the *equivalent voltage* to which F_0 is proportional is of the same form as in a parallel plane triode, but that the denominator in the expression for F_0 is changed.

22. Amplification Factor and Spacing of Equivalent Diode for a Cylindrical Triode. If the transforming relations of the preceding section are entered in Equation (131), the following expressions for the amplification factor of a cylindrical triode are obtained:[1, 8, E 178]

$$\mu = \frac{2\pi n r_g \log \dfrac{r_p}{r_g} - \log \cosh 2\pi n R}{\log \coth 2\pi n R} \qquad (134)$$

or, for very small values of $2nR$,

$$\mu = \frac{2\pi n r_g \log \dfrac{r_p}{r_g}}{\log \dfrac{1}{2\pi n R}} \qquad (135)$$

The equivalent space-charge-free cylindrical diode has the same cathode and cathode charge as the true cylindrical triode. It must have a plate radius r_1 such that a voltage $E_g + \dfrac{E_p}{\mu}$ between its two electrodes produces the value of cathode charge (and of $-F_0$) that exists in the true triode with grid and plate at E_g and E_p volts. Therefore in the equivalent diode, of plate radius r_1, cathode radius r_c,

$$E_g + \frac{E_p}{\mu} = -2\tau_C \log \frac{r_1}{r_c} \qquad (136 \text{ esu})$$

where τ_C is, as in the preceding section, the charge per centimeter length of true cathode. Equation (136) can be written

$$2\tau_C = -\frac{E_g + \dfrac{E_p}{\mu}}{\log \dfrac{r_1}{r_c}} \qquad (137 \text{ esu})$$

Using this in Equation (128),

$$-F_0 = \frac{E_g + \dfrac{E_p}{\mu}}{r_c \log \dfrac{r_1}{r_c}} \qquad (138)$$

for the equivalent space-charge-free diode, having plate radius r_1.

If the transforming relations are introduced into Equation (133), it appears that for the real triode

$$-F_0 = \frac{E_g + \dfrac{E_p}{\mu}}{r_c \left[\log \dfrac{r_g}{r} + \dfrac{1}{\mu} \log \dfrac{r_p}{r_c} - \dfrac{1}{2\pi n r_g} \left(1 + \dfrac{2}{\mu} \right) \log \cosh 2\pi n R \right]} \qquad (139)$$

From a comparison of Equations (138) and (139) it is seen that r_1 must be such that

$$\log \frac{r_1}{r_c} = \log \frac{r_g}{r_c} + \frac{1}{\mu} \log \frac{r_p}{r_c} - \frac{1}{2\pi n r_g} \left(1 + \frac{2}{\mu} \right) \log \cosh 2\pi n R \qquad (140)$$

In many cases the term containing $\log \cosh 2\pi n R$ may be neglected without introducing appreciable error, then, approximately,

$$\log \frac{r_1}{r_c} = \log \frac{r_g}{r_c} \left[1 + \frac{1}{\mu} \frac{\log \dfrac{r_p}{r_c}}{\log \dfrac{r_g}{r_c}} \right] \qquad (141)$$

This approximate specification of the ratio r_1/r_c of plate radius to cathode radius in the space-charge-free equivalent diode should be compared with Equation (111).

23. Limitations to the Validity of Triode Geometrical Relations. The derivations of potential distribution equations, charge densities, and tube properties that have been given in the preceding sections have all been based on the assumption that the array of three charges, $-\tau_c$ at cathode center, $+\tau_c$ at grid kernel, and $-\tau_p$ at grid center, make the Z-plane cathode, grid, and plate figures true equipotentials. Actually it makes them only approximate equipotentials; the closer the approximation the more satisfactory are the results of the method.

Ordinarily the approximation is close enough so that uncertainties due to entirely unrelated circumstances are of greater consequence than those inherent in the method itself. End effects at top and bottom of the electrode structure may distort the field, also, there is an element of uncertainty in the relation, derived later in Chapter V, between F_0 and the tube current; parallel-plane grid and plate structures usually extend along both sides of the cathode, and must therefore bend around at each end of the cathode, with consequent complication of the field properties. Also, most tubes with parallel-plane grid and plate arrangements have filamentary cathodes which are more nearly cylindrical than infinite-plane surfaces, and must be treated by a concentric geometry method in any close-up analysis of current flow; yet farther out toward the grid the geometry is not concentric.

Although these various incidental aspects of electrode arrangement and use are almost certain to introduce more serious discrepancies with calculated predictions than arise from failure of the Z-plane line charge system to match the true equipotentials perfectly, it is nevertheless useful to review the limitations inherent in the analytical method itself, in order to avoid absurdities in application. It is convenient to state these inherent limitations as follows:

(1) The Z-plane grid figure of Fig. 15 is treated in the analysis as though perfectly circular; actually it is very nearly circular as long as the screening fraction $2nR$, which satisfies the relation

$$2nR = \frac{\text{grid projected area}}{\text{total grid-plane area}} \qquad (142)$$

does not exceed $\frac{1}{8}$. Fair results are probably obtainable as long as $2nR$ does not exceed $\frac{1}{5}$.

(2) The analysis assumes that the plate circle is remote from the grid location, that is, s_p is presumed to be much greater than s_g. As s_g is rarely much greater than unity (when $2nR = \frac{1}{4}$ it is only about 1.3), the requirement that $s_p \gg s_g$ is not greatly different from a requirement that $s_p \gg 1$. Sometimes the plate is close

enough to the grid to make this assumption untenable; it may still be possible to use the general method by more extensive imaging.

(3) The analysis assumes that the cathode is far enough from the grid so that the Z-plane cathode circle is not much more than a point, on a scale that represents the grid circle and its location satisfactorily. The requirement that $s_c \ll s_g$ is not greatly different from a requirement that $s_c \ll 1$, since s_g is never much more than unity. Again more elaborate imaging may be helpful if in a particular case this assumption is untenable.

Practically all ordinary commercial triodes whose geometry is such as to permit any attempt to apply the simple analysis as here presented have dimensions such as to make the results reasonably satisfactory. Even when the potential lines as plotted do not show as near ideal behavior at the boundaries as might be desired, the average results, relating to the amplification factor and average value of the space-charge-free cathode gradient, may still be satisfactory.

24. Mapping the Fields. Potential variations in electrostatic devices can, like the topography of a hilly country, be studied either by means of profiles or of contours. The space-potential lines of Figs. 4, 20, and 26 represent the simplest type of profile treatment, in which only the two extreme potential profiles are shown. Such potential distribution diagrams are, in electronic work, ordinarily more useful and more easily prepared than are field maps of flux lines and equipotentials, which correspond to topographical contour maps.

It is nevertheless interesting and occasionally useful to know how to map the space-charge-free fields. In some cases field maps are obtainable only, or most easily, by the freehand field sketching method that has been found valuable in electrical design problems.[4, C] With irregular boundaries, such as that of the transformed grid when the screening fraction is more than about $\frac{1}{3}$, such procedure may be a necessary preliminary to a knowledge of the potential variation. On the other hand, the best method of obtaining some field maps is a graphical transfer of values from a mathematically determined potential distribution diagram; Fig. 22 was prepared in this way. Triode field maps have also been obtained by various experimental methods.[E 174]

Whenever the Z-plane field can be satisfactorily mapped, either by field sketching or by mathematical means, the actual field can easily be obtained, for graphical transformation of the field intersections from the Z-plane to the W-plane is very easily accomplished. Figs. 18 and 21 illustrate two situations in which the Z-plane field is of a familiar type. Fig. 21 shows the shapes of the equipotentials and flux lines at cut-off, that is, with zero charge on the cathode. Except in the immediate neighborhood of the cathode circle it is a family of concentric circles intersected by a family of radial lines. Fig. 18 represents the field when

the plate charge is zero. The equipotentials and flux lines in Fig. 18 are families of nonconcentric circles except in the neighborhood of the plate.[4][68] The field maps in Figs. 18 and 21 are interchanged if $-x$ is used in place of x, and vice versa, in the original transformation, for then the plate becomes a small circle around the origin and the cathode a remote outer circle.

PROBLEMS

Chapter II

1. A " grid " of parallel wires, each 1 cm in diameter, is supported 10 meters above the ground; the wires are 2 meters apart.

(a) Sketch and dimension the corresponding Z-plane figure, treating the ground as the cathode of Fig. 14; there is of course no " plate " and no plate charge.

(b) Sketch and dimension similarly the Z-plane figure obtained by treating the ground as the plate of Fig. 14; there is in this case no " cathode," and no cathode charge.

(c) The wires are 30,000 volts above ground potential. Find the charge on each, first in statcoulombs per centimeter length, then in coulombs per centimeter length.

(d) Find the potential difference between ground and a point midway between any two adjacent grid wires, using the Z-plane figure sketched in (a).

(e) Repeat, using the Z-plane figure sketched in (b).

2. A certain Z-plane representation consists of (1) a plate, for which $s_p = 200$, (2) a grid wire for which $s_g = 1.071$, $R' = 0.389$, (3) a very small cathode circle with its center at $\phi' = \pi$, $r' = 1.002$ (approximately 1.00), and having a radius $R'' = 0.0628$. Note that this cathode $is\ not\ located\ at\ the\ Z\text{-}plane\ origin$. In fact, there is no electrode at the Z-plane origin.

(a) Sketch the corresponding W-plane figure, and indicate on the sketch the dimensions of and between the various electrodes. Use $n = 10$ grid wires per cm.

(b) A point T lies, $in\ the\ W\text{-}plane\ diagram$, midway between the plate and the plane of the grid wires and, also, one-quarter of the distance from one grid wire to the next. State the values of x and y for this point, also the values of r' and ϕ', and show clearly its location on a Z-plane sketch.

(c) In this Z-plane figure the cathode circle is small enough so that three line charges, $-\tau_c$ inside the cathode, and $+\tau_c$, $-\tau_p$ inside the grid, can be used to establish the boundary equipotentials in the usual way. Show that the radius s_k from the Z-plane origin to the grid kernel is now 1 instead of the usual value $1/s_g$.

(d) Derive expressions for the potential differences between plate and cathode and between grid and cathode in terms of τ_c, τ_p, and the Z-plane dimensions. Use $s_k = 1$, not $1/s_g$.

(e) Derive an expression for the amplification factor μ, following the general method outlined in Section 18.

(f) If $E_p = 500$ volts, and $E_g = -40$ volts, what is the space-charge-free off-cathode gradient?

3. A parallel plane triode like that illustrated in Fig. 14 has the following dimensions:

$$a = 0.16 \text{ cm} \qquad 2nR = 0.12$$
$$b = 0.08 \text{ cm} \qquad n = 10$$

Plate-to-cathode potential difference E_p is 250 volts. Select a grid-to-cathode

potential difference E_g from among the following values: -17.5, -35, -52.5, -70. Then, using the selected value,

(a) Determine the values of $-\tau_c$ and τ_p, as used in Sections 12–18.

(b) Using the results obtained in (a) calculate points for and plot curves of space-charge-free potential distribution similar to those in Figs. 4 and 20. It is suggested that the following values of x be used: -0.16, -0.12, -0.08, -0.04, -0.02, $+0.02$, $+0.04$, $+0.08$.

(c) Calculate μ from the dimensions, and find what value of the equivalent voltage corresponds to the E_p and E_g being used.

(d) Determine the value of d_1, the space-charge-free equivalent diode spacing, by a graphical construction similar to that indicated in Fig. 20, and check by numerical calculation, following Section 19.

(e) Find the value of $-F_0$, the space-charge-free off-cathode gradient; also find the surface charge densities on both cathode and plate.

(f) What is the grid voltage at cut-off, if $E_p = 250$?

4. (a) For a triode like that of Figs. 14 and 15, write an r', ϕ' equation for the potential relative to the cathode at any point between the electrodes, that is, for any pair of values of r' and ϕ' (see Section 13). Show how your equation reduces to Equations (70) and (71) for specific values of ϕ'.

(b) Convert the general equation obtained in (a) to an x, y equation for the potential distribution curve corresponding to $y = \frac{1}{4}n$, and plot such a curve, using the dimensions and voltages of Problem 3.

(c) Convert the general equation obtained in (a) into an x, y equation for the potential distribution in the plane of the grid wires, that is, $x = 0$. Plot such a curve, using the dimensions and voltages of Problem 3.

5. (a) Using the principles outlined in Section 21, determine values of r_c, r_g, and r_p for a cylindrical triode (like that of Fig. 25) that corresponds to the parallel-plane triode of Problem 3. There are to be 18 equally spaced parallel-to-the-axis grid wires. The screening fraction, true grid-wire radius, and values of s_p and s_c, are to be the same as in Problem 3.

(b) Draw, for this cylindrical triode, the potential distribution curves that correspond to those prepared in Problem 3, using the values of plate and grid voltages employed there. To do this, it is only necessary to calculate values of true radius r that correspond to the values of x suggested in (b), Problem 3. The voltage values will be the same as those used before, because the Z-plane figure is the same for the cylindrical triode as for the parallel-plane triode.

(c) Determine the amplification factor, the value of r_1, the charge per centimeter length of true cathode, the charge per square centimeter of plate surface, and $-F_0$, the space-charge-free off-cathode potential gradient, for this cylindrical triode.

6. (a) A cylindrical triode is similar to that of Problem 5, except that the grid consists of rings, concentric with the cathode, and having the same radius as the grid-wire cylinder of Problem 5. The total length of all grid wires per centimeter of axial length of tube structure is the same as in Problem 5. Find the value of n, grid wires per centimeter, and compare with that in Problem 5. If the grid-wire radius is the same as before, what is the screening fraction, and how does it compare with the former value?

(b) If the grid is a helix with a pitch of $1/n$, instead of rings, and the grid wire the same size as before, by what per cent are the total length of grid wire and screening fraction increased over their values in (a)?

7. A cylindrical triode consists of a cathode of circular cross-section, radius 0.008

cm. The plate radius is 0.20 cm. The screening fraction of the grid is $\frac{1}{8}$, and the radius to the grid cylinder is 0.08 cm. There are 20 grid wires per cm.

(a) Make a sketch of the transformed (Z-plane) figure, marking on it numerical values for s_c, s_p, s_g, R'.

(b) Determine the amplification factor.

(c) Determine the radius r_1 to the plate of the equivalent space-charge-free diode.

8. Calculate μ and d_1 for the geometry of Fig. 20, and check against the construction shown in that figure.

9. Calculate μ and r_1 for the geometry of Fig. 26, and check against the construction used in that figure.

10. The cathode of a General Electric FP-110 triode consists of a W-shaped filament located in a plane midway between two parallel-plane electrodes. One of these plane electrodes is the plate, while the other performs the functions of the grid of a triode of the usual type. Thus Fig. 14 is a fairly satisfactory representation of the true geometry of the FP-110 tube, except that the left plate is a " plane-surfaced grid " rather than a cathode, and n, R, are respectively the spacing and the filament radius of the grid-like cathode.

Suppose that in a tube of this general type the plate and plane-surfaced grid are 0.2 cm apart, and the grid-like cathode is located midway between them. Radius of cathode wires is .004 cm and there are 3 cathode wires per centimeter.

(a) Cathode and plane-surfaced grid are at a common potential, plate 80 volts positive relative to them. Sketch carefully the space-charge-free potential distribution curve for a path, from the plane-surfaced grid to the plate, passing through a cathode wire, also for a similar path passing midway between cathode wires. No mathematics should be used, but the principles illustrated in the development and discussions of Figs. 4, 20, 22, and 26 should be borne in mind.

(b) Repeat with plate 80 volts above cathode potential and plane-surfaced grid 150 volts below cathode potential.

(c) State whether or not plate current is to be expected in either the (a) or (b) condition, or in both, explaining the reasons for your statements by reference to the diagrams.

(d) Plate at +80 volts relative to cathode. Estimate the approximate potential of the plane-surfaced grid at cut-off, explaining your calculations by reference to a diagram similar to those asked for in (a) and (b).

11. Sketch and dimension the Z-plane figure for the tube of Problem 10. Derive a mathematical expression for the amplification factor of such a tube, using the general method outlined in Section 18, but modifying it to correspond to the changed electrode functions. Calculate the amplification factor for the tube of Problem 10. What dimensions are of chief importance in determining the value of μ?

12. For a cylindrical triode, what value of the ratio r_g/r_p will make μ a maximum, if r_c, r_p, n and R remain fixed as r_g is varied?

CHAPTER III

ELECTRON BALLISTICS*

25. Acceleration Due to an Electric Field. Under certain conditions electrons may escape from the confinement of metal boundaries. If electric or magnetic fields exist in the region into which they escape, and if there is a vacuum sufficiently high so that collisions with gas particles are extremely rare, these fields will control the subsequent movements of the electrons. This chapter analyzes various types of electronic motion that take place when the number of electrons escaping is so small that their charges and movements have either negligible or very small effects on the electric and magnetic fields.

Imagine a few electrons to be released from the more negative of two parallel electrodes, such as those of Fig. 5a, between which there is a considerable difference of potential. Each electron experiences a force f whose magnitude in dynes is the product of the charge it carries $(-e)$ by the electric field intensity F, thus

$$f = -eF = +e\frac{dE}{dx} \qquad \text{(143 esu)}$$

or in practical electrical units:

$$f = -eF10^7 = +e\frac{dE}{dx}10^7 \qquad \text{(144 p)}$$

The practical unit of F is the volt per centimeter. The numerical values of the charge on an electron, and of the electron's mass m_e, are as follows:[X 713]

$$e = 4.767 \times 10^{-10} \text{ statcoulomb per electron} \qquad \text{(145 esu)}$$

$$e = 1.590 \times 10^{-19} \text{ coulomb per electron} \qquad \text{(146 p)}$$

$$e = 1.590 \times 10^{-20} \text{ abcoulomb per electron} \qquad \text{(147 emu)}$$

$$m_e = 9.038 \times 10^{-28} \text{ gram, or approximately } 9.00 \times 10^{-28} \text{ gram} \quad \text{(148)}$$

The ratio m_g/m_e of the mass of a gas particle or ion to the mass of an electron is determined by multiplying the particle's atomic or molecular weight by 1824, which is the ratio of the mass of a particle of unit atomic weight to the mass of an electron. Mathematically, the mass ratio is

$$m_g/m_e = 1824 \times \text{atomic weight} \qquad \text{(149)}$$

* The title used for this chapter was suggested by Dr. G. S. Timoshenko.

No particles having unit atomic weight exist, the nearest to it being a hydrogen atom, with an atomic weight of 1.008, mass ratio 1847.

Just as a ball on a hillside is accelerated toward the bottom of the hill at a rate dependent upon its mass and the slope (i.e., elevation gradient) of the hill, so such an electron is accelerated toward the region of high potential with a rate of change of velocity, dv/dt, that is dependent on its mass and the potential gradient in which it lies; that is,

$$\frac{dv}{dt} = -\frac{eF}{m_e}\left(\frac{\text{cm}}{\text{sec}^2}\right) \qquad \text{(150 esu)}$$

Equation (150) is a rearrangement of the force equation

$$f = -eF = +m_e\frac{dv}{dt} \text{ (dynes)} \qquad \text{(151 esu)}$$

If the electric field is uniform, the acceleration is constant. This corresponds to a ball rolling down a hill of constant slope; the time to travel a given distance, and the average and final velocities, may be calculated by the familiar laws governing uniformly accelerated motion. Behavior in a nonuniform field may be treated by integration, as in the case of a ball on a curved slope.

26. Velocity and Potential; the Electron Volt. The kinetic energy acquired by a ball on a hill is independent of local variations in the slope of the hill, being related rather to the total vertical distance traveled, that is, to the decrease in potential energy. A "falling" electron is, like such a ball, continually converting potential energy of position into kinetic energy, and the gain of one kind of energy must equal the loss of the other.

Electric potential difference between two points is by definition the work expended in moving a unit charge from one point to the other, hence the change in ergs of potential energy corresponding to the movement of an electron through a field is the product of the potential difference E through which it moves, in statvolts, by the electron's charge e, in statcoulombs. Equating this to the kinetic energy gained

$$Ee = \tfrac{1}{2}m_e v^2 \text{ (ergs)} \qquad \text{(152 esu)}$$

Both sides of this equation are in ergs. To make a quick conversion into practical units it is only necessary to recall that the product of volts by coulombs gives energy, but in watt seconds, not ergs. A watt second is 10^7, ergs; therefore, in practical electrical units:

$$Ee10^7 = \tfrac{1}{2}m_e v^2 \text{ (ergs)} \qquad \text{(153 p)}$$

Equations (152) and (153) express mathematically the extremely useful general principle that an electron which has "fallen" through a

certain potential difference E has perfectly definite values of kinetic energy and velocity, regardless of the manner of variation of potential between the starting and end points, or of the direction of the velocity.

The energy possessed by an electron that has passed through a potential difference of one volt is a convenient *unit of energy*, called the *electron volt*, sometimes abbreviated to "*volt.*" The use of this unit has been extended to very many circumstances in which the energy measured is not at all the result of movement of a charged particle in an electric field. For example, the average kinetic energy of thermal motion possessed by molecules of air at 40° Centigrade is conveniently specified as 0.0675 volt. This means that their average kinetic energy is the same as that which would be gained by an electron or ion during acceleration through a potential difference of 0.0675 volt. It does not mean at all that an electron or an electric field has any part in creating an air molecule's energies. The electron volt is an extremely convenient unit for measuring all kinds of energies of electronic magnitude.

If the energy in "volts" of an electron is known, its velocity in centimeters per second can be calculated by solving Equation (153) for the velocity, giving

$$v = \sqrt{\frac{2Ee10^7}{m_e}} = 5.93 \cdot 10^7 \sqrt{E} \qquad (154 \text{ p})$$

Anyone engaged in electronic work or study is greatly helped by having permanently available in a not too remote corner of his mind the derivation and form of Equation (154), if for no other reason than that of having familiarity with the order of magnitude of electronic velocities.

It is frequently desirable to be able to determine quickly the velocity of an *atom* or *molecule* whose energy in electron volts is known. The initial form of Equation (154) indicates that the velocity is inversely proportional to the square root of the mass, so that for these heavier particles:

$$v = \frac{5.93 \cdot 10^7 \sqrt{E}}{\sqrt{m_g/m_e}} \qquad (155 \text{ p})$$

Here m_g is the mass, and m_g/m_e the *mass ratio*, of the particle concerned; see Equation 149.

27. Directed Energies; Velocity Measurable in Square Root Volts. The direction as well as the magnitude of the velocity possessed by an electron is of importance. Suppose, for example, that an "electron gun" (devices so described are used in cathode-ray apparatus) is employed to shoot a beam of electrons at an angle into the uniform field between

the plates of Fig. 27. Each entering electron initially possesses an
x-directed velocity component u, and a y-directed component v. The
former remains constant, the latter is modified by uniform downward
acceleration due to the electric field. The trajectory is parabolic, as in
the exactly analogous case of a ball thrown upward at an angle against
the force of gravity.

Only the y-directed velocity is effective in aiding the electrons to
reach the upper plate. For
example, if E_s is 200 volts,
the beam can reach the upper
plate only if initially $v \geqq 5.93 \cdot$
$10^7 \times \sqrt{200}$, that is, the ini-
tial vertical velocity must
exceed $\sqrt{200} = 14.14$ *square
root volts*. The horizontal
velocity may be anything at
all without affecting the ver-

Fig. 27. Electron trajectory in a uniform
electric field.

tical travel. If the initial vertical energy is exactly 200 electron volts the
vertex of the parabola occurs at the upper electrode, and the contact
there is a grazing one.

The two initial velocity components may conveniently be described
as $\sqrt{E_x}$ and $\sqrt{E_y}$ square root volts. The total velocity is $\sqrt{u^2 + v^2}$
or, in square root volts

$$\text{total velocity} = \sqrt{\overline{\sqrt{E_x}}^2 + \overline{\sqrt{E_y}}^2} = \sqrt{E_x + E_y} \qquad (156)$$

If for example $\sqrt{E_x}$ is $\sqrt{400} = 20$ and $\sqrt{E_y}$ is $\sqrt{300} = 17.32$, the total
initial velocity in centimeters per second is $5.93 \cdot 10^7 \times \sqrt{400 + 300} =$
$5.93 \cdot 10^7 \times \sqrt{700}$. The initial angle with the horizontal is arc tan
$\sqrt{\frac{300}{400}} = $ arc tan 0.866. The electron gun is said to be delivering
" seven-hundred-volt electrons " and the electron beam is called a
" seven-hundred-volt beam." The gun must contain provisions for
accelerating electrons through a potential difference of 700 volts.

The time required to reach the upper plate depends entirely upon the
initial upward velocity, and can be obtained by integration after writing
Equation (154) in the form

$$dt = \frac{1}{5.93 \cdot 10^7} \frac{1}{\sqrt{E_y}} dy \qquad (157 \text{ p})$$

Here $\sqrt{E_y}$ is a variable that describes the vertically upward velocity
at any point after release from the muzzle of the gun. In this example
E_y is a linear function of y, because the field is uniform; if the initial
value of E_y is 300, the vertical velocity at any ordinate y is $+ \sqrt{300 - yF}$

square root volts, and the time of flight to the top plate is

$$t = \frac{1}{5.93 \times 10^7} \int_{y=0}^{y=s} \frac{dy}{\sqrt{300 - yF}} \qquad \text{(158 p)}$$

This integral is easily evaluated. The distance traveled in the x direction meanwhile is of course ut, where $u = 5.93 \cdot 10^7 \sqrt{E_x}$.

28. Electron Deflection in Passing Through Grids.

Fig. 28a illustrates the space-charge-free potential distribution in a triode with plate and cathode at a common potential, and the grid positive. Fig. 28b is a cross-section showing electrode arrangement. The general nature of the potential distribution in the neighborhood of the grid is not materially altered by the passage of the small electron current that is presumed to exist owing to thermionic emission at the cathode. Most of the electrons that leave the cathode cannot reach the plate, for as indicated by the trajectories shown in Fig. 28b, each one experiences as it passes the grid a slight sideways deflection which converts part of the energy of horizontally directed motion that it has received from the field into vertically directed motion. It must miss the plate by just the amount of energy so converted.

For example, imagine an electron at T, traveling along the trajectory drawn through that point. Since it started from the cathode with zero initial velocity, it must of course possess at T just E_T electron volts of kinetic energy, which is enough, if properly directed, to get it to the plate. But E_y electron volts of this is associated with vertical motion. The electron

FIG. 28. Deflection of electrons in passing through a grid.

possesses, at T, $E_T - E_y$ electron volts of horizontally directed motion, but the plate is E_T volts away, so that the electron must fail to reach the plate by just E_y volts.

One of the trajectories shown is the straight-line path of electrons that pass exactly between grid wires, so experiencing no deflection; only these can reach the plate. All others follow parabolic trajectories whose vertex is slightly to the left of the plate. They are unable subsequently to return to the cathode for the same reason that prevents them from reaching the plate. They must oscillate back and forth, like a ball rolling down one side and up the other of a trough, until finally they happen to strike the grid and pass on into the grid circuit. The period of oscillation can be calculated by proper use of Equation (158).

A situation very similar to this exists near zero plate voltage in screen-grid tubes and pentodes. The slopes of plate characteristic curves for such tubes (see Chapter VI) as they rise from zero with increasing but still small plate voltages, are dependent partly on the degree of dispersion experienced by the electrons in traversing the grid region, for each increment of plate voltage makes it easier for deflected electrons to reach the plate.

If the plate voltage is considerably positive, all electrons may be expected to reach and enter it, but it is still true as before that any vertical velocity component acquired by an electron in passing the grid does not subsequently change in value. If the plate is at a much higher potential than the grid, the vertical velocity acquired may become almost completely masked by the large increase in horizontal velocity, but is not destroyed. If the plate potential is just enough below that of the grid to make the charge on the plate and the field outside of the plate zero, the direction of flight will not change after leaving the grid region. Manipulation of grid and plate voltages permits focusing the electrons into narrow bands at various distances from grid toward plate.

In studying electron flow in cylindrically arranged triodes these concepts must be somewhat modified. For example, a tangential component of motion is in general similar to y-directed motion in Fig. 28, but tangential motion can, if it persists, bring electrons to a cylindrical plate. However, in most real devices proportions are such as to make the general overall behavior somewhat similar to that described for parallel-plane geometry.

29. Force on an Electron Moving in a Magnetic Field. A moving electron is a form of electric current, and, like a conductor carrying a current, experiences in a magnetic field a force normal to the field and to the direction of movement of charge. The force f on an electron

having velocity v in a field of magnetic flux density B normal to the velocity is

$$f = Bev \text{ (dynes)} \tag{159 emu}$$

Note that e is in abcoulombs. Equation (159) is very closely related to the familiar expression

$$f = BiL \text{ (dynes)} \tag{160 emu}$$

which describes the force on a conductor L centimeters long carrying i abamperes at right angles to the magnetic field. For if the conductor contains, in each centimeter of length, N electrons that are free to travel, and if their average velocity is v, Equation (159) requires that the force on the L centimeters of conductor length shall be

$$f = Bev \cdot NL = B \cdot Nev \cdot L \tag{161 emu}$$

This is identical with Equation (160), for Nev is the current i. The identification of Nev with current is analogous to obtaining the air flow in a duct in pounds per second by multiplying the number of cubic feet of duct volume per foot of its length by the pounds of air per cubic foot and the velocity in feet per second.

In practical units Equation (159) is

$$f = \frac{Bev}{10} \tag{162 p}$$

In electrostatic units it is

$$f = \frac{Bev}{c} \tag{163 esu}$$

where c is the ratio of the numerical values of the electronic charge in the two systems of units, and is numerically the same as the velocity of light in centimeters per second, that is, $c = 3 \times 10^{10}$.

Fig. 29. Direction of magnetic force on a moving electron.

The direction of the magnetic force on an electron is obtained by the same method that is used for determining the direction of the force on a conductor in a magnetic field, bearing in mind that an upward-moving electron corresponds to a downward-flowing current. As illustrated in Fig. 29, an upward motion of the electrons in a magnetic field whose flux lines are pointed toward the reader results in a force toward the left.

30. Path Circular or Helical in a Uniform Magnetic Field; Superposition of Magnetic Motions. The force on and acceleration of an

electron in a magnetic field are at every instant at right angles to the existing velocity, therefore the total velocity remains constant. If the magnetic field is uniform the acceleration normal to the velocity is of constant magnitude. The simultaneous existence of constant linear electron velocity and uniform electron acceleration at right angles to the direction of motion requires travel in a circular path, in entire agreement with experimental electronic observations.

If the original velocity is not normal to the direction of the field, the *velocity component parallel to the field* is not affected either in direction of magnitude, while the *velocity component normal to the field* continually changes direction, remaining constant in magnitude. The resulting path is a helix combining uniform circular motion with translation parallel to the axis of rotation.

Equations for the circular motion are derived by equating centripetal force to the magnetic force. If r is the radius of the circle

$$\frac{m_e v^2}{r} = Bev \qquad\qquad (164 \text{ emu})$$

or

$$r = \frac{m_e}{e}\,\frac{v}{B} \qquad\qquad (165 \text{ emu})$$

or, using Equation (163)

$$r = \frac{m_e c}{e}\,\frac{v}{B} \qquad\qquad (166 \text{ esu})$$

Equation (154) permits expression of the radius in terms of the electron's kinetic energy, E_v electron volts; thus

$$r = \frac{9 \cdot 10^{-28} \times 5.95 \cdot 10^7}{1.591 \cdot 10^{-20}}\,\frac{\sqrt{E_v}}{B} = 3.37\,\frac{\sqrt{E_v}}{B} \qquad\qquad (167 \text{ p})$$

For example, 400-volt electrons, in a field of 100 lines per square centimeter, pursue a circular path of 0.6 cm radius; the radius is 6 cm for 40,000-volt electrons. This example illustrates the very great sensitivity of an electron's motion to even weak magnetic fields. The radius of the path for heavier particles having the same energy is larger in proportion to the *square root* of the mass ratio; this fact is frequently made use of in the determination of the masses of particles of atomic magnitude.

Angular velocity ω (radians per second) in the circular path is v/r; on substituting this in Equation (165) it appears that

$$\omega = \frac{Be}{m_e} = 1.759 \times 10^7 B \qquad\qquad (168 \text{ emu})$$

Thus the angular velocity depends only on the field strength, and not at all on the linear velocity. The time for completing one circuit around the circular path is obtainable from the angular velocity; it is:

$$t = \frac{2\pi m_e}{e}\, \frac{1}{B} = \frac{0.357 \cdot 10^{-6}}{B}\ \text{(second)} \qquad \text{(169 emu)}$$

In a field of 100 lines per sq cm, this period is about $\frac{1}{300}$ of a microsecond. The period for heavier particles is of course longer in proportion to the mass ratio.

Magnitudes for a helical path can be calculated by splitting the total velocity v into two components, one normal to the magnetic field (v_N), one parallel to it (v_B), the latter remaining constant. The angle, θ, constant in a uniform field, between the directions of total velocity and field is related to the radius of the helix as follows:

$$r = \frac{m_e v_N}{eB} = \frac{m_e}{e}\, \frac{v}{B} \sin\theta \qquad \text{(170 emu)}$$

The angular velocity is the same as in a circular path, for it is wholly independent of the electron's energy. Hence Equation (169) gives the time for one circuit around the helix *for all values of θ*. The pitch of the helix is the product of the cyclic time by the velocity component parallel to the direction of the field, and is therefore quite independent of the velocity normal to the field. It is given by

$$p = \frac{2\pi m_e}{e}\, \frac{v_B}{B} = \frac{2\pi m_e}{e}\, \frac{v}{B} \cos\theta \qquad \text{(171 emu)}$$

Equations (170) and (171) can be expressed in terms of the electronic energy by the aid of Equations (152) or (154).

An attempt to analyze helical motion by splitting the *magnetic field* into components perpendicular and parallel to the initial velocity leads to results which are not correct. Radius of curvature and distance traveled parallel to the axis of the helix are directly proportional to the respective components of the initial velocity. The circular and translational motions calculated from the velocity components can therefore be superposed on one another to determine the true helical path as long as the motion remains within the extent of the uniform field. However, radius of curvature is an inverse, not linear function of the magnetic field strength, and travel along the helix not dependent on field strength at all. Hence superposition of motions determined by resolving the field into components is not permissible.

31. Opposing Electric and Magnetic Fields; Moving Magnetic Fields. Fig. 30 illustrates a region which contains a uniform electric field and a uniform magnetic field, the two being at right angles to one

another. The magnetic flux lines are directed toward the reader, that is, in the $+z$ direction. If an electron beam is shot upward from below, the magnetic field tends to deflect it toward the left, the magnetic force on each electron being given by Equations (159) or (163). The electric field exerts an oppositely directed force, its magnitude being given by Equation (143).

If these two forces are equal, the electrons of the beam are not deflected, and travel vertically upward with constant velocity, just as though neither field were present. The conditions for no deflection are obtained by equating the two force expressions, with the following result:

$$Bv = Fc \qquad \text{(172 esu)}$$

where c is as before 3×10^{10}. In practical units

$$Bv = F10^8 \qquad \text{(173 p)}$$

In the above illustration the electrons have been thought of as moving upward with constant velocity, and the magnetic field as standing still. It is reasonable to expect that the important matter is the relative velocity between the two, and such a point of view does in fact lead to a correct prediction of the motion. If the magnetic field is made to move downward with a velocity v, related to B and F according to Equations (172) and (173), an electron released into the field with zero velocity must stand still; the stationary electric and moving magnetic fields completely neutralize one another.

Fig. 30. Straight-line electron motion with balanced uniform electric and magnetic fields.

It is more in accord with the usual presentation of electromagnetic theory to say that *the motion of the magnetic field " generates " an electric field that is equal and opposite to that already existing*, that is, of value

$$F_G = -Bv10^{-8} \qquad \text{(174 p)}$$

and that the electron stands still because the net electric field is zero. Since the electron's velocity is zero the magnetic field exerts no force; the effect of the motion of the magnetic field itself has been taken account of in terms of a " generated electric field." The " generated voltage " E_G, over the entire region between the plates, their separation being l centimeters, is of course F_Gl, that is

$$E_G = -Blv10^{-8} \qquad \text{(175 p)}$$

The form of Equation (175) is familiar from the study of electric machinery.

Thus the relative motion between magnetic field and electron that is needed to avoid movement in response to the applied electric field is exactly that needed to generate a voltage numerically equal to E in a wire stretched from one electrode to the other. The internal electric field produced by the moving magnetic field is the exact analogue of the "generated voltage," sometimes called "back voltage" or "counter electromotive force," in a direct current motor. The electron's motion is zero when the generated field is equal to the applied field, for exactly the same reason that the armature current in a motor is zero when the generated voltage is equal to the applied voltage. The voltage that appears between the ends of a wire stretched from one electrode to the other has the same polarity as the plates for the identical reason that the generated voltage of a motor produces the same polarity at machine terminals as that due to the applied voltage.

Analysis of the effect of a moving magnetic field in terms of the electric field that it generates has the distinct merit of avoiding the confusion that grows out of attempts to visualize or create experimental arrangements in which magnetic fields "move." Spinning the poles from which flux lines emerge does not necessarily spin the magnetic field. Whatever kind of motion is set up must, to have any effect on electrons, be of a type that would generate a voltage by magnetic flux-cutting in a conductor located in the region being considered.

32. Cycloidal and Trochoidal Motion in the Presence of Uniform Electric and Magnetic Fields. Only under the very special conditions described by Equations (172) or (173) is the force on an electron zero when, as between the plates of Fig. 30, it is subject to the joint action of uniform electric and magnetic fields. In general the initial velocity is not such as to make the opposing electric and magnetic forces equal and opposite, and the electron pursues some sort of curved path. The type of motion in any particular case can be predicted by recognizing that it must be a composite of (a) *the motion necessary to counteract magnetically the electric force and* (b) *circular or helical motion as produced by the magnetic field alone.*

Suppose the magnetic field of Fig. 30 to be moving vertically downward with just the velocity necessary to generate a " back " field equal and opposite to the applied field; the velocity must be related to F and B in accordance with Equations (172) and (173). If now an electron is introduced with some initial velocity it must pursue a circular path, for the motion of the magnetic field just destroys the effect of the applied electric field; only the normal effect of velocity in a magnetic field

remains.   The *relative* motion between the electron and the magnetic field is now a composite of circular and translational types.

A case of particular interest is that in which the circular motion has the same velocity as the translational, both having the balanced-force value

$$v = \frac{Fc}{B}$$ (176 esu)

also written

$$v = \frac{F}{B} 10^8$$ (177 p)

The radius of the circular motion is obtained by using this in Equation (166); it is

$$r = \frac{m_e c^2}{e} \frac{F}{B^2} = \frac{cF}{\omega B}$$ (178 esu)

or, in practical units,

$$r = 5.68 \frac{F}{B^2}$$ (179 p)

The angular velocity ω produced by the magnetic field is independent of the electron's linear velocity.   The value of ω, converted into electrostatic units from Equation (168), is given by

$$\omega = \frac{e}{m_e c} B = 1.759 \times 10^7 B$$ (180 esu)
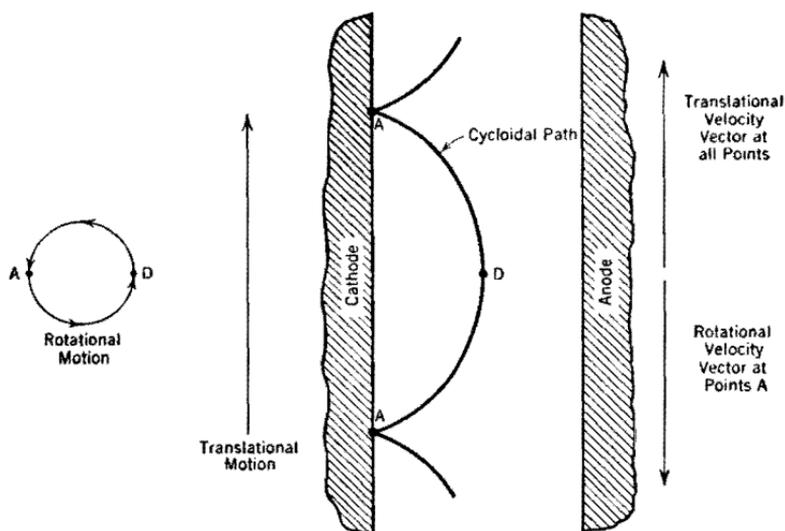
Fig. 31.   Cycloidal path of an electron in uniform electric and magnetic fields; zero initial velocity.

Now suppose that an observer rides with the magnetic field.   Not being conscious of his own downward motion, he thinks the electron is

moving upward. To him it appears to have circular motion super-posed on straight-line upward travel of equal velocity, with resultant cycloidal motion of the type illustrated in Fig. 31. The cusps at points A indicate periodic recurrence of zero *relative* velocity, owing to the two relative-velocity components being momentarily equal and opposite.

In physical problems only relative motion has real significance. Therefore the electron's movement relative to the observer and to the magnetic field must be the same whether they are moving or stationary. Thus the cycloid in Fig. 31 also illustrates the path pursued by an electron released with zero initial velocity into a *stationary* combined field. The point of release is the first of the periodically recurring points A.

Such a path as is illustrated in Fig. 31 would be traced by a point on the rim of a wheel, of radius given by Equation (179), rolling along the plane surface (cathode) from which the electron is initially released. The center of the wheel translates at just the velocity, given by Equations (176) and (177), necessary to counteract magnetically the electric force due to the applied field.*

If the electron's initial velocity is not zero, its path is that traced by the end of a spoke that either extends beyond or fails to reach the rim of the wheel, which must have the same radius and roll at the same speed as before. See Fig. 32. There are several distinct steps in the prediction of the type of motion that must result from specified values of field strength and initial velocity, as follows:

(1) The translational velocity, radius to rim of the rolling wheel, and angular velocity are all independent of the electron's initial velocity, so are determined, exactly as in the simpler zero-initial velocity case, from Equations (176) to (180).

(2) As illustrated for the start-off point A, Fig. 32, the initial magnitude and direction of the circular velocity component must be such that vector combination of it with the translational velocity gives the initial velocity whose direction and magnitude are specified in the statement of the problem.

* As pointed out at the end of the previous section, the true motion of an electron in a uniform magnetic field can be predicted by splitting the initial electron velocity into components, determining the motion for each component, then superposing the motions. When an electric field is also present, an electron's motion in a plane per-pendicular to the magnetic field should be thought of as consisting of two com-ponents, chosen so that one of them has the balanced-force direction and value specified by Equations (176) and (177). If the initial velocity is in fact zero, the second or remainder component is equal but opposite to the balanced-force component.

Three forces must then be considered: (1) that due to the balanced-force velocity component, (2) that due to the electric field, (3) that due to the remainder velocity component. Forces (1) and (2), being equal and opposite, neutralize one another so that the balanced-force component of motion continues unchanged. Force (3) results in a circular component of motion, of radius and angular velocity as in Section 30. Superposition therefore predicts translational combined with circular motion.

(3) The length L of the spoke that traces the path is determined from the general relation $\omega = v/L$, where ω is known from (1) and v is the circular velocity, whose magnitude remains constant at the initial value found as described in (2). The initial position of the tracing spoke is of course perpendicular to the initial circular-velocity vector.

(4) The far end of the initial position of the spoke identifies the position of the line of motion of the rolling wheel's center. Since the radius to the rim is known, the track can be located and the path graphically or mathematically constructed. The periodicity of the motion is in all cases given by Equation (169).
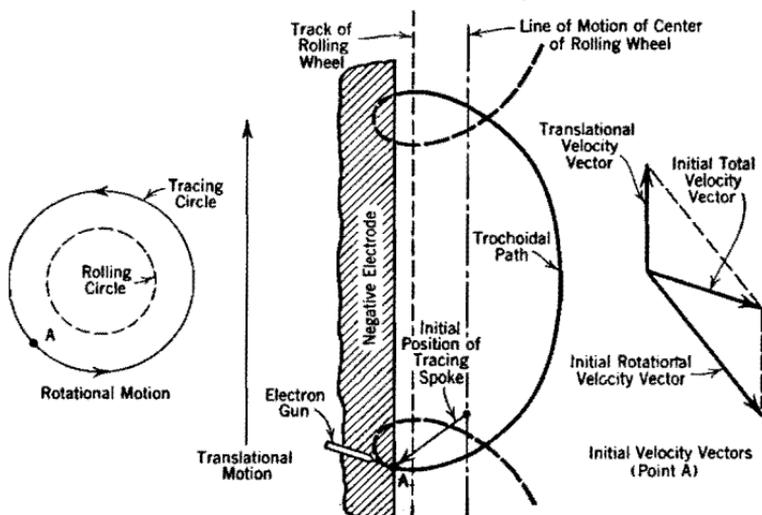
FIG. 32. Trochoidal path of an electron in uniform electric and magnetic fields, initial velocity not zero.

The most interesting property of the simple cycloidal motion of Fig. 31 is that for any given electrode spacing there is a particular value of magnetic field strength that makes the electrons just graze a plate surface located parallel to the cathode. The value of this magnetic field can be obtained by observing that the kinetic energy at the outer extreme, D in Fig. 31, is all due to vertical velocity of twice the translational value, for at this point the circular and translational motions are equal in direction as well as magnitude. The electric field strength can be expressed as E_p/s, where s is the spacing between the plates, and E_p the plate potential. The resulting relation between voltage, spacing, and field strength necessary to produce grazing contact at the plate surface is:

$$B = \frac{\sqrt{E_p}}{s}\sqrt{\frac{2m_e c^2}{e}} \qquad \text{(181 esu)}$$

If the magnetic field is weaker than this, all electrons originating at the cathode enter the plate; if stronger than this, no electrons enter the

plate. In actual operation this behavior may be somewhat modified by the presence of space charge, which has been neglected in this derivation, but the modification is slight and the general principles unchanged.

All of the relations that have been derived in this section can be obtained without having recourse to the " relative velocity " imagery by solving the following pair of differential equations of the electron's motion:

$$m_e \frac{d^2x}{dt^2} = -\frac{Be}{c} \frac{dy}{dt} + eF$$

$$m_e \frac{d^2y}{dt^2} = +\frac{Be}{c} \frac{dx}{dt}$$

(182 esu)

Here t is time, and x and y horizontal and vertical coordinates, all measured from the instant and point of electron release or ejection. The second derivatives represent accelerations and the first derivatives velocities. For the zero initial velocity case, Fig. 31, dx/dt and dy/dt are both zero when t is zero, so that the solution is very nicely handled by the operational calculus. The results of the solution are

$$x = \frac{Fc}{B} (1 - \cos \omega t)$$

$$y = \frac{Fc}{B} (\omega t - \sin \omega t)$$

(183 esu)

where ω is as defined by Equation (180).

These same equations for the electron's motion can be written directly from the knowledge that the path is a cycloid. Determination of equations for the more complicated path of Fig. 32 is much more easily accomplished by superposition of the translational and rotational travel than by solution of the differential equations of motion.

33. Motion Between Concentric Cylinders with Magnetic Field Parallel to Axis. Cylindrical electrode arrangements similar to that shown in Fig. 33a, with a magnetic field parallel to the axis, cathode and anode small and large cylinders respectively, have been used as ultrahigh-frequency *magnetron oscillators*.[9, 10, 11, 12, 13, 14] The oscillating properties are incidental to an electron behavior somewhat similar to that between parallel plates with uniform field, as described in the preceding section.

If as is usually true the cathode radius is small relative to that of the plate, an electron released at the cathode with negligible initial velocity pursues a path that is approximately circular, and returns ultimately to the cathode surface, as shown in the figure. The reason for this is that except in the immediate neighborhood of the cathode the linear velocity

is nearly constant.   Fig. 33*b* illustrates the variation of potential in such
a region (no space charge) and Fig. 33*c* the corresponding variation
in an electron's total velocity.   There is a very rapid growth of velocity
in the near neighborhood of the cathode; but after the electron emerges
from that region its velocity, therefore also the radius of curvature of its
path, changes very little.

(a) Concentric electrodes with magnetic field
(not shown) parallel to axis.

(b) Potential distribution diagram.

(c) Velocity diagram.
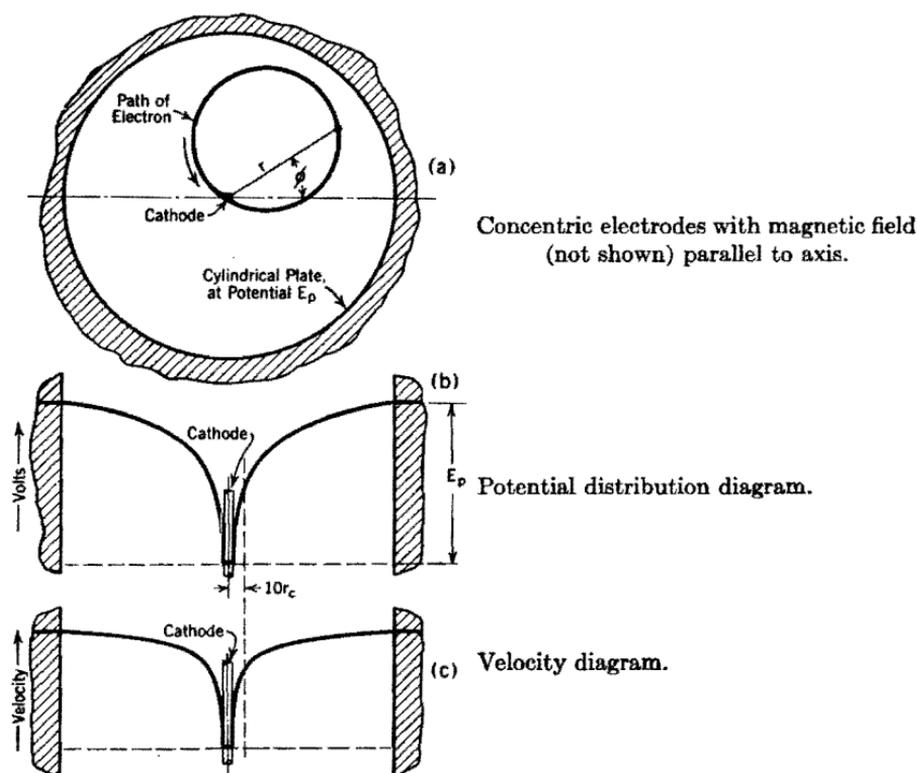
FIG. 33.   Circular path of an electron in radial electric field and axial magnetic field.

The important property of such a device is the relation between mag-
netic field strength, anode potential, and tube geometry that is neces-
sary to cause the path just to graze the anode.   The velocity at the
grazing point is of course that corresponding to plate potential, and the
radius r_p of the plate is twice the radius of the grazing path.   The radius
of the grazing path is that due to movement at a velocity $\sqrt{E_p}$ square
root volts in a magnetic field of strength B.   The desired overall inter-
relation, obtained directly by the use of Equation (167) is

$$B = 2\frac{3.37 \sqrt{E_p}}{r_p} \tag{184 p}$$

This is an approximate relation that is valid only when the plate radius is considerably larger than the cathode radius. It is correct within about one per cent when the outer radius is ten times the inner.

In actual devices there is considerable space charge near the cathode; more elaborate theory predicts, and the prediction is confirmed by experiment, that the resulting change in potential distribution in the neighborhood of the cathode does not appreciably alter the validity of Equation (184).

It has been stated that Equation (184) can be applied only in case the plate radius r_p is considerably greater than the cathode radius r_c. It is desirable to show how to determine how much greater r_p should be than r_c in order to keep the error small. Just as between parallel plates, there are two differential equations of electronic motion; they relate angular and radial motion respectively to the field forces. Since at the grazing point the only kinetic energy is that due to angular motion, only the angular-motion equation is needed to obtain an exact expression for the critical magnetic field.

The electric field can produce only *radial* acceleration; the *angular* acceleration must be the result of a torque produced by the action of the magnetic field on a radial component of velocity, v_r. This torque is $r \cdot Bev_r$; it can also be written $Ber\, dr/dt$, which is the same as $\frac{1}{2}Be\, dr^2/dt$. Force is usually equated to mass times acceleration; it can also, and with considerably greater generality, be equated to rate of change of momentum. The two relations are identical as long as the mass is constant. Torque can similarly be equated to rate of change of angular momentum. Angular momentum is in this case the product of angular velocity $d\phi/dt$ by moment of inertia $m_e r^2$. The following equation [9,13] between the torque and the time derivative of angular momentum

$$\frac{d}{dt}\, m_e r^2 \frac{d\phi}{dt} = \tfrac{1}{2} Be \frac{dr^2}{dt} \qquad (185\ \text{esu})$$

can be integrated with respect to time by canceling a pair of time differentials and adding a constant of integration, thus

$$m_e r^2 \frac{d\phi}{dt} = \tfrac{1}{2} \frac{Ber^2}{c} + k \qquad (186\ \text{esu})$$

The electron is assumed to start from rest at the cathode surface, so that $r = r_c$ when $d\phi/dt = 0$, which makes $k = -Ber_c^2/2c$. Using this expression for k and rearranging,

$$\frac{d\phi}{dt} = \frac{\omega}{2}\left(1 - \frac{r_c^2}{r^2}\right) \qquad (187)$$

Here ω has the meaning assigned to it in Equation (180), and is not angular velocity referred to the polar axis.

Translational kinetic energy is $\frac{1}{2}m_e v^2$. Rotational kinetic energy can be similarly expressed by substituting moment of inertia for mass, angular velocity $d\phi/dt$ for velocity. At the grazing point all the kinetic energy is due to rotational motion, and is that due to plate potential, so that

$$eE_p = \frac{1}{2}(m_e r_p^2)\left[\frac{\omega}{2}\left(1 - \frac{r_c^2}{r_p^2}\right)\right]^2 \qquad (188 \text{ esu})$$

which solves, for B, involved in the factor ω, into

$$B = \sqrt{\frac{8m_e c^2}{e}} \; \frac{1}{\left(1 - \dfrac{r_c^2}{r_p^2}\right)} \; \frac{\sqrt{E_p}}{r_p} \qquad (189 \text{ esu})$$

or, in practical units

$$B = 2\frac{3.37 \sqrt{E_p}}{\left(1 - \dfrac{r_c^2}{r_p^2}\right)r_p} \qquad (190 \text{ p})$$

When r_p is ten times r_c, the simpler expression, Equation (184), is correct to within one per cent.

Hull and Langmuir[9, 13] have developed complete mathematical solutions for the electronic path when, as in the above discussions, space charge is neglected; also when space charge due to uniform electron emission from the cathode is considered. Their work shows that for either analysis Equation (184) gives a satisfactory approximation to the critical value of the magnetic field, if the ratio r_p/r_c is reasonably large. This is true because the important modifications introduced by space charge occur near the cathode, where the electron concentration is high, whereas the value of the critical magnetic field depends chiefly on the electron velocity at points remote from the cathode, which is dependent almost solely on the plate potential; see Fig. 33c.

It is interesting to note that the term $-\dfrac{\omega}{2}\dfrac{r_c^2}{r^2}$ in Equation (187) plays the same part in this concentric geometry that the balanced-force velocity does in the analysis of an electron's motion in uniform combined fields between parallel plates. It describes the angular velocity that the magnetic field must have at each radius r to generate a " back field " equal to the existing electric field at that radius. This angular velocity is exactly $\omega/2$ at the cathode surface, and becomes rapidly smaller as the radius increases. Spinning the uniform magnetic field as a whole can compensate for the electric force *only at some one radius*.

The greater complexity of the electronic motion in concentric than in parallel-plane geometry arises principally from this fact.

With B, E_p, and r_p related as in Equation (184), an oscillatory current may be generated in an external resonant circuit, the period of oscillation being the same as that required for an electron to traverse the nearly circular path illustrated in Fig. 33a. If the path were exactly circular, and traversed at uniform velocity, the period of oscillation would be determinable from Equation (169). However, the true path is similar to a cardioid, and is slightly longer than a circular path. Furthermore, at initial emergence from the cathode, and again just before return to it, the electron's velocity is distinctly less than that elsewhere (see Fig. 33c). The actual cyclic time is greater than as given by Equation (169). Experiments have shown that the period may be expected to be about 23 per cent in excess of that predicted by Equation (169).

An often-used empirical equation between B and the wave length λ of the electromagnetic radiation from a magnetron oscillator is as follows:

$$\lambda B = 13,000 \qquad (190.1)$$

Either of at least two mechanisms may conceivably account for the generation of oscillatory currents in an external resonant magnetron plate circuit, as follows:

(1) If the plate voltage rises a little above the critical value, B and r_p being unchanged, electron flow into the plate and through the plate circuit causes a voltage drop in the resonant circuit element, with resulting decrease in plate voltage below the critical value. Electrons emerging from the cathode immediately thereafter experience a weakened electric field, consequently fail to reach the plate, so that current flow stops. The plate voltage therefore rises again; electrons subsequently emerging from the cathode do reach the plate, and so on. The *transit time* required for electron travel from cathode to plate is a half period of the cyclic behavior.

(2) Of course electrons start out initially from all points on the cathode; the electrons traversing many nearly circular paths like that in Fig. 33a constitute an electric current flowing clockwise around the inside of the plate (electron flow is counterclockwise). This circular sheet of current will set up a magnetic " reaction " field which opposes the applied magnetic field. Electrons starting from the cathode after B is weakened by the presence of the reaction field pursue paths of enlarged radius and so reach the plate. The current-sheet strength is consequently cut to half or less of its former value, and the reaction field is correspondingly reduced; the total field is therefore strengthened. Electrons subsequently leaving the cathode pursue a small-radius path,

and fail to reach the plate.   The cycle is then repeated, the transit time being a half period of the cyclic behavior.

In a " split-anode magnetron " the anode consists of two semi-cylinders, one on each side of the axis, connected through a resonant circuit.   Split-anode magnetrons are more commonly used than any other type of short-wave magnetron oscillator.

34. Mass: a Property Due to Electric and Magnetic Fields.   The electric field between the plates of a condenser represents storage of energy; so also does the electric field due to its own charge that surrounds an electron.   Likewise the magnetic field that surrounds a moving electron represents energy storage, in a form similar to that around an inductance carrying a current.

According to modern physical theory mass is a manifestation of energy;[B 535] the nature of the interdependence is indicated by the equation

$$m_e = \frac{W_e}{c^2} \tag{191}$$

which relates the electron's mass m_e to the energy storage W_e (ergs) in its electric field. c is as usual the ratio between the esu and emu values of the electrical units, also the velocity of light.

If the charge on an electron is assumed to be distributed uniformly over a hollow sphere of radius a

$$W_e = \frac{\text{charge} \times \text{potential}}{2} = \frac{1}{2} e \frac{e}{a} = \frac{e^2}{2a} \tag{192}$$

This permits an estimate of an electron's dimensions.   Combining Equations (191) and (192), and using $m_e = 9 \times 10^{-28}$, the radius is found to be

$$a = 1.41 \times 10^{-13} \text{ cm} \tag{193}$$

It is very unlikely that the charge on an electron is distributed uniformly around a spherical shell, so that $e^2/2a$ cannot be more than a first approximation to the true expression relating dimensions to energy storage.   These calculations merely indicate that an electron's radius should be in the neighborhood of 1 or 2×10^{-13} cm.   Field calculations show that 99 per cent of the electric energy storage must lie within 2×10^{-11} cm of the electron's center, because of the very great field concentration.

35. "Rest Mass" of an Electron;   Increase of Mass at Large Velocities.   The energy stored in a moving electron's magnetic field also contributes to its mass.[38]   In general then the total mass m of an

electron is

$$m = \frac{W_e}{c^2} + \frac{W_m}{c^2} \tag{194}$$

where W_m is the magnetic energy storage. At standstill there is no magnetic field, so that W_m is zero; only W_e/c^2 remains. Hence m_e, which was identified with W_e/c^2 in the preceding paragraph, is called the " rest mass " of an electron.

In case an electron's velocity approaches that of light, as happens in some engineering devices, the contribution of the magnetic term W_m/c^2 is not negligible. Determination of mass at these high velocities starts with the observation that in all cases W_m is the stored-up form of the kinetic energy Ee acquired by the electron in traversing the potential difference E of an accelerating field. That is,

$$\frac{W_m}{c^2} = \frac{Ee}{c^2} \tag{195 esu}$$

The electron's mass can therefore be expressed as

$$m = m_e + \frac{Ee}{c^2} \tag{196 esu}$$

Since the rest mass m_e is a constant,

$$\frac{dm}{dt} = \frac{d}{dt}\left(\frac{Ee}{c^2}\right) = \frac{1}{c^2}\frac{d}{dt}(Ee) \tag{197 esu}$$

But the rate of change of energy $d\,(Ee)/dt$ is power, and the power input to the electron is the accelerating force times the velocity, so that

$$\frac{dm}{dt} = \frac{fv}{c^2} \tag{198 esu}$$

When acting on a particle of changing mass, force is rate of change of momentum, therefore

$$\frac{dm}{dt} = \frac{v}{c^2}\frac{d(mv)}{dt} \quad \text{or,} \quad dm = \frac{v}{c}d\left(\frac{mv}{c}\right) \tag{199}$$

If the indicated differentiation is carried out, the resulting expression separated according to the variables m and v/c, and integrated using an integration constant $\log A$, there results

$$\log m = \log \frac{1}{\sqrt{1 - \dfrac{v^2}{c^2}}} + \log A \tag{200}$$

or

$$m = \frac{m_e}{\sqrt{1 - \dfrac{v^2}{c^2}}} \qquad (201)$$

for when $v = 0$, $m = m_e = A$. Thus the electron's mass approaches an infinite value as its velocity approaches that of light.

36. "Transverse Mass" and "Longitudinal Mass." Equation (201) permits calculation of the mass and momentum of an electron of known velocity. If an electron that has already acquired a high velocity is introduced into an electric field, its acceleration is obtained from an equation between the field force and the rate of change of momentum. Mathematically:

$$f = \frac{d}{dt} (mv_F) \qquad (202)$$

where v_F is the velocity in the direction of the force. The electron's mass m is dependent on its total velocity v. It is convenient to treat two distinct conditions: (1) the electric field perpendicular to the direction of the electron's initial motion, and (2) the electric field in the direction of the initial motion.
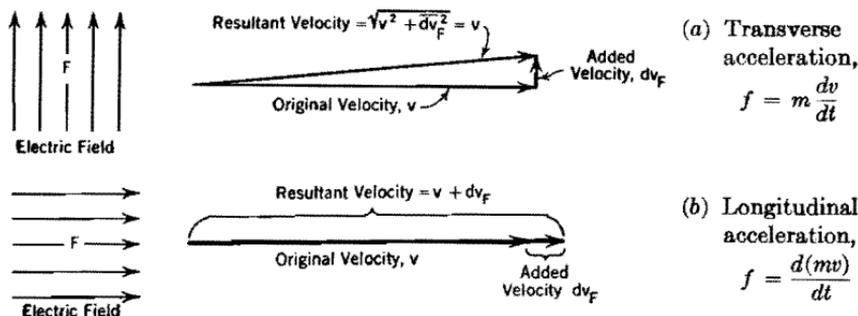
FIG. 34. Transverse and longitudinal acceleration of a high-velocity electron.

In the first case the added velocity, being a differential increment normal to the original, does not change the total velocity, for the vector resulting from the addition of a very small velocity at right angles to a very large one has the same length as the large one (see Fig. 34a). Hence the mass remains constant at the value given by Equation (201), and Equation (202) becomes

$$f = \frac{m_e}{\sqrt{1 - \dfrac{v^2}{c^2}}} \frac{dv_F}{dt} \qquad (203)$$

This has the general form: force = mass times acceleration, and gives rise to the name and symbol " transverse mass," m_t, of value:

$$m_t = \frac{m_e}{\sqrt{1 - \dfrac{v^2}{c^2}}} \tag{204}$$

which must be used in determining a high-speed electron's acceleration in a direction normal to its existing motion. m_t is 10 per cent greater than m_e when $v = 1.26 \times 10^{10}$ cm per second, 42 per cent of the velocity of light. In the next section m_t is expressed in terms of the electron's energy in electron volts.

In the second case the change in velocity adds its full value to the existing velocity, for the very small vector is added as an extension to a very large one (see Fig. 34b); dv_F becomes simply dv, and since m is now affected by the change in total velocity, the time derivative of the momentum contains dv/dt in one term, dm/dt in the other. The force expression simplifies to

$$f = \frac{m_e}{\left(1 - \dfrac{v^2}{c^2}\right)^{\frac{3}{2}}} \frac{dv}{dt} \tag{205}$$

which gives rise to the concept " longitudinal mass," m_l, of value

$$m_l = \frac{m_e}{\left(1 - \dfrac{v^2}{c^2}\right)^{\frac{3}{2}}} \tag{206}$$

which can be used in calculating instantaneous acceleration along the line of existing motion.

37. Relation of Velocity and Mass to Accelerating Potential. The longitudinal mass is usually of academic interest only. Calculations of the nature suggested by the name are accomplished by setting up an expression similar to Equation (152) but more general in application. The familiar expression $v = ds/dt$, in which ds stands for a differential increment of distance, can be rearranged to read $dt = ds/v$, and this value for dt substituted into Equation (205). Both sides can then be integrated, the left side (in the form $eFds$) with respect to distance, the right side with respect to velocity; there is no constant of integration if the electron is considered to be initially at rest. The result is [38]

$$Ee = m_e c^2 \left(\frac{1}{\sqrt{1 - \dfrac{v^2}{c^2}}} - 1 \right) \tag{207 esu}$$

For velocities small enough so that $\frac{v}{c} \ll 1$ a binomial expansion reduces Equation (207) to Equation (152).

Equation (207) can be rearranged into the form

$$\sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{1 + \frac{Ee}{m_e c^2}} \qquad (208 \text{ esu})$$

or, in practical units

$$\sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{1 + \frac{E}{510,000}} \qquad (209 \text{ p})$$

If this expression is solved for v it gives the velocity as a function of the accelerating potential, as follows:

$$v = c\sqrt{1 - \frac{1}{\left(1 + \frac{E}{510,000}\right)^2}} \qquad (210 \text{ p})$$

which reduces to Equation (154) if $E \ll 510,000$ volts.

Equation (209) also permits expression of the "transverse mass," which must be used in estimating response to magnetic or transverse electric fields, in terms of the accelerating potential, in the following form:

$$m_t = \frac{m_e}{\sqrt{1 - \frac{v^2}{c^2}}} = m_e\left[1 + \frac{E}{510,000}\right] \qquad (211 \text{ p})$$

Equation (211) indicates that the per cent increase in mass is directly proportional to the accelerating voltage. For example, an electron that has been accelerated through a potential difference of 51,000 volts has a mass 10 per cent greater than its rest mass, and experiences in a crosswise electric field a transverse acceleration that is 91 per cent of the value to be expected on the basis of rest mass. If the accelerating potential is 510,000 volts, the mass increase is 100 per cent, and the transverse acceleration is halved. These relations have been demonstrated experimentally.

38. Motions in Irregularly Curved Fields. The motion of an electron in a curved field is complicated by the carry-over into each new part of the field of the directed inertia gained earlier. The general differential

equations for motion in an electrostatic field, in rectangular coordinates, are:

$$\frac{d^2x}{dt^2} = -\frac{e}{m_e}\frac{\partial E}{\partial x}$$

$$\frac{d^2y}{dt^2} = -\frac{e}{m_e}\frac{\partial E}{\partial y} \qquad\qquad \text{(212 esu)}$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m_e}\frac{\partial E}{\partial z}$$

Of course

$$-\frac{\partial E}{\partial x} = F_x, \quad -\frac{\partial E}{\partial y} = F_y, \quad -\frac{\partial E}{\partial z} = F_z$$

where F_x, F_y, F_z are the components of electric field strength in the x, y, and z directions.

The motion of an electron in a curved magnetic field is subject to another set of differential equations, as applied to the coordinate system illustrated in Fig. 35, as follows:

$$\frac{d^2x}{dt^2} = \frac{e}{m}\left[B_z\frac{dy}{dt} - B_y\frac{dz}{dt}\right]$$

$$\frac{d^2y}{dt^2} = \frac{e}{m}\left[B_x\frac{dz}{dt} - B_z\frac{dx}{dt}\right] \qquad\qquad \text{(213 emu)}$$

$$\frac{d^2z}{dt^2} = \frac{e}{m}\left[B_y\frac{dx}{dt} - B_x\frac{dy}{dt}\right]$$

where B_x, B_y, B_z are the components of magnetic flux density in the x, y, and z directions. Treatises on partial differential equations and on theoretical mechanics contain methods for the mathematical solution of these equations for special cases. In combined fields, add Equations
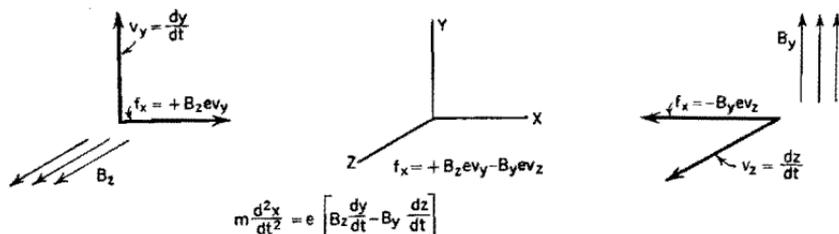
FIG. 35. Components of force on an electron in a curved magnetic field.

(212) in proper sequence to Equations (213), after converting the latter to electrostatic units by dividing the right-hand members by c.

Simplification is sometimes possible by conversion into cylindrical or polar coordinates. It is always important, as in all field problems,

to consider symmetries and to take advantage of the known behavior in certain types of localized regions.

If the shape of the field, either electric or magnetic, has been obtained by flux mapping or other graphical means, so that its mathematical expression is not known, or if the mathematical solution for electron motions is hopelessly difficult, step-by-step graphical methods may lead to a sufficiently accurate knowledge of the path. Such processes are facilitated by recalling that, no matter where an electron may cross a given equipotential, the magnitude of its velocity at crossing is fixed by Equation (152) or an equivalent. And of course passage through a magnetic field neither adds to nor subtracts from the magnitude of an electron's velocity; it affects only the direction.

In combined nonuniform fields the motion within any given *local* region is (*a*) motion at right angles to the electric field with just the velocity necessary to counteract magnetically the electric force at that point, vectorially added to (*b*) circular or helical motion of angular velocity dependent entirely on the magnetic field, and with radius and initial direction such that when vectorially added to (*a*) it gives the actual velocity at entrance to the local region. This is illustrated by initial-condition vectors in Figs. 31 and 32.

PROBLEMS

CHAPTER III

1. It is stated in Section 26 that the average kinetic energy of thermal motion possessed by molecules of air at $40°$ C is 0.0675 electron volt. Only $\frac{2}{3}$ of this average energy is due to translational motion, the remainder being due to rotation and vibration of the dumbbell-like molecules that air consists of. What is the velocity, in centimeters per second, of a nitrogen molecule that possesses the average translational energy? (Note that this is not the same as the average velocity; see Chapter X.) The atomic weight of nitrogen is given in Table V.

2. Suppose that, in Fig. 27, $s = 2.5$, and $E_s = 50$ volts. Assume that both plates are infinite in extent. State where, with what velocity components and after what time interval, the electrons first strike one or the other of the electrodes if the gun shoots 100-volt electrons at angles with the horizontal of (*a*) $30°$, (*b*) $45°$, (*c*) $60°$, (*d*) $90°$.

3. Suppose that an electron starts with zero initial velocity from a point on the cathode of Fig. 14 almost opposite a grid-wire center, so experiences considerable y-directed acceleration before reaching the plane of the grid. Assume that it crosses the plane of the grid at a point that is 64 volts above cathode potential, that its y-directed component of velocity at the point of crossing is 40×10^7 cm per sec, and that it experiences no y-directed acceleration after crossing the plane of the grid. Plate potential is 300 volts above cathode potential, and the plate is 0.4 cm beyond the grid plane.

(*a*) State the electron's x-directed velocity component at the grid-plane crossing point.

(b) What will be the electron's y-directed velocity when it reaches the plate?

(c) What will be its x-directed velocity when it reaches the plate?

(d) How long after crossing the grid plane will it reach the plate?

(e) How far will it move in the y direction meanwhile?

(f) How much energy, in electron volts, will it deliver to the plate on arrival there?

4. An electron crosses the grid plane with the same velocity components as in Problem 3, and experiences no y-directed acceleration thereafter; but the plate is now only 10 volts above cathode potential.

(a) How close to the plate will the electron get before turning back?

(b) To what potential must the plate be raised in order to permit this electron to reach it?

5. The magnetic field strength inside a hollow cylindrical current-carrying coil is 100 flux lines per square centimeter. An electron gun protrudes through the coil in just the way one protrudes through the lower plate in Fig. 27.

(a) If the gun is perpendicular to and aimed directly at the axis, what must be the energy of the electrons in the beam in order to make them emerge through a hole in the coil just 90° around its circumference from the gun's location?

(b) At what angle with the axis must the gun be inclined, and with what energy must the electrons leave it, in order to make them emerge through the hole if it is still 90° around the coil's circumference but shifted 5 cm axially from the gun's location?

6. In a *cyclotron* a strong magnetic field makes *ions* pursue successive semicircular paths of increasing radii, so that the overall effect is that of spiral motion. After each semicircle of path the ions pass through a very short portion of arc within which there is a tangential accelerating electric field, so that during each new semicircle of path the ion's energy is greater than during the last one. The electric field must be periodically reversed in order to encourage rotation of the same kind at the beginning and end of each semicircle. Suppose that heavy hydrogen ions (atomic weight 2) are used, and that the magnetic field strength is 16,000 flux lines per sq cm. What must be the frequency of the electric power supply for the accelerating field? If each passage through the accelerating field increases the ionic energy by 25,000 electron volts, how much time will elapse between introduction of an ion at the center of the spiral and its emergence at the rim as a 5,000,000-volt ion? What will be the radius of the last semicircle before such emergence? What would the radius be if argon ions were used?

7. Suppose that, in Fig. 30, $F = 1,500$ volts per cm and $B = 100$ flux lines per sq cm. What must be the energy, in electron volts, of the electrons that emerge from the gun in order to:

(a) Make them pursue the straight-line path indicated in Fig. 30?

(b) Make them pursue a cycloidal path like that illustrated in Fig. 31?

(c) How far to one side of the gun's axis would the cusps of the cycloid lie? (The gun delivers electrons at a point in the path corresponding to D, Fig. 31.)

8. Electric and magnetic fields and an electron gun are located as in the previous problem, and F and B have the values specified there. The electrons pursue a trochoidal path (Fig. 32) with tracing-spoke radius as long as the distance from the gun to the plane of the cusps in the previous problem.

(a) What must be the kinetic energy, in electron volts, of the electrons as they emerge from the gun?

(b) Find the velocity of the electrons when at the extreme left point of this trochoidal path.

(c) At what angle, and with what energy, would such electrons strike a plane parallel to the gun's axis, and to the left of it a distance equal to the radius of the tracing spoke?

(d) Same as (c), except that the plane is only half as far to the left of the gun.

9. (a) What is the transverse mass of an electron when traveling with half the speed of light?

(b) Through how large a potential must it "fall" to acquire this velocity?

(c) What will be the arc radius and time for one complete rotation in the circular path pursued by such an electron when it enters a magnetic field of 1,000 flux lines per sq cm?

10. Show that Equation (190.1) calls for a cyclic period of electron motion approximately 23 per cent greater than that obtained from Equation (169), using the fact that the product of λ by frequency must equal the velocity of light, 3×10^{10} cm per sec.

CHAPTER IV

CATHODE RAYS

39. Cathode Ray: a Name for a Beam of Electrons. A small jet or beam of electrons in a vacuum, which may be ejected from an " electron gun "[15] is often called a *cathode ray* because it initially emerges from the more negative electrode or cathode of the high-vacuum apparatus in which it originates, and possesses many properties similar to those of a ray of light. Just as a light beam can be reflected, refracted, and focused by a system of lenses and mirrors, so an electron beam can be reflected, refracted, and focused by a system of electric and magnetic fields. The branch of science called *Electron Optics*[16, 17, 18, 19] deals in detail with the laws of behavior of these electron beams.

Cathode rays are used in laboratories for measuring very rapidly changing voltages and currents, in television apparatus as part of phenomenally high-speed switching devices, in hospitals for the production of X-rays, and in general for a variety of experimental services. It is impossible to predict the extent of future applications. The major quantitative aspects of the behavior of cathode-ray devices can be satisfactorily analyzed by studying the trajectory of individual electrons according to the principles developed in Chapter III. The effects of the space charge produced by the electrons themselves, that is, the mutual repulsions of the electrons in the beam, make focusing arrangements necessary.

40. The Cathode-Ray Oscillograph. Probably the best-known device employing an electron beam is the cathode-ray oscillograph,* used for measuring electrical quantities; Fig. 36a illustrates a simple but typical geometrical arrangement. The cathode ray originates at the cathode, passes through an accelerating field between cathode and anode, through a hole in the anode, between one or two pairs of " deflecting plates," and finally terminates on a sensitized surface, which is either fluorescent material or a photographic film, according to whether visual observation or direct photographic recording of the beam's trace is desired.

The voltage to be measured is connected to a pair of deflecting plates, so that as the beam passes between them the electrons experi-

---

* See reference books *L, M, N, O*, and reference articles 16 to 56, inclusive, in the bibliography.

ence a transverse acceleration. The resulting bend in the trajectory makes them strike the screen at one side or the other of the zero position, the distance from the zero being proportional to the potential difference between the deflection plates. If instead of a pair of deflecting plates, a current-carrying coil is used, so placed that its magnetic field produces

(a) General arrange-   (b) Detail of deflecting   (c) True deflecting plate
        ment.   plates.   field.
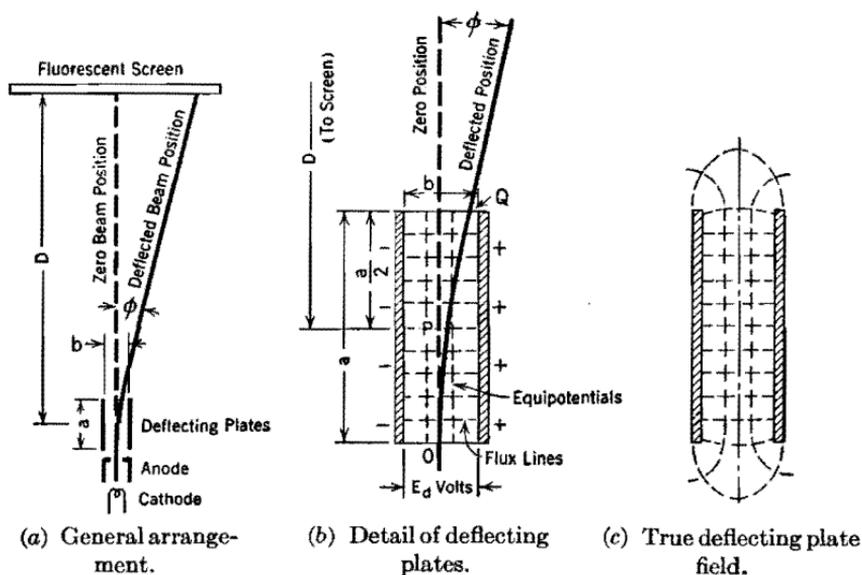
Fig. 36.   Electrostatic deflection of the beam of a cathode-ray oscillograph.

a bend in the beam's path, the deflection from the zero position measures the current in the coil. The time axis necessary for any oscillographic record is obtained by a movement of the beam at right angles to that due to the deflecting plates or magnetic field. This time-axis or " sweep " motion may be produced either electrically, by a second set of deflecting plates, or magnetically.

41. Voltage Sensitivity. Electrostatically produced deflection of the beam of a cathode-ray oscillograph is proportional to the voltage to be measured, inversely proportional to the accelerating voltage, and dependent on the geometrical arrangement of the parts.[23, 33] Figure 36b is a detail of the electron's path in the immediate vicinity of deflecting plates of length a, spacing b, in centimeters. The actual form of the field at the two extremities of the plates is as shown in Fig. 36c, but a sufficiently accurate analysis of the response of the beam can be obtained by assuming a uniform field up to the ends of the plates and zero field beyond.

The electrons enter the deflecting field with a velocity that depends upon the accelerating voltage according to Equations (154) or (210).

While in the deflecting field each electron has: (1) constant velocity in the original direction, and (2) uniformly accelerated motion in the direction of the field force, that is, at right angles to the original direction. The resulting path between the plates is parabolic, like that of a ball which is subjected to a horizontal velocity plus uniform gravitational acceleration, by being thrown outward from the top of a tower.

The vertex of the parabola is at O, the entrance to the transverse field. The vertical component v_a of the velocity remains constant, the transverse component v_t increasingly uniformly with time. The path continues to be parabolic to the point of emergence from between the plates. There the deflecting field ends and v_t becomes constant at a value which will be called v_d, and the trajectory straightens out into a direct line at an angle ϕ with the original direction. The deflection of the beam's terminus on the screen from the zero location depends on the angle ϕ and the distance D from the center of the plates to the screen. For a truly parabolic path the tangent to the trajectory at the point of emergence Q passes through P, the midpoint of the field.

One component of the ultimate straight-line motion is v_a, the other is v_d, the transverse velocity acquired while passing between the plates. The deflection x on the screen is dependent on these velocity components by way of the angle ϕ, as follows:

$$x = D \tan \phi = D \frac{v_d}{v_a} \tag{214}$$

If the accelerating voltage is small, the deflection for a given transverse field is correspondingly large, for each electron then moves very slowly through the deflecting field, and is exposed for a long time (a/v_a) to the transverse acceleration dv_t/dt. The dependence of v_d on time of exposure is expressed mathematically by the equation

$$v_d = \frac{a}{v_a} \frac{dv_t}{dt} \tag{215}$$

Using Equation (143) for relating the transverse acceleration to the field, recognizing that the strength of the electric field between the plates is E_d/b, and using Equation (152) to express the incoming velocity in terms of the accelerating or anode voltage E_a, Equation (214) becomes

$$x = D \frac{a}{2b} \frac{E_d}{E_a} \tag{216}$$

E_d is of course the voltage that has been applied to the plates in order to be measured. If E_a is large enough so that the electron's mass is appreciably more than the rest mass, the incoming velocity and the

mass factor in the acceleration should be expressed as in Equations (210) and (211), giving:

$$x = D \frac{a}{b} \frac{E_d}{E_a} \frac{(E_a + \quad 510{,}000)}{(E_a + 1{,}020{,}000)} \qquad (217)$$

This reduces to Equation (216) if $E_a \ll 510{,}000$ volts.

42. Magnetic Sensitivity. Magnetically produced deflection in a cathode-ray oscillograph varies in direct proportion to the strength of the field, inversely as the square root of the accelerating voltage, and is dependent on geometrical proportions.[23, 33]   Figure 37 is a detail of the

bend in the electron beam as it passes through a magnetic field substituted for the electric deflecting field of Fig. 36.   The magnetic field is of course directed at right angles to the desired deflection.   Analysis is based on the reasonably satisfactory assumption of a sharply bounded magnetic field having a value B lines per square centimeter within a restricted region and zero value elsewhere.

Each electron pursues a circular arc of radius r while in the magnetic field; radii drawn to the path at O and Q are per-
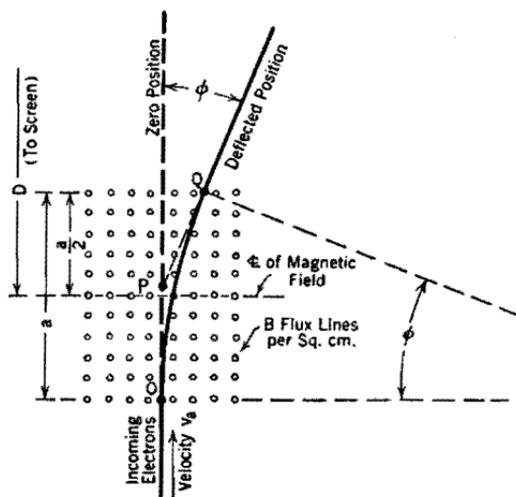
Fig. 37.   Magnetic deflection of a cathode-ray oscillograph beam.

pendicular respectively to the zero and deflected beam directions.   Hence the total angle of arc traversed between these radii is the same as the deflection angle ϕ.   The deflection is usually small enough so that the point P obtained by backward projection of the out-going beam line may be identified with the midpoint of the field from which D is measured; it is also small enough so that the angle ϕ is indistinguishable from $\tan \phi$ and from a/r, where a is the length of the magnetic field in centimeters.   As with the electric field

$$x = D \tan \phi \qquad (218)$$

Since $\tan \phi = \phi = a/r$, the deflection can be immediately evaluated by using Equation (167) for the radius, as follows:

$$x = D \frac{a}{r} = D \frac{aB}{3.37 \sqrt{E_a}} \qquad (219\ p)$$

43. Measuring-Circuit Relations.   The deflecting plates of a cathode-ray oscillograph have a finite electrostatic capacitance, of value, in statfarads

$$C = \frac{S}{4\pi b} \text{ (statfarads)} \qquad\qquad (220 \text{ esu})$$

or, in practical units

$$C = \frac{S}{3.6\pi b} \text{ (micro-microfarads)} \qquad\qquad (221 \text{ p})$$

S being the area of one of the plates and b their distance apart, in centimeter units.   Although this capacitance is rarely more than a very few micro-microfarads, it may, if the frequency is very high or transients very abrupt, have an appreciable effect on the circuit whose behavior is being measured.

In some cases the distributed inductance and capacitance of the deflection plate leads, combined with the terminal impedance of the plates, form an oscillating system whose effects mask the voltage variations being investigated.   Proper lead-wire analysis treats the incoming wires as a transmission line, which the impulses to be recorded traverse, as traveling waves, with a finite velocity of propagation. These waves are subject to forward or reversed reflection at the terminals, depending on the nature of the terminal impedances.[55, 56]

In general, troublesome lead-in-circuit oscillations are not to be expected unless the duration of the transient to be measured is less than one cycle of the natural frequency of the lead-in system.   Damping resistors, of magnitude dependent on transmission constants, placed at the source end of the lead-in transmission system, are sometimes helpful in stabilizing the response to very abrupt transients, and in eliminating lead-wire circuit oscillations initiated by pick-up from auxiliary circuits.

In making magnetic-deflection measurements of currents, it is often important to use deflecting coils with inductances small relative to that of the circuit whose current is to be measured.   The cathode-ray deflection is proportional to the extent of the coils along the line of beam travel, but not to their width, so that by using long rectangular coils, no wider than needed to make the field as wide as the transverse motion of the beam requires, the total flux, and with it the inductance, may be kept small.   Also, two small coils placed close in on either side of the beam can produce the same deflection as two larger ones farther out, and with very much less inductance.   For given coil shape, inductance is proportional to linear dimension.   Bureau of Standards Handbook 74 is helpful in calculating inductances and field strengths of coils, and electrostatic and magnetic properties of lead-in systems.[P]

44. Photographic and Visual Sensitivity; Penetration of High-Velocity Electrons. The blackening of a photographic film by an electron beam varies in proportion to the concentration of electrons in the beam, as the square of the accelerating voltage, and inversely as the rate of travel of the trace on the film.[23, 33] Photographic response to the electrons continues as long after they enter the sensitive layer as they are in motion, and the depth of penetration varies as the fourth power of the velocity on arrival at the surface. An approximate mathematical statement[23] is:

$$v_a{}^4 - v_s{}^4 = 5.05sd10^{42} \tag{222}$$

where v_a is the velocity on arrival at the surface, v_s that after penetrating to the depth s, and d the density of the photographic emulsion.

If the electron energies are expressed in electron volts, by the aid of Equation (154), this takes the form

$$E_a{}^2 - E_s{}^2 = 4.04sd10^{11} \tag{223 p}$$

where E_a and E_s are the kinetic energies in electron volts that correspond to the velocities v_a and v_s. The extreme penetration is reached when the energies of the electrons vanish; the maximum depth of penetration is therefore

$$s = \frac{E_a{}^2}{4.04d10^{11}} \tag{224 p}$$

It is not unreasonable to suppose that the density of the gelatin that forms the base of a photographic emulsion is approximately unity. In that case the photographic penetration of 100-volt electrons is about 2.5×10^{-8} cm, or about the thickness of one or two layers of molecules, while 1,000-volt electrons may be expected to penetrate through about one hundred such layers.

Sometimes the evacuated chamber is arranged so that the beam strikes a thin metal foil (" Lenard ") window through which electrons can pass outward but no air pass inward, and registration is made on a photographic film just outside the window. Wood gives values of the quantity $5.05d \times 10^{42}$ and corresponding densities for a few substances, as follows[23]:

MATERIAL	$5.05d \times 10^{42}$	DENSITY
Air	0.02×10^{42}	0.004
Aluminum	7.32×10^{42}	1.45
Gold	25.4×10^{42}	5.03

Using Wood's figure for aluminum in Equations (223) and (224), it appears that, although 60,000-volt electrons would be stopped completely by an aluminum window 0.006 cm thick, a window three-fourths of that thickness would permit them to emerge as 30,000-volt electrons, capable of traveling a few millimeters in air before losing their entire energy.

Visual observation of the beam's trace on a metal or glass screen is made possible by a chemical coating on the screen; the following substances are among those that may be used:

Zinc silicate (" willemite ") ($Zn_2 SiO_4$) . green trace

Calcium sulphide (CaS) . . . . $\begin{cases} \text{nearly white when strong,} \\ \text{green when faint} \end{cases}$

Calcium tungstate ($CaWO_4$) . . . blue

An electron beam of moderate concentration and a few hundred volts energy, too weak to register successfully on a photographic film, may make a visible trace on a fluorescent screen. The light produced may be so faint that it can be photographed only if the trace is cyclically repeated many times a second, permitting a time exposure.

Cathode-ray apparatus is available with screens so sensitive that the trace made by a beam of a few thousand volts energy in crossing a six-inch screen in 10 to 20 microseconds is intense enough to be photographed with an ordinary camera.[46, 53] Photographic registration directly on film with a 60,000-volt beam can under favorable conditions record a similar single trace produced in a few hundredths of a microsecond.

45. Production and Focusing of the Beam. The cathode from which the electrons emerge may be a heated filament similar to those used in thermionic vacuum tubes, or it may be the unheated negative terminal of a glow discharge at a very low gas pressure. The first-mentioned arrangement provides a beam with high electron concentration, and with good visual and photographic sensitivities at moderate accelerating potentials, but requiring special focusing provisions in order to secure a fine trace. A cold-cathode beam operates with a high accelerating voltage, usually more than 40,000 volts, so has rather limited voltage sensitivity. It has satisfactory photographic sensitivity, and is kept focused easily because of its low electron concentration.

Flexible control of the intensity of a beam from a thermionic filament may be obtained by the use of a grid between filament and anode, the mechanism of control being the same as that of the plate current in a triode. The life of a filament of the oxide-coated type used for this service is rapidly shortened by continuous use of an intense beam, yet

not seriously affected by occasional very brief intensification, by means of grid control, to permit photographing the fluorescent trace.

Proper focusing of the beam to produce a brilliant, narrow trace on the screen requires that the radially outward acceleration of the electrons due to their own space charge be either negligible or compensated for by a reverse acceleration of correct magnitude. In cold-cathode instruments satisfactory focusing can be obtained by using a tiny pinhole to select a fine pencil of rays, providing a vacuum is maintained in which dispersion by collision with gas particles is negligible. The electron concentration in such beams is small enough so that only a very moderate enlargement of the ray occurs after passing through the pinhole.

The highly concentrated beam from a hot filament is less easily focused than that from a cold-cathode source. The radial variation of potential caused by the space charge of a cylindrical electron beam of uniform concentration, in the absence of any external field, is illustrated in Fig. 38. The potential gradients are such as to spray the electron stream. In the absence of any special provision for focusing there should result a moderately increasing divergence of electron paths, as illustrated in Fig. 39a, the space-charge density and resultant outward acceleration becoming less as the spread increases.

Fig. 38. Potential variation within and near an electron beam, due to the beam's own space charge.

Such a beam may be electrostatically focused[34, 45, 47, 48, 49] by a *converging field* which gives the outer electrons a radially inward component of velocity. The left half of Fig. 39b illustrates the shape of a converging field between a focusing cylinder and, above and beyond, a cylindrical anode. The right half shows the response to this field of an otherwise divergent electron beam. Several radial cross sections of the space-charge-free potential in this region are shown in Fig. 39c; they indicate the presence of a pronounced central high-potential "channel" just above the focusing cylinder. The steepness of the slopes of the sides of the channel in the lower-potential sections measures the potential gradient that produces radially inward electron acceleration. The lower boundary of the shaded region underneath the lowest potential curve illustrates the nature of the 1-1 potential variation when the electron beam is present. The downward shift from the top to the bottom boundary of the shaded region is the result of the space charge of the beam. A similar shift, not shown in the figure, occurs in the other sections.

The helical nature of the electron's path in a magnetic field is used in producing *magnetic focusing* [28, 48, 49] of the beam.   The left half of Fig. 40a illustrates the nature of a magnetic focusing field, the right half the response to this field of an otherwise divergent beam.   As the electrons emerge from a hole in the anode, already with somewhat divergent velocities, they enter a magnetic field that is substantially parallel to the axis of the tube.   Each electron whose radially outward velocity is appreciable pursues a helical path which in one complete revolution

(a) Beam normally diverging; no focusing field.

(b) Flux pattern and beam, in convergent focusing field.

(c) Potential profiles at locations 1-1, 2-2, 3-3, 4-4, of Fig. 39b.
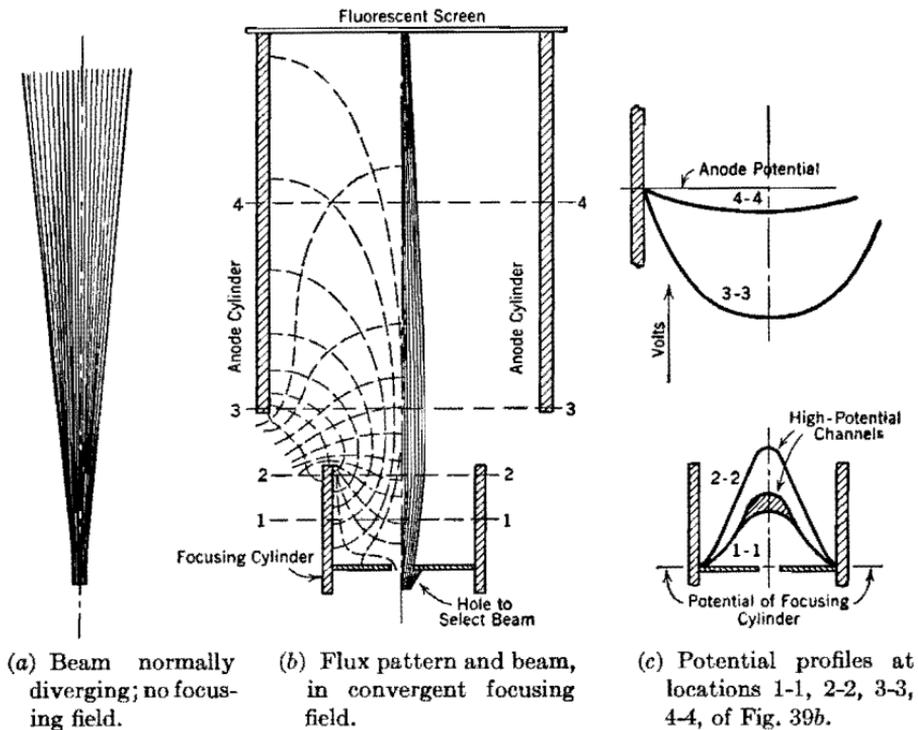
FIG. 39.   Electrostatic focusing of a cathode-ray oscillograph beam.

brings it back to a point on the axis.   Electrons having large initial outward velocities have large diameter paths, but since the time for completing a turn around the helix is dependent only upon the field strength and not at all on velocity, all return to the axis after the same length of time.   Since all have the same axial velocity, they arrive at a common focus.   If the magnetic field is of sufficient extent there may be more than one focus.   Satisfactory results have been obtained with instruments in which the deflection plates are placed near the first focus, the screen at the second.

Beam concentration may be obtained or aided by *gas focus-*

ing.[20, 40, 42, 43, 44, 48] The mechanism of gas focusing is not completely understood, but it can probably be at least partly attributed to electron and ion activity somewhat as follows:

If there is an appreciable amount of gas present, yet not enough to cause serious scattering of the electrons, some of the atoms lying in the path are ionized, each atom so affected introducing one positive ion and one electron. The electrons thus produced travel radially outward very rapidly, but the ions are so heavy that they remain practically stationary within the beam. The ions tend to make the beam's space charge positive rather than negative, while the electrons give rise to a negative charge on and near the tube walls. If ionization occurs rapidly enough the beam becomes positively charged, therefore at a higher potential than the tube walls. A converging field then exists, with definite focusing action.

The rate of ionization and consequent strength of the focusing action is dependent on the nature and concentration of the gas, the concentration of electrons within the beam, and the accelerating voltage. The electron concentration may be adjusted by filament-temperature or grid control until the focus is satisfactory; with too weak or too intense a beam the electrons do not converge at the screen. Some devices which depend primarily on electrostatic focusing contain an inert gas at a very low pressure to contribute to the sharpness of focus.
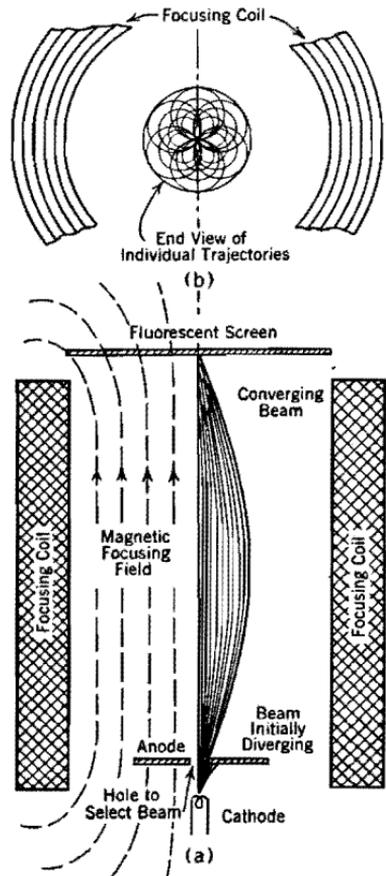
FIG. 40. Magnetic focusing of a cathode-ray oscillograph beam.

46. Time-Axis Motion. The need for a satisfactory time axis on a screen or photographic record makes it desirable for the beam to move with constant velocity at right angles to the deflection produced by the voltage or current to be measured. This motion can be produced by a second pair of deflecting plates, called *sweep plates*. If this time-axis or sweep motion takes place at a uniform velocity, the record has a *linear* time scale, which is helpful in interpreting the significance of

voltage and current variations.   For photographic registrations of a
single transient the beam need only sweep across the exposed film once,
but for the observation of recurring phenomena the motion must be
repeated periodically at a controlled frequency.

A saw-tooth sweep-plate voltage, illustrated in Fig. 41, is necessary to
produce satisfactory time-axis motion for observing cyclic phenomena.
Ideally the rising front of the wave should be a straight line of controll-
able slope.   The drop to zero should begin at a definite voltage that is
independent of the rate of rise, and should occur with extreme rapidity
in order to make the trace of the beam on the return movement faint
enough to avoid confusing the forward image.

A controlled rate of growth of the front of the saw tooth may be ob-
tained by using the voltage of a condenser that is being charged from a
high-potential direct-current source through a resistance.   The full
course of such a charging operation is indicated by the dotted line of
Fig. 41.   At some point on this dotted line, which must be low enough so
that departure from a straight line is not appreciable, the condenser is
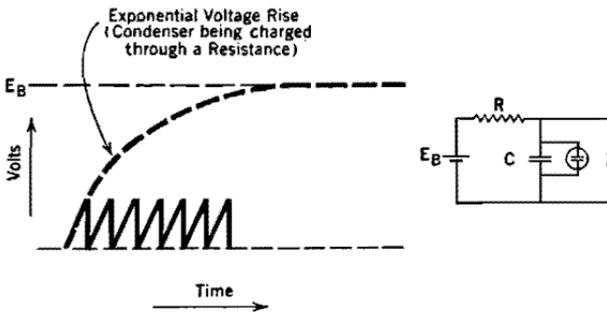
FIG. 41.   Saw-tooth voltage wave form suitable for producing time-axis
("sweep") movement of a cathode-ray oscillograph beam.

abruptly discharged by means of a suitably adjusted spark gap or elec-
tronic device.   Immediately after discharge the process begins again.
If R and C are the resistance in ohms and capacitance in farads that are
in series in such a sweep circuit, and e_C and E_B the instantaneous and
ultimate values of condenser voltage, which is applied to the sweep
plates, the circuit equation for the dotted line of Fig. 41 is

$$e_C = E_B\left(1 - \epsilon^{-\frac{t}{RC}}\right)$$
(225 p)

time t being measured from the beginning of the rise.   This makes the
initial rate of rise of each saw tooth have the value

$$\left(\frac{de_C}{dt}\right)_{t=0} = \frac{E_B}{RC}$$
(226 p)

in volts per second or per microsecond, depending on whether the *time constant RC* is expressed in seconds or microseconds.   The sweep frequency is this rate divided by the crest voltage of each tooth.

A great variety of sweep circuits, some wholly different from that just described, have been devised to suit special conditions.   For some purposes a magnetic sweep circuit, that employs the straight-line portion of a sine wave of alternating current as it rises through zero, gives satisfactory results and is easily provided.

47. Cathode Rays as Current Carriers; Television.   A cathode ray carries an electric current, flowing within the tube in whatever direction the internal electric and magnetic fields may order;   these fields are subject to the control of the operator.   The current carried by a cathode-ray oscillograph beam may be extremely small, or it may be as much as fifty microamperes;   the choice of magnitudes is based entirely upon convenience and effectiveness in producing a visual or photographic image.   For other requirements it may be increased.   Even fifty microamperes is ample to operate many sensitive electronic devices, so that the beam has all the potentialities of an extremely fast and flexible switching device, depending on current-receiving arrangements within the tube.

Up to the present time the only important application of the current-carrying property of cathode rays has been in television scanning operations.[52]   For this service the fluorescent screen is replaced by what is actually an almost infinite array of tiny condensers, having one, the outer, plate in common, the other plate of each consisting of a localized bit of photoelectrically sensitive material.   Under influence of field controls the beam *scans* this surface line by line as a draftsman cross-hatches a section;   the screen must be completely scanned from top to bottom many times a second.

Lenses like those in a camera focus an image of the action to be televised on the photoelectrically sensitive surface.   Between successive visits of the electron scanning beam each localized part of the surface acquires, photoelectrically, a charge proportional to the local light intensity and to the time since the last visit.   At the beam's passage, each tiny condenser is discharged through the beam and an external circuit connected to the metal backing plate common to all the tiny condensers.

The current in this external circuit thus varies rapidly, corresponding to the changing light intensity at points successively scanned.   This variation is broadcast over an extremely high-frequency carrier wave and made to control the beam intensity in a cathode-ray television receiver.   The receiver's beam scans a fluorescent screen in synchronism with the scanning by the transmitter's beam of the action image.   There

is thus produced on the receiving screen a light image that follows faithfully the local brightnesses and darknesses of the action image.

PROBLEMS

CHAPTER IV

1. If, in Fig. 36b, $a = 10$ cm, $b = 2$ cm, and the electron beam consists of 60,000-volt electrons, how large a value of E_d would make the electron beam just hit the upper end of one of the deflecting plates?

2. If, in Fig. 37, $a = 4$ cm, and $\phi = 10°$, how far above the center line of the magnetic field does the intersection point P lie? Demonstrate that *in Fig.* 36b the vertex P of the angle ϕ lies half-way between the ends of the plates.

3. If in a given cathode-ray oscillograph a deflecting voltage E_d and a magnetic flux density B produce equal deflections when the accelerating potential is 3,000 volts, what will be the ratio of the deflections produced by the same E_d and B when the accelerating voltage is 2,000 volts, but the geometry unchanged?

4. Write an equation for a potential distribution curve, of the type shown in Fig. 38, for a 1,000-volt electron beam 2 mm in diameter, carrying a current of 20 microamperes. Assume uniform electron concentration within the beam. Also write an equation for the radial force on the electrons in terms of the radius.

5. If Fig. 40 represents a 1,000-volt beam, what magnetic focusing field strength is necessary to bring the electrons to a focus at a distance of 10 cm from the anode? If a given electron reaches a maximum of 0.5 cm radial distance from the axis at the point of greatest beam divergence, at what angle with the axis did that electron leave the anode? Neglect effects of the radial electric field due to the beam's own space charge.

CHAPTER V

SPACE-CHARGE FLOW

48. Equilibrium between Energy, Flow, and Poisson's Equations.
Poisson's equation requires that in a region containing space charge the
electric field must vary from point to point, and cannot therefore be
everywhere zero. *It must in fact have some value almost everywhere.*
The particles that make up the space charge experience a force due to
this field, and, if they are free to move, their motion constitutes an
electric current. Any region that contains space charge due to unat-
tached particles must therefore also contain an electric current, and there
is a definite interdependence between current density, space-charge
density, and potential distribution.

Suppose that electrodes like those of Fig. 5 are maintained at constant
potential difference by an external power source; then imagine the
region between them to contain a large number of electrons that originate
at the surface of the cathode and are driven to the anode by the electric
field. In such a one-dimensional region the simple form of *Poisson's
equation* describes the local relation between potential and space charge
as

$$\frac{d^2E}{dx^2} = -4\pi\rho \qquad (227 \text{ esu})$$

The *energy equation:*

$$Ee = \tfrac{1}{2}m_ev^2 \qquad (228 \text{ esu})$$

also written

$$v = \sqrt{\frac{2Ee}{m_e}} \qquad (229 \text{ esu})$$

describes an electron's velocity at any point for which the potential is
E; this velocity evidently increases during flight.

The *equation of flow*

$$J = \rho v \qquad (230)$$

relates current density J of space-charge flow (statamperes per square
centimeter), to space-charge density and velocity. There is an analogous
equality between air flow rate and the product of air density by velocity.
The equation of flow signifies that *for fixed current flow the space-charge
density must vary inversely as the velocity*, the comparison usually being
made from point to point along the path of flow.

97

Imagine the electron flow to be initiated by opening an electron gate at the cathode; there must exist at first a current *transient*; the current initially increases with time, but at a decreasing rate. Eventually stability is reached, the current and space-charge densities no longer changing.

Since space-charge density no longer changes as time passes, electrons must be leaving any selected volume element, traveling toward the right, at just the same rate that others are entering from the left; hence there can be nowhere any change from point to point in current density. In this equilibrium state *the current density J is the same for all values of x,* for much the same reason that the current must be the same in all parts of a closed electric circuit.

But v is not the same for all values of x, and, according to Equation (230), ρ must vary inversely with v. Hence in the equilibrium state the equation of flow requires a *lower* space-charge density in the *high-velocity region* near the anode than in low-velocity, low-potential regions.

The statement of the equation of flow for transient conditions is

$$J = \rho v + \frac{1}{4\pi} \frac{dF}{dt} \qquad (231 \text{ esu})$$

where J is now total current density, not merely current density due to space-charge flow. The first term, called the conduction current density, represents the movement of space charge, while the second, called the *displacement current density,*[A 290, B XVII] represents changes in flux density that grow out of redistribution of charge density between the plates. In a one-dimensional region the J of Equation (231) must be the same for all values of x, for both transient and equilibrium conditions. Under equilibrium conditions, however, the displacement current vanishes, making the conduction current alone invariant with x.*

49. Zero Gradient at the Cathode: a Condition for Maximum Space Charge Consistent with Steady Current Flow. If the cathode is a very plentiful source of electron supply, the intermediate space charge grows very rapidly (Fig. 42a) after an imaginary " electron gate " is first opened at the cathode. In the presence of the resulting negative space charge the potential line between the electrodes must be convex downward, and the greater the density of space charge, the greater the flexion must be.

---

* See papers and references given by Llewellyn[131] for analysis of vacuum tube operation with frequencies high enough so that the transit time of the electrons between cathode and plate is an important fraction of the cyclic period of the circuit frequency. Under these conditions the equilibrium condition is never reached, and the displacement term in Equation (231) becomes important.

It is possible to imagine enough space charge existing between the plates to force the potential line to dip below the cathode potential, as in Fig. 42c, so producing a negative potential gradient just outside the cathode surface. But even a slight dip of this nature shuts off the emergence of electrons from the cathode, for the cathode surface merely releases electrons. It 'does not to any appreciable extent eject them, and any electron released into a negative gradient experiences a force which pushes it back into the cathode. So space charge large enough in amount to produce an appreciable negative dip outside the cathode is not consistent with steady current flow between the plates. (This statement will require modification; see Section 97.)
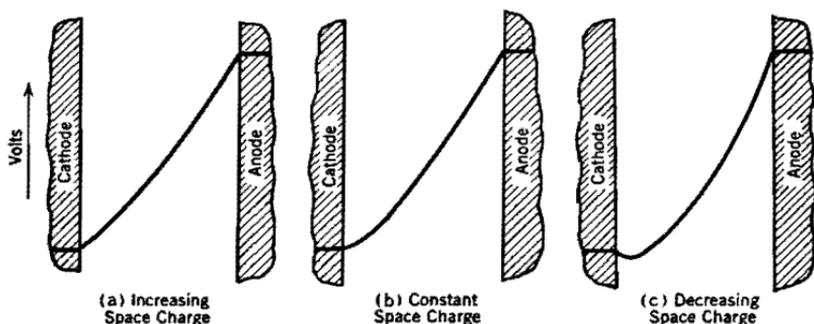
Fig. 42.   Zero gradient at the cathode on equilibrium requirement for steady space-charge flow with plentiful electron supply at the cathode.

Thus if the cathode can supply electrons at a practically unlimited rate, the space-charge density and flexion of the potential line must grow until the potential gradient at the cathode vanishes (Fig. 42b). They cannot grow beyond that point, because any further increase of space charge shuts off electron emergence from the cathode altogether. Zero gradient at the cathode is therefore the condition normally reached if the cathode supplies electrons plentifully, and corresponds to the maximum space charge consistent with steady current flow.

Since current flow in such a region is due to the movement of space charge, this condition for maximum space charge is also the condition for maximum steady current flow between the electrodes at a given potential difference. No matter how profuse the electron supply at the cathode may be, *the energy, flow, and Poisson relations in combination establish, for given geometry and electrode potentials, a definite upper limit to the steady current that can be carried by space-charge movement.*

50. Space-Charge-Limited Current Proportional to the Three-Halves Power of the Voltage. The maximum current which, according to the considerations of the preceding section, can be carried by space-

charge movement between a pair of electrodes is commonly called the *space-charge-limited current*. The current in most thermionic vacuum tubes is normally space-charge-limited, the filaments being capable when hot enough of supplying many more electrons then actually flow. The magnitude of the space-charge-limited current that actually does flow in an individual tube is controlled by electrode geometry and potentials, not by filament temperature.

Under these usual conditions, typified by Fig. 42b, there is no surface charge on the cathode, while the positive surface charge on the anode is just equal to the total negative space charge. All the negative charge lies between the plates. Now Poisson's equation indicates a general *proportionality* between potential and space charge if the integration constants are zero, as they are when, as here, the potential and potential gradient are zero at zero distance from the cathode. So it is to be expected, and is in fact true, that *the total space charge under equilibrium conditions is proportional to the total voltage, for given geometry.*

The *shape* of the potential line for the equilibrium space-charge-limited state is the same for any overall voltage, the potential at a given distance from the cathode being a definite percentage of the total voltage, regardless of what the total may be. Electron velocity is proportional to the square root of the potential, hence *for a large overall voltage, the velocity of space-charge movement at a given location must be large, but in proportion to the **square root** of the total voltage.*

The space-charge content being proportional to the first power of the voltage, and the velocity of its movement to the square root of the voltage, the current, being the product of space-charge density and velocity, should and does vary as the *three-halves power* of the voltage. This reasoning is equally applicable to the current between concentric cylinders, and, in fact, to that in any geometrical arrangement in a high vacuum in which the electric flux lines and stream-lines of current flow are coincident and straight. It is nearly true for many geometries that seem only very roughly to approximate this requirement.

The proportionality factor depends on both the general form and detail dimensions of the geometry; it is not the same for concentric cylinders as for parallel plates.

51. Space-Charge-Limited Volt-Ampere Relation, Parallel Plane Electrodes. Poisson's, the energy, and the flow equations in combination give for one-dimensional geometry the differential relation

$$\frac{d^2E}{dx^2} = 4\pi J \sqrt{\frac{m_e}{2e}} \frac{1}{\sqrt{E}} \qquad (232 \text{ esu})$$

The negative sign of Poisson's equation has disappeared by virtue of

ρ and therefore J being negative. The direction of motion of electrons is positive, that is, for each electron x becomes progressively larger, but this corresponds to a current flow in the opposite direction, for the carriers are negative.

This equation is solved by using a mathematical trick that is employed in handling many other differential equations that are encountered in space-charge-flow problems. Both sides of the equation are multiplied by dE/dx; then it is recognized that, on the left

$$\frac{dE}{dx}\frac{d^2E}{dx^2} = \frac{1}{2}\frac{d}{dx}\left(\frac{dE}{dx}\right)^2 \qquad (233 \text{ esu})$$

and that on the right

$$\frac{1}{\sqrt{E}}\frac{dE}{dx} = 2\frac{d\sqrt{E}}{dx} \qquad (234 \text{ esu})$$

so that the first integration results in

$$\left(\frac{dE}{dx}\right)^2 = 16\pi J\sqrt{\frac{m_e}{2e}}\sqrt{E} \qquad (235 \text{ esu})$$

The second integration is straightforward after taking the square roots of both sides. The integration constant for the first integration, and the lower limit for the second, are zero because both potential and potential gradient are zero when $x = 0$. The second integration has the upper limit $x = s$, the spacing of the plates.

The final result when solved for J gives the following expression for space-charge-limited current density between parallel plane electrodes:[57, 58, 74,]

$$J = \frac{\sqrt{2}}{9\pi}\sqrt{\frac{e}{m_e}}\frac{E_p^{\frac{3}{2}}}{s^2} \qquad (236 \text{ esu})$$

In practical units, and for either electrons or ions,

$$J = \frac{2.331 \times 10^{-6}}{\sqrt{m_g/m_e}}\frac{E_p^{\frac{3}{2}}}{s^2} \qquad (237 \text{ p})$$

E_p is the voltage between the plates, m_g the mass of an ion. The inclusion of the mass-ratio factor in Equation (237) makes it applicable to space-charge-limited currents made up of moving ions. For electrons the mass ratio is of course unity.

No matter how many electrons a cathode may be able to supply, geometry and potential do not permit the current to exceed the value given by Equations (236) and (237). Of course, it may have a less value, if the electron supply from the cathode is restricted.

The second integration can equally well be made between $x = 0$ and

$x = x$, in order to describe E at any point between fixed plates. Making this change, and solving the corresponding equivalent of Equation (236) for E in terms of J and x, there results the potential distribution equation

$$E = 3\left(\frac{3\pi^2 m_e J^2}{2e}\right)^{\frac{1}{3}} x^{\frac{4}{3}} \qquad \text{(238 esu)}$$

Fig. 42b has been drawn in accordance with this equation.

Fig. 42a illustrates the type of curve that is obtained from a solution of Equation (232) in case J is limited, by low cathode temperature, to some value less than the space-charge-limited current. In that case the first integration constant is not zero, for there is a potential gradient just outside the cathode, and a surface charge on it.

52. Space-Charge-Limited Current in a Parallel-Plane Triode.* The possibility of grid control of the space-charge-limited current in a triode arises out of the fact that such current is dependent in a definite, calculable way on the potential gradient that would exist at the cathode if no electrons were present in the interelectrode space, that is, on the space-charge-free off-cathode gradient $-F_0$. In a parallel-plane triode like that of Fig. 42 the space-charge-free gradient is of course

$$-F_0 = \frac{E_p}{s} \qquad (239)$$

By the use of this relation Equation (236) can be changed into the form

$$J = \frac{\sqrt{2}}{9\pi} \sqrt{\frac{e}{m_e}} \frac{(-F_0)^{\frac{3}{2}}}{\sqrt{s}} \qquad \text{(240 esu)}$$

or

$$J = M \frac{(-F_0)^{\frac{3}{2}}}{\sqrt{s}} \qquad \text{(241 esu)}$$

The meaning of M is obvious. In practical units Equation (241) becomes

$$J = 2.331 \times 10^{-6} \frac{(-F_0)^{\frac{3}{2}}}{\sqrt{s}} \qquad \text{(242 p)}$$

F_0 must itself be negative in order to permit any electron emergence.

In order to apply the relation described by these equations to a parallel-plane triode, $-F_0$ must be determinable, and a value to be used for the spacing s selected. The proper value to use for $-F_0$ is of

---

* The analysis here presented of the manner of dependence of plate current on triode geometry and potentials is very different from the one given by R. W. King[59] at a much earlier stage in the development of electronic science. However, the theoretical formulas on which Kusunose's work[8] in 1929 was based are close approximations to those derived here.

course that existing between the cathode and imaginary equivalent diode of Fig. 20, its value being

$$-F_0 = \frac{E_g + \dfrac{E_p}{\mu}}{d_1} \tag{243}$$

where d_1 is determinable from the reasoning presented in Section 19. The correct value for s must certainly be less than the distance d_1, for the uniform gradient corresponding to F_0 does not exist that far out in the actual triode; at about the grid location the potential line bends sharply upward, under the usual condition of positive plate and negative grid.

53. Magnitude of the Spacing-Factor s, Parallel-Plane Triode. A study of Fig. 43 is helpful in arriving at an understanding of the underlying meaning of the spacing-factor s. In this figure three different plate spacings, s_1, s_2, s_3, are compared at a common value of F_0, there being no grid. The three potential distribution curves are similar to
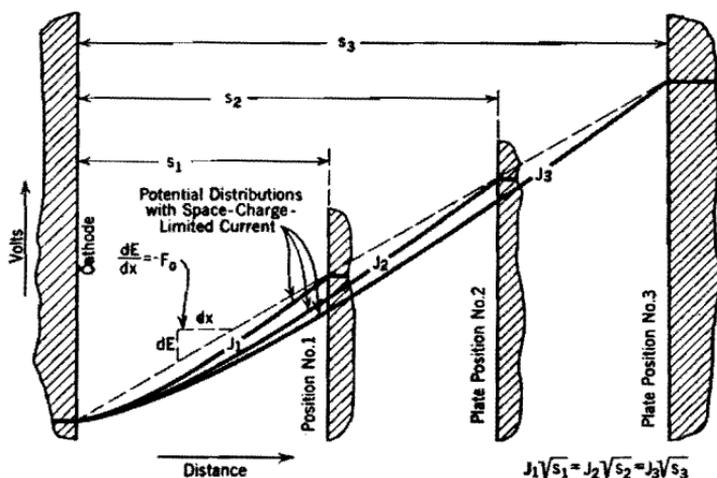
FIG. 43.   Effect on space-charge-limited current of varying the spacing between two parallel plane electrodes, keeping space-charge-free gradient constant.

one another, as shown in the figure, but intersect the F_0 line at different points. The current densities vary inversely as the square roots of the spacings, in accordance with Equations (240, 241, 242). Evidently the value of J is completely determined when (1) F_0 is stated, and (2) the distance is specified at which intersection occurs between the F_0 line and a $\frac{4}{3}$-power potential distribution curve [see Equation (238)].

This must be true in a parallel-plane triode as well as between cathode and lone plate, for regardless of what happens farther out, the actual potential line near the cathode, where the field is one-dimensional, must

be a part of some one $\frac{4}{3}$-power curve that is aimed toward some definite point of intersection with the F_0 line. The local situation must be of this same general type whether the electron acceleration is due to the effect of a lone plate or of a plate and grid acting jointly.

The accurate mathematical determination of the proper $\frac{4}{3}$-power curve and of the manner in which it merges into a similar but much steeper curve beyond the grid is a little difficult, but a close graphical estimate is usu-ally possible. The solid between-grid potential dis-tribution line shown in Fig. 44 is the result of such a graphical estimate, as applied to space-charge-free conditions identical with those of Fig. 20. The process of locating it may be illustrated by reference to Fig. 45. The first step is the placement of a family of $\frac{4}{3}$-power curves, like those of Fig. 43, with their common origin at zero volts on the cathode surface. It is then neces-sary to select the one

FIG. 44. Space-charge-flow potential distribution in a parallel-plane triode.

$\frac{4}{3}$-power curve that provides the best all-round match with conditions at and beyond the grid, and to estimate the nature of the departure, near to and beyond the grid, from the initial $\frac{4}{3}$-power form.

All the deviations of the true potential curve from the dotted space-charge-free potential line are of course due to the presence of space charge, which is related to potential according to the two-dimensional Poisson equation:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} = -4\pi\rho \qquad (244 \text{ esu})$$

The upper dotted line in Fig. 45 describes the variation of E with x along a path between grid wires when ρ is zero, that is, when $\partial^2 E/\partial x^2 = -\partial^2 E/\partial y^2$. The introduction of a small amount of space charge makes $\partial^2 E/\partial x^2$ more positive, and $\partial^2 E/\partial y^2$ less negative, than without space charge. Therefore to fit space charge-flow conditions properly a po-tential curve must have at each value of x a flexion greater than that of

the space-charge-free line, about in proportion to the local negative space-charge density.

Of course space-charge density decreases rapidly as potential and velocity grow ($J = \rho v$). Space charge density must be much less in the neighborhood of the grid than near the cathode, and must become smaller still as the electrons speed toward the plate. A low $\frac{4}{3}$-power curve, like No. 3 in Fig. 45, must, near to or just beyond the grid, bend upward considerably more sharply than the dotted curve in order to reach the plate at the correct potential. The corresponding increase of its flexion over that of the dotted curve indicates a much more pronounced negative space charge than can exist except close to the cathode; the choice is evidently wrong. A high $\frac{4}{3}$-power curve like No. 1 must either bend downward, or cross the dotted space-charge-free line before the grid region is reached, in order to reach the plate at the correct potential. A downward bend along the path between grid wires makes its

FIG. 45. Trial and error method of selecting true space-charge-flow potential distribution curve in a parallel plane triode.

flexion negative, so less than that of the dotted line, calling for a positive space-charge density, which is absurd. The other alternative, crossing, only masks the same kind of absurdity, for once across, it must bend considerably less sharply than the dotted line in order to reach the plate at the correct potential, and such a less flexion also calls for positive space-charge density.

Only a curve like No. 2, that lies very close to the space-charge-free line in regions near to and beyond the grid, requires neither positive nor excessive negative space charge to justify its shape. The increase in the flexion of the true solid curve over that of the dotted one must become less as the potential rises above zero, the change occurring approximately inversely as the square root of the potential growth.

The chosen $\frac{4}{3}$-power line, if mathematically prolonged according to its initial form, must intersect the straight F_0 line, as at S in Fig. 44. The

distance from the cathode to this intersection is the quantity s in the denominator of Equations (240–241–242), so that the value of s is determinable as soon as the proper $\frac{4}{3}$-power curve is selected.

The most important result of this discussion of the magnitude of the spacing-factor s is the determination of the fact, illustrated by the location of the point S in Fig. 44, that s is approximately the distance to the region of sharpest bend in the between-grid space-potential line, so usually a little less than the cathode-to-grid spacing. s may be expected for a given tube to be *nearly but not exactly the same for various combinations of grid and plate voltages*, within the normal operating range. The fact that s occurs to the one-half power in the expressions for J minimizes the importance of the small changes in its value that result from variations in tube voltages.

Knowledge of the magnitude of s is chiefly of value in comparing current expectations for various geometries. In particular, a small valve of s results in a low *plate resistance*, defined in Chapter XII. For comparisons of the type suggested the use of the distance from cathode to grid as an approximation to s is often justified. Thus the smaller the cathode-to-grid spacing, the less the plate resistance.*

Of course parallel-plane triodes with actual plane-surfaced cathodes are not used, yet this discussion of them is of more than theoretical interest, because the principles it illustrates carry over with proper modifications into the consideration of cylindrical triodes, and of triodes with filamentary cathodes and parallel-plane grid-plate structures.

54. Space-Charge-Limited Volt-Ampere Relationship, Concentric Cylinders. For space-charge flow between concentric cylinders the three basic equations can, for the sake of convenience, be written in the following forms:

$$\frac{d^2 E}{(d \log r)^2} = -2\rho' r \quad \text{(Poisson's equation; see Section 9)} \quad \text{(245 esu)}$$

$$Ee = \tfrac{1}{2}m_e v^2 \quad \text{(the energy equation)} \quad \text{(246 esu)}$$

$$J' = \rho' v \quad \text{(the equation of flow)} \quad \text{(247)}$$

* If it is assumed that Equation (111) reduces to approximately $d_1 = a$, (Fig. 20) because μ is ordinarily considerably greater than 1, and that the spacing-factor s is also equal to a, the following approximate expression for J in a parallel-plane triode is obtained:

$$J = 2.331 \times 10^{-6} \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{a^2} \quad \text{(244.1 } p)$$

The *plate resistance* R_p of a triode is defined in Section 133 as $\partial E_p / \partial I_p$ (grid voltage held constant, E_p and I_p varied). In general, any change in dimensions that increases I_p relative to E_p will decrease R_p. Decreases in d_1 and in s are changes that have such an effect. The cathode-to-grid spacing is the most important part of d_1, also of s. Therefore the cathode-to-grid spacing must be small if R_p is to be small.

Here J' is the current flow per centimeter of cylinder length, and $\rho' = 2\pi r\rho$, the space-charge density per centimeter length integrated around the axis at radius r. A cylindrical shell of radius r, thickness dr, and unit height, contains $\rho'\, dr$ statcoulombs of charge, as illustrated in Fig. 46. Solution of the differential equation into which these may be combined

Ring of Space Charge Density ρ

dr

1 cm.

Total Space Charge in Ring $= 2\pi r\rho\, dr$
$= \rho' dr$
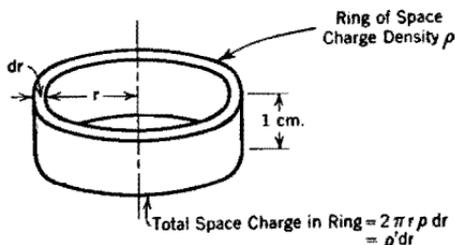
FIG. 46. Relation between ρ', r, and ρ.

leads, if the inner cylinder is the cathode, to an approximate form very similar to Equation (236), as follows: [60, 61, 74]

$$J' \cong 2\pi r_s \left[\frac{\sqrt{2}}{9\pi} \sqrt{\frac{e}{m_e}} \frac{E_p^{\frac{3}{2}}}{r_s^2} \right] = \frac{2\sqrt{2}}{9} \sqrt{\frac{e}{m_e}} \frac{E_p^{\frac{3}{2}}}{r_s} \qquad (248 \text{ esu})$$

or, in practical units, and for either electrons or ions,

$$J' \cong 2\pi \frac{2.331 \times 10^{-6}}{\sqrt{m_g/m_e}} \frac{E_p^{\frac{3}{2}}}{r_s} \qquad (249 \text{ p})$$

r_s being the radius of the outer cylinder. These expressions are reasonably accurate if r_s is ten or more times the radius r_c of the inner cylinder. The exact expression is

$$J' = 2\pi \frac{2.331 \times 10^{-6}}{\sqrt{m_g/m_e}} \frac{E_p^{\frac{3}{2}}}{r_s\beta^2} \qquad (249.1 \text{ p})$$

in which β^2 is a known function of r_s/r_c. Values for β^2, derived as described in the following paragraphs, appear in Table I and Fig. 47.

The *current density* J_c, at radius r_c, can be obtained by dividing Equation (249.1) by $2\pi r_c$, with the following result:

$$J_c = \frac{2.331 \times 10^{-6}}{\sqrt{m_g/m_e}} \frac{E_p^{\frac{3}{2}}}{r_c r_s \beta^2} \qquad (250 \text{ p})$$

Equation (250) is interesting because of its similarity to Equation (237).

A differential equation whose solution leads to the relations just stated is obtained by combining Equations (245), (246), and (247), and is

$$\frac{d^2E}{(d\log r)^2} = 2rJ' \sqrt{\frac{m_e}{2e}} \frac{1}{\sqrt{E}} \qquad (251 \text{ esu})$$

By making the substitution $r = r_c \epsilon^u$, which may also be expressed as $u = \log r/r_c$, and letting $M = 2J'\sqrt{m_e/2e}$, where r_c is the radius of the

inner (cathode) cylinder, Equation (251) is converted into

$$\frac{d^2E}{du^2} = \frac{Mr_c\epsilon^u}{\sqrt{E}} \qquad (252 \text{ esu})$$

If E is eliminated in favor of a new variable β, defined by the relation

$$E^{\frac{3}{2}} = \tfrac{9}{4} Mr_c\epsilon^u\beta^2 \qquad (253 \text{ esu})$$
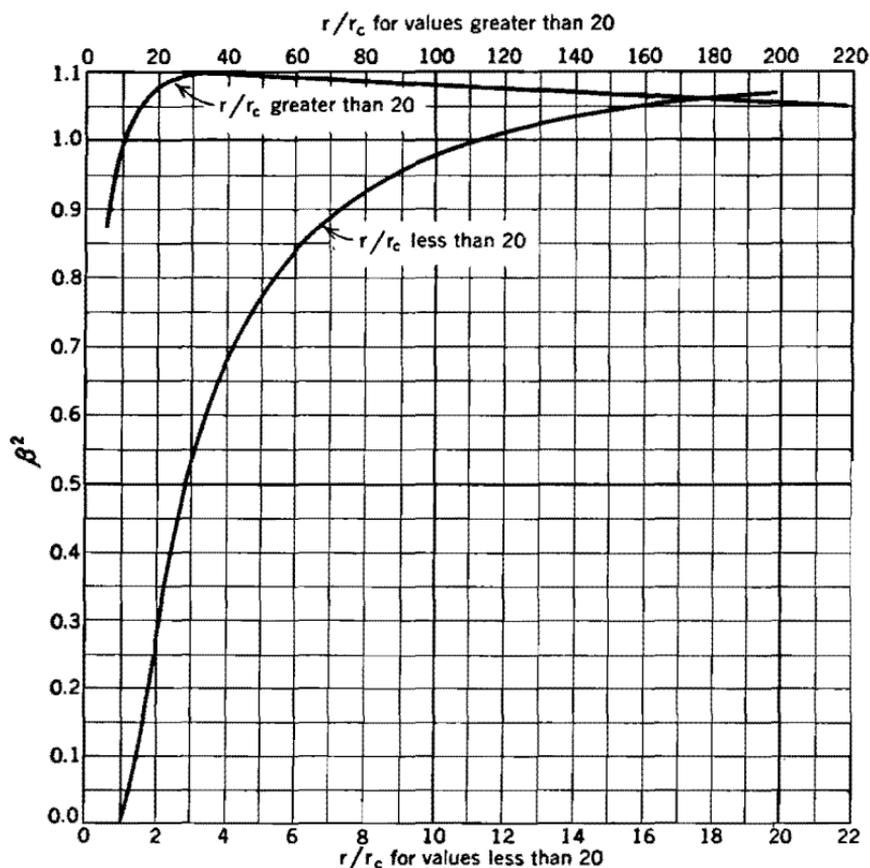
FIG. 47.   Variation of β^2 with r/r_c, in the space-charge-limited current equations for cylindrical geometry.   See Table I, and Equations (249) to (264).

a differential equation between β and u is obtained, having the form

$$3\beta\frac{d^2\beta}{du^2} + \left(\frac{d\beta}{du}\right) + 4\beta\frac{d\beta}{du} + \beta^2 - 1 = 0 \qquad (254 \text{ esu})$$

β can be evaluated in terms of u by assuming a power series, as follows

$$\beta = C_0 + C_1u + C_2u^2 + C_3u^3 + \cdots \qquad (255)$$

then evaluating the derivatives of β in terms of u and the C's, entering

them in Equation (254) and equating to zero the coefficients of each power of u. The value of C_0 is zero, because $E = 0$ when $u = 0$. The solution for β, obtained by Langmuir,[60, 61, 62, 74] is

$$\beta = u - \tfrac{2}{5}u^2 + \tfrac{11}{120}u^3 - \tfrac{47}{3300}u^4 + \cdots \qquad (256)$$

The useful form of the final solution is obtained by putting into Equation (253) the original meaning of M and solving for J'. The following expression results:

$$J' = \frac{2\sqrt{2}}{9}\sqrt{\frac{e}{m_e}}\frac{E_p^{\frac{3}{2}}}{r_s\beta^2} \qquad (257 \text{ esu})$$

Equation (257) becomes Equation (248) if $\beta^2 = 1$. Values for β^2 may be found from Table I and Fig. 47. It will be noted that, for all values of r_s/r_c greater than ten, β^2 is within ten per cent of unity. Therefore Equations (248), (249), and (250) are approximately true when $r_s/r_c > 10$. The following expression, similar in form to Equation (249), is valid for all values of r_s/r_c:

$$J' = 2\pi \frac{2.331 \times 10^{-6}}{\sqrt{m_g/m_e}}\frac{E_p^{\frac{3}{2}}}{r_s\beta^2} \qquad (257.1 \text{ p})$$

The equation for the potential distribution between concentric cylinders passing a space-charge-limited current is obtained directly from Equation (253), and is

$$E = (\tfrac{9}{4}M)^{\frac{2}{3}}(r\beta^2)^{\frac{2}{3}} \qquad (258 \text{ esu})$$

or, in practical units

$$E = \left(\frac{J'}{2\pi \times 2.331 \times 10^{-6}}\right)^{\frac{2}{3}}(r\beta^2)^{\frac{2}{3}} \qquad (258.1 \text{ esu})$$

55. Space-Charge-Limited Current in a Cylindrical Triode. According to Equation (257.1) the general equation for space-charge-limited electron flow between concentric cylinders, no grid, outer and inner radii r_s and r_c, is

$$J' = 2\pi \times 2.331 \times 10^{-6}\frac{E_p^{\frac{3}{2}}}{r_s\beta^2} \qquad (259 \text{ p})$$

Here β^2 is as before a function of r_s/r_c, having values as given in Table I. The space-charge-free off-cathode gradient in this geometry [(compare Equation (138)] is

$$-F_0 = \frac{E_p}{r_c\log\dfrac{r_s}{r_c}} \qquad (260)$$

Therefore, in terms of F_0, the current per unit length of axis is

$$J' = 2\pi \times 2.331 \times 10^{-6} \frac{\left(-F_0 r_c \log \frac{r_s}{r_c}\right)^{\frac{3}{2}}}{r_s \beta^2} \qquad (261\ \mathrm{p})$$

The application to a cylindrical triode is made by expressing F_0 in terms of (1) an equivalent voltage, and (2) the radius r_1 to the plate of a space-charge-free equivalent diode (see Section 22), as follows,

$$-F_0 = \frac{E_g + \dfrac{E_p}{u}}{r_c \log \dfrac{r_1}{r_c}} \qquad (262)$$

Equation (262) may be substituted in Equation (261), giving

$$J' = 2\pi \times 2.331 \times 10^{-6} \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{r_s \beta^2 \left[\dfrac{\log \dfrac{r_1}{r_c}}{\log \dfrac{r_s}{r_c}}\right]^{\frac{3}{2}}} \qquad (263)$$

β^2 is still a function of r_s/r_c.   The radius r_s in Equation (263) is analogous to the spacing-factor s in the parallel-plane triode discussion (Section 53), and r_1 to the distance d_1.   The true potential line, if prolonged indefinitely in accordance with Equation (258.1), must intersect the space-charge-free equivalent diode potential line at a radius r_s; see Fig. 48.

Equation (263) can be rewritten to express J_c, current density *at the cathode of the triode* by dividing by $2\pi r_c$, with the following result:

$$J_c = 2.331 \times 10^{-6} \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{r_c r_s \beta^2 \left[\dfrac{\log \dfrac{r_1}{r_c}}{\log \dfrac{r_s}{r_c}}\right]^{\frac{3}{2}}} \qquad (264)$$

Equation (264) should be compared with Equation (250), and with the expression obtained by combining Equations (242) and (243).

The sharpness of the bend in the between-grid potential line is if anything accentuated by the change from parallel-plane to cylindrical geometry, as a comparison of Figs. 20 and 26 will show.   See also Fig. 48. It is therefore reasonable to expect, as before, the correct value of r_s to be about the distance to the point of sharpest bend in the space-charge-free potential line (compare Section 53).

Actual estimate of the position of the true space-charge-limited potential line is made by graphical selection from among a family of curves that start from the cathode in accordance with Equation (258.1).   The principles used in making the selection are essentially the same as those discussed in connection with Figs. 44 and 45.   Detailed comparison
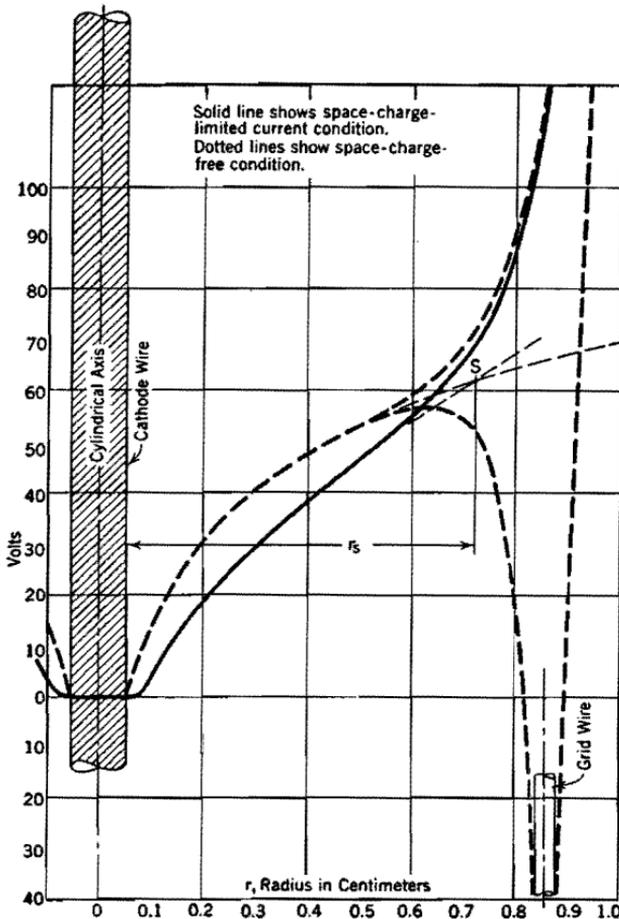
FIG. 48.   Space-charge-flow potential distribution in a cylindrical triode.   The solid line describes the potential variation along a path between grid wires when space-charge-limited current is flowing.   The dotted lines describe the space-charge-free potential distribution.

of flexions does not have the simple significance that it had in Section 53, because Poisson's equation for this geometry [Equation (27) with the left-hand term in z omitted] does not contain the r-section flexion directly.

It should be borne in mind that in the arrangement of electrodes

usually employed the grid wires are wound around rather than being placed parallel to the cathode wires. The radii r_1 and r_s may however reasonably be expected to have, as μ is known to have, about the same values whether ϕ or z is the second variable, for given grid-wire radius and screening fraction. Equation (264), with r_s chosen somewhat smaller than r_g, may therefore be expected to represent the actual behavior reasonably closely.

If, as a first approximation, r_g is used for r_s in Equation (264), and $\log r_1/r_c$ eliminated by the use of Equation (141), the following expression is obtained:*

$$J' = 2\pi \times 2.331 \times 10^{-6} \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{r_g \beta^2 \left[1 + \dfrac{1}{\mu}\dfrac{\log \dfrac{r_p}{r_c}}{\log \dfrac{r_g}{r_c}}\right]^{\frac{3}{2}}} \qquad (264.1\ p)$$

Here β^2 is a function of r_g/r_c.

56. Effect of Potential Variation along the Cathode on Space-Charge-Limited Current. The equations of recent sections all indicate that the space-charge-limited current density J_c from a cathode, whether plane or cylindrical, with or without a grid, should be proportional to the $\frac{3}{2}$ power of a voltage. In a diode, the voltage to be used is simply that of the anode; in a triode, it is the equivalent voltage $E_g + \dfrac{E_p}{\mu}$. These

* Equation (264.1) may be rearranged into the following form:

$$I_p = 2.331 \times 10^{-6} A \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{r_p r_g \beta^2 \left[1 + \dfrac{1}{\mu}\dfrac{\log \dfrac{r_p}{r_c}}{\log \dfrac{r_g}{r_c}}\right]^{\frac{3}{2}}} \qquad (264.2)$$

Here I_p is total plate current, and A is *plate* surface area. Equation (264.2) is in substantial agreement with the similar relation given on page 1709 of Kusunose's paper,[8] which in complete form for a uniform-potential cylindrical triode is, in the present notation:

$$I_p = 2.331 \times 10^{-6} A \frac{\left(E_g + \dfrac{E_p}{\mu}\right)^{\frac{3}{2}}}{r_p r_g \left[1 + \dfrac{1}{\mu}\right]^{\frac{3}{2}}} \qquad (264.3)$$

As r_g/r_c is usually 10 or greater, β^2 in Equation (264.2) is not greatly different from 1. The ratio $\dfrac{\log \dfrac{r_p}{r_c}}{\log \dfrac{r_g}{r_c}}$ is usually not large; for the geometry of Figs. 26 and 48 it is 1.17.

relations can be summarized in the form

$$J_c = GE_b^{\frac{3}{2}}$$ (265)

where E_b is now either the plate or an equivalent voltage, according to the geometry used, and G is a quantity dependent on geometry in the manner indicated by whichever of the various earlier equations applies. In all cases G contains the numerical factor 2.331×10^{-6}, if practical units are employed. If J_c is everywhere the same, the total current I_p to the plate is obtained by multiplying J_c by the area of the cathode surface.

All of the derivations leading to the general relation described by Equation (265) are based on the assumption that the cathode is an equipotential surface. In many real devices the cathode is heated by the passage of an electric current, which causes a potential drop between its two extremes. Since this drop may be several volts, hence comparable in magnitude with a triode's equivalent voltage, it often produces marked variations in J_c along the cathode surface. This results in a distinct modification of the volt-ampere relationship.

The nature of the effect of potential variation along the cathode can be analyzed by assuming that Equation (265) holds for each incremental length of cathode surface, but that the voltage to be used varies from point to point.[8, Q 64, I 105] This assumption is satisfactorily near to the truth as long as the spacing between electrodes is considerably smaller than the extent of the cathode.

Suppose the voltage E_b, plate voltage if a diode, equivalent voltage if a triode, to be measured from the more negative end of the cathode; also let

l = length of cathode (also of plate).

z = distance along cathode from negative toward positive end; z varies from 0 to l.

w = width of cathode surface ($=2\pi r_c$ if cylindrical).

$-F_z$ = potential gradient along cathode due to heating current.

E_f = potential drop between $z = 0$ and $z = l$ due to heating current; of course

$$E_f = -F_z l$$

S = area of cathode surface ($=wl$).

I_p = total current to plate.

Now apply Equation (265) to an increment of length dz, width w, with current dI_p. In a diode, which of course has no grid

$$dI_p = wG[E_b - (-F_z z)]^{\frac{3}{2}} dz$$ (266)

If, as in a triode, there is a grid, so that E_b is an equivalent voltage, the grid and plate voltages change by equal amounts, so that

$$dI_p = wG\left[E_g + F_z z + \frac{E_p + F_z z}{\mu}\right]^{\frac{3}{2}} dz$$

giving

$$dI_p = wG\left[E_b + F_z\left(1 + \frac{1}{\mu}\right)z\right]^{\frac{3}{2}} dz \qquad (267)$$

For the sake of simplicity the following analysis has in most cases been stated for a diode. To make it applicable to a triode,

$$F_z\left(1 + \frac{1}{\mu}\right) \quad \text{and} \quad E_f\left(1 + \frac{1}{\mu}\right)$$

must replace F_z and E_f respectively. This is illustrated by Equations (266) and (267). Otherwise there is no change.

The total current is obtained by integration; two cases must be considered:

Case I. $E_b > E_f$ in a diode, $E_b > E_f\left(1 + \frac{1}{\mu}\right)$ in a triode; the integration extends between $z = 0$ and $z = l$, as follows for the diode

$$I_p = wG \int_{z=0}^{z=l} (E_b + F_z z)^{\frac{3}{2}} dz$$

$$= \frac{2}{5}\frac{wG}{F_z}\left[(E_b + F_z l)^{\frac{5}{2}} - E_b^{\frac{5}{2}}\right] \qquad (267.1)$$

When numerator and denominator are multiplied by l, this becomes

$$I_p = \frac{2SG}{5E_f}[E_b^{\frac{5}{2}} - (E_b - E_f)^{\frac{5}{2}}] \qquad (268)$$

which can also be written

$$I_p = \frac{2SG}{5} E_f^{\frac{3}{2}}\left[\left(\frac{E_b}{E_f}\right)^{\frac{5}{2}} - \left(\frac{E_b}{E_f} - 1\right)^{\frac{5}{2}}\right] \qquad (269)$$

Expansion of the quantity $\left(\frac{E_b}{E_f} - 1\right)^{\frac{5}{2}}$ by the binomial theorem shows that when $E_b/E_f \gg 1$, Equation (269) becomes

$$I_p = SGE_b^{\frac{3}{2}} \qquad (270)$$

as is to be expected. The infinite series that results from this expansion converges rather rapidly; so much so that when E_b is more than twice E_f the following expressions, employing only two terms of the series, are very satisfactory approximations:

For a diode

$$I_p = SGE_b^{\frac{3}{2}}\left[1 - \frac{3}{4}\frac{E_f}{E_b}\right] \tag{271}$$

For a triode

$$I_p = SGE_b^{\frac{3}{2}}\left[1 - \frac{3}{4}\left(1 + \frac{1}{\mu}\right)\frac{E_f}{E_b}\right] \tag{272}$$

Case II. $E_b < E_f$ for a diode, $E_b < E_f\left(1 + \frac{1}{\mu}\right)$ for a triode; the integration extends only to the point along the cathode at which

$$E_b - E_f \quad \text{or} \quad E_b - E_f\left(1 + \frac{1}{\mu}\right),$$

therefore F_0, become zero, for beyond there the off-cathode gradient is negative, and no current can flow from increments exposed to a negative gradient.

The integration yields, for the diode

$$I_p = \frac{2}{5}\frac{SG}{E_f}E_b^{\frac{5}{2}} \tag{273}$$

which can also be written

$$I_p = \frac{2}{5}SGE_f^{\frac{3}{2}}\left(\frac{E_b}{E_f}\right)^{\frac{5}{2}} \tag{274}$$

The following general relationship, stated for a triode, includes Cases I and II:

$$I_p = \frac{2}{5}SG\left[E_f\left(1 + \frac{1}{\mu}\right)\right]^{\frac{3}{2}}\ f\left[\frac{E_b}{E_f\left(1 + \frac{1}{\mu}\right)}\right] \tag{275}$$

For application to a diode, use 1 in place of $\left(1 + \frac{1}{\mu}\right)$.

The function $f\left[\frac{E_b}{E_f}\left(1 + \frac{1}{\mu}\right)\right]$ has the numerical values given in Table II,[3] which are in accord with Equation (269) when $E_b > E_f\left(1 + \frac{1}{\mu}\right)$, and with Equation (274) when $E_b < E_f\left(1 + \frac{1}{\mu}\right)$. The effect of contact differences of potential on I_p can be accounted for by including in the equivalent voltage a term E', defined by Equation (368), Section 102.

The two contrasting extremes are Equations (274) and (270). These equations indicate a $\frac{5}{2}$-power variation when the equivalent voltage is less than $E_f[1 + (1/\mu)]$, and a $\frac{3}{2}$ power variation when the equiva-

lent voltage is considerably larger than the filament voltage. Actual tube characteristics ordinarily extend through both ranges, and when plotted logarithmically are often found to follow quite accurately, and over a considerable range, a power curve whose exponent is between $\frac{3}{2}$ and $\frac{5}{2}$. Thus Fig. 49b shows that the Western Electric 101-F triode average characteristics follow a 2.2-power curve rather closely.

An interesting feature of Equation (275) is the indication it gives that plate current remains a function of the equivalent voltage $E_g + \dfrac{E_p}{\mu}$ even when potential variation along the cathode is taken into account.

57. Capacitance between Electrodes Carrying a Space-Charge-Limited Current. It was pointed out in Section 50 that the application of Poisson's law to a region containing a space-charge-limited current indicates a direct proportionality between total space-charge content and overall voltage. In this respect such regions behave as electrostatic condensers, the total charge of one sign being in space-charge form, that of the other residing, as in ordinary condensers, on the surface of the more positive electrode.

The capacitance between two parallel plates in the presence of a space-charge-limited current, no grid, is the ratio of the total charge Q of each sign (most easily expressed in terms of the plate charge) to the plate voltage. The total charge on a plate of area S square centimeters is $S\sigma$, where σ is surface charge density in statcoulombs per square centimeter. Therefore, using Equation (20),

$$Q = \frac{SF}{4\pi} \qquad \text{(276 esu)}$$

where F is the field strength just outside the plate. An expression for Q in terms of geometry and fundamental constants can be obtained by differentiating Equation (238) with respect to x to obtain F, then giving to x in the resulting expression the value s, spacing between the plates. Equation (238) can be used to express E_s, the value of E when $x = s$. The capacitance C is then determinable as the ratio Q/E_s. The result is

$$C = \frac{S}{3\pi s} \qquad \text{(277 esu)}$$

as compared with the value

$$C = \frac{S}{4\pi s} \qquad \text{(278 esu)}$$

in the absence of space charge.

A similar treatment of the region between two concentric cylinders in the presence of space-charge-limited current, no grid, begins with

differentiation of Equation (258) with respect to r in order to obtain an expression for F. The plate-surface charge is $l\tau$, the product of total axial length l in centimeters by the charge τ per unit length, and of course $F = 2\tau/r_p$ at plate radius r_p. The ratio of total charge to potential gives the capacitance as

$$C = \frac{l}{3\beta^2} \frac{d}{dr} (r\beta^2) \qquad (279 \text{ esu})$$

to be evaluated at $r = r_p$. If as in most cases of interest the radius of the plate is more than ten times that of the cathode, β^2 is not greatly different from unity and changes slowly; in that case

$$C = \frac{l}{3} \qquad (280 \text{ esu})$$

is a very close approximation to the true value of capacitance. The corresponding space-charge-free capacitance is

$$C = \frac{l}{2 \log\left(\dfrac{\text{outer radius}}{\text{inner radius}}\right)} \qquad (281 \text{ esu})$$

Numerical calculations from these figures show that concentric-cylinder diodes whose diameter ratio is greater than 10 *all* have a capacitance, in the presence of space-charge-limited current, that is the same as for a pair of space-charge-free cylinders whose diameter ratio is about 4.5 to 1.

All of this discussion applies directly, of course, to arrangements with but two electrodes. The introduction of grids between parallel plates or concentric cylinders may introduce important modifications as to magnitudes if not as to principle.

The energy stored in the electric field can be determined by volume integration of $F^2/8\pi$; for the parallel-plane case, again without a grid, it is $\frac{3}{5} QE$ instead of the space-charge-free value $\frac{1}{2} QE$.

58. Energy Dissipation at the Plate. A space-charge-limited current consists of electrons which fall freely from their point of origin at the cathode to the point of entry into the plate. Each one arrives with an amount of energy measured, in electron volts, by the potential difference between the electrodes. Upon striking the plate this entire amount of energy is converted into heat, just as the kinetic energy of a falling pebble is converted into heat when it strikes the ground.

The total rate of energy conversion is the product of the energy brought in per electron by the number of electrons arriving per second. The rate of electron arrival is measured by the current I_p between the

electrodes, and the energy of each by their potential difference E_p, so that the rate of generation of heat at the surface of the plate is E_pI_p, volts times amperes, giving watts.

The temperature of the plate must rise; as it is usually located in a vacuum the major amount of heat dissipation from it is by radiation. Radiant heat output from the plate increases as the fourth power of the temperature, so that in general only a moderate rise is enough to produce equilibrium between electrical power input and radiant heat output. Since graphite is a better radiator of heat than most metals, plates made of this material are sometimes used to permit operation at high current densities and high voltages. It is of course desirable with the same objective in view to arrange the geometry of the parts to permit direct outward heat radiation from the plate to surrounding objects. Tubes with water-cooled anodes are sometimes used in high-power radio broadcasting circuits.

PROBLEMS

Chapter V

1. (a) Assume that the potential distribution between two parallel plates, like those illustrated in Fig. 42, is described by the equation $E = Ax^2$, where A is a constant. Distance between the plates is s. Find the ratio of the quantity ρv at $x = \frac{3}{4}s$ to that at $x = \frac{1}{4}s$. Is this an equilibrium potential distribution?

(b) Same as part (a), except that the potential distribution equation is $E = Bx^{\frac{4}{3}}$, where B is a constant.

2. Imagine a hot plane-surfaced cathode, 4 sq cm in area, whose thermionic emission cannot exceed 4.3 milliamperes per sq cm at the actual operating temperature. Parallel to this cathode, but at a distance of 0.5 sq cm, there is a " plate " at a potential 50 volts higher than the cathode.

(a) What is the actual current flow between the cathode and plate when the plate potential is 50 volts above that of the cathode?

(b) When the plate potential is 100 volts higher than that of the cathode? (This problem is intended to illustrate the contrast between space-charge limitation and temperature limitation of plate current. In most actual tubes the current is space-charge-limited rather than temperature-limited.)

3. A tungsten wire, 0.02 cm in diameter and 5 cm long, is to serve as the filament (cathode) of a diode (two-electrode tube). The plate is to be a cylinder concentric with the filament.

(a) Select the radius of the plate so that the space-charge-limited current will be 40 milliamperes per cm of axis when the plate potential is 30 volts above that of the cathode. Compare the results obtained by using Equation (249) with that obtained by using Equation (257.1). (Use of this latter, more exact, expression requires a trial and error solution.)

(b) How much power, in watts, must the plate radiate?

(c) Find the potential at a point whose radius is twice that of the cathode.

(d) Find the velocity of travel of the electrons at that radius.

(e) Find the values of ρ, and of ρ' as defined in Section 54, at that radius.

4. Consider a cylindrical diode in which *the outer cylinder is the cathode*, the inner one the anode or "plate." Radius of the anode is 0.01 cm, of the cathode 0.2 cm. Anode-to-cathode potential difference is 1,000 volts.

(a) Find the space-charge-limited current per centimeter length (J').

(b) Plot points for and draw the space-charge-free potential distribution curve.

(c) Plot points for and draw the space-charge-flow potential distribution curve.

5. Combine Equations (242) and (243) into a single equation for space-charge-limited current density in a triode. Then show that the manner of occurrence of μ in the resulting expression is consistent with the definition of μ given by Equation (490) in Chapter XII. Also show that the manner of occurrence of μ in Equation (264.2) is in accordance with Equation (490).

6. The correct values of d_1 and μ, asked for in parts (d) and (e) of Problem 3, Chapter II, are 0.207 centimeter and 4.86 respectively. What is the space-charge-limited current density in the triode of that problem when $E_p = 400$ volts, $E_g = -40$ volts? Explain the nature of any uncertainties in your solution.

7. Suppose that, in the triode of Figs. 20 and 44, $E_g = E_p = +15$ volts. Sketch space-charge-free and space-charge-flow potential distribution lines, and calculate the space-charge-limited current density flowing from the cathode. Use $\mu = 4.05$, $d_1 = 0.843$ cm. Be careful to make use of the principles discussed in Section 53. Explain the nature of any uncertainties in your solution.

8. Fig. 44 is drawn for the geometry and potentials used in Fig. 20. The values of d_1 and μ for that geometry are 0.843 cm and 4.05 respectively. Calculate the space-charge-limited current density that corresponds to Fig. 44, assuming that S is correctly located in that figure. How large a per cent difference in the current prediction would result if S were assumed to lie in the plane of the grid?

9. Fig. 48 is drawn for the geometry and potentials used in Fig. 26. The values of r_1 and μ for this geometry are 1.90 cm and 4.05 respectively. Answer questions relative to this cylindrical geometry similar to those asked in Problem 8 relative to the parallel-plane geometry of Figs. 44 and 26.

10. Space-charge-limited current flows between parallel plane electrodes, no grid, as for example in position No. 1, Fig. 43. $F_0 = 600$ volts per cm, $s_1 = 0.75$ cm. How long does it take an electron to travel from cathode to plate? How long would it take a lone electron to travel from cathode to plate in the corresponding space-charge-free field? The method of solution is suggested by Equation (157).

11. (a) Estimate the current J' per centimeter length of cathode for the tube described in Problem 2, Chapter II, when the space-charge-free off-cathode gradient $-F_0$ is 100 volts per cm. Explain any uncertainties in your answer.

(b) To the tube of part (a) there is added a second plate, on the opposite side of the grid-and-cathode plane, but at the same distance from that plane as the first one. Estimate the value of J' if $-F_0 = 100$ volts per cm. Explain any uncertainties.

12. Cylindrical diode, $r_c = 0.02$ cm, $r_p = 0.10$ cm; the diode is 3 cm long, and the drop E_f along the cathode is 10 volts. Current is space-charge-limited. Find the plate current when the plate voltage relative to the negative end of the cathode is (a) 5 volts, (b) 10 volts, (c) 20 volts.

13. Calculate the plate current for the tube and potentials of Figs. 48 and 26 if its length is assumed to be 5 cm, and the drop E_f along its cathode 40 volts. Point S is located as in Fig. 48. Plate and grid potentials as in Fig. 26, measured relative to the negative end of the cathode. $r_1 = 1.90$ cm, $\mu = 4.05$.

14. Calculate the plate current for the geometry and potentials of Figs. 44 and 20

if the cathode is 5 cm wide and 10 cm long, on the assumption that there is a voltage drop E_f of 20 volts between one end and the other of the cathode. Plate and grid potentials as in Fig. 20, measured relative to the negative end of the cathode.

15. Carry through the mathematical operations that lead from Equations (252) and (253) to Equation (254).

16. Derive the differential equation, for space–charge-limited current flow between concentric spheres, that corresponds to Equation (251) for concentric cylinders. Then convert to an expression similar to Equation (254) by proper choice of substitutions generally similar to but not the same as those used in deriving Equation (254). (See reference No. 74 in Bibliography.)

17. If the electron emission from the cathode is limited by low cathode temperature to some fixed *temperature-limited* current density J_{th}, the potential distribution between parallel-plane electrodes is as illustrated in Fig. 42(a). Here dE/dx is not zero when $x = 0$, so that an integration constant A must be added on the right-hand side of Equation (235). Derive an expression relating E and x for the case of a temperature-limited current between parallel-plane electrodes. It will be found convenient to use the substitution $\sqrt{E} = u$, where u is a new variable.

CHAPTER VI

TRIODES, TETRODES, PENTODES

59. Grids Permit Electrostatic Control of Space-Charge-Limited Triode Current. In the great majority of thermionic vacuum tubes the electron current that flows between filament (cathode) and plate (anode) is a space-charge-limited current; that is, its magnitude is determined by geometry and potentials within the tube. The filament temperature is kept high enough to supply much more current, as far as the number of electrons available is concerned, than actually flows.

By modifying the potential applied to a grid located between cathode and plate the equivalent voltage $E_g + \dfrac{E_p}{\mu}$ can be changed, and so the plate current modified, in accordance with the space-charge-limited current analysis of Chapter V. Since the grid is usually maintained at a potential negative relative to the cathode, electrons do not reach it; it is not an emitter, so cannot release electrons. Carrying no current in either direction, its potential, and therefore the tube current, can be controlled with the expenditure of extremely minute amounts of power in the controlling circuit. The control of plate current so provided is essentially electrostatic in nature, in that it is accomplished entirely by means of modification of the electrostatic field within the tube, the only current demanded of the controlling circuit being the " charging current " to the grid, dependent in magnitude upon interelectrode capacitances.

60. Current-Voltage Relations in Triodes. For the reasons summarized in Section 56, the total space-charge-limited current of a triode should be expected to vary as the $\frac{3}{2}$ power, or as the $\frac{5}{2}$ power, of the equivalent voltage, or according to some intermediate relationship. The dependence should be different for a cathode heated directly by passage of current through it than for one that is indirectly heated and therefore at uniform potential. A tube of the latter type may be expected to follow a $\frac{3}{2}$-power law. The small effect of *contact differences of potential* between electrodes on the equivalent voltage is discussed in Section 102.

It is found experimentally that plate currents of many triodes can be satisfactorily expressed in terms of grid and plate voltages by an empiri-

cal expression of the following form:

$$I_p = B\left(E_g + \frac{E_p}{\mu}\right)^n \tag{282 p}$$

If this can be done, the current is actually — as expected — dependent on an equivalent voltage. The amplification factor μ in such empirical equations is found to agree closely with that predicted from dimensions of grid and plate structures by the methods of Chapter II. The actual value of the exponent for any particular device can be found by making use of the logarithmic form of Equation (282), as follows:

$$\log I_p = \log B + n \log\left(E_g + \frac{E_p}{\mu}\right) \tag{283 p}$$

Equation (283) has the general form

$$y = b + nx \tag{284}$$

Here n and $\log B$ are respectively the slope and I_p axis intercept of a logarithmic graph like that illustrated in Fig. 49b.
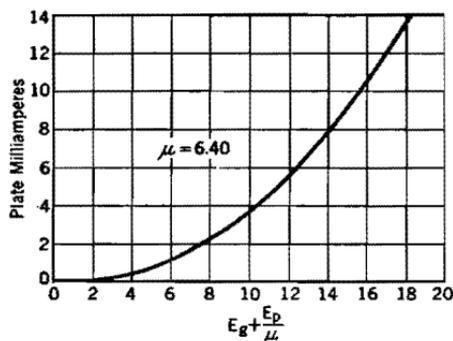

FIG. 49a. Dependence of triode plate current on equivalent voltage, Western Electric type 101-F triode. The curve in this figure represents any of those in Fig. 3b.
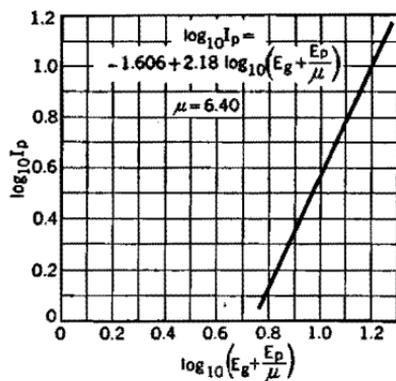
FIG. 49b. Logarithmic relation between plate current and equivalent voltage, Western Electric type 101-F triode. The straight line shown here is a logarithmic graph of the curve shown in Fig. 49a.

To the extent that the different *plate characteristic curves*, illustrated in Figs. 3b and 50b, are parallel to one another, that is, identical except for a horizontal displacement, the plate current is dependent on an equivalent voltage. To the extent that they are not so identical, the plate current is not dependent] purely on the equivalent voltage. Fig. 50b illustrates a kind of deviation from dependence on equivalent voltage that is typical in lesser degree of most actual triodes: the high-plate-

voltage characteristics show progressively gentler slopes, particularly near cut-off.

To understand the reason for this, it must be remembered that the analysis predicting dependence on an equivalent voltage presupposes a space-charge-free potential distribution of the general type illustrated by Figs. 20 and 26, in which the potential lines between grid wires merge with those through grid wire centers *before the cathode surface is reached.* This makes the space-charge-free off-cathode gradient everywhere uniform, and its value directly proportional to the equivalent voltage. However, as the plate voltage rises and the grid voltage falls, there arises a tendency for the two potential lines not to merge, but to enter the cathode surface at different angles. F_0 is then no longer uniform, and the degree of nonuniformity is dependent on the *spread*

(a) Mutual characteristics.   (b) Plate characteristics.
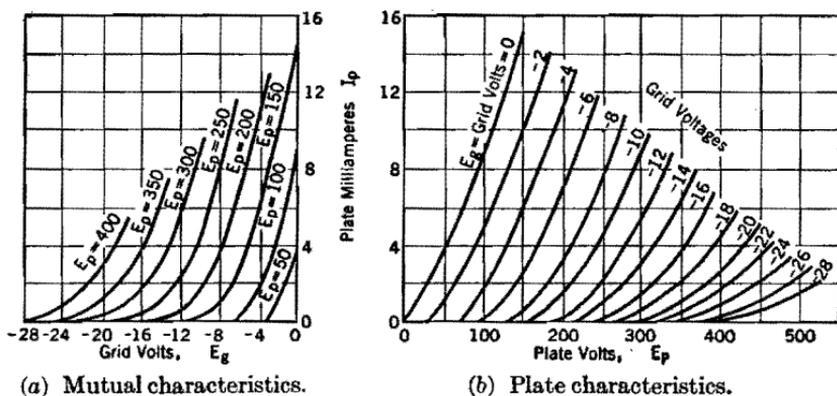
FIG. 50.   Average triode characteristic curves, for the RCA Type 6C5 cylindrical triode (metal envelope).

between plate and grid potentials.   One result of this is that F_0 *for the between-grid-wires path* does not become zero until the grid potential is substantially below the value necessary to make $E_g + \dfrac{E_p}{\mu}$ zero, and of course cut-off cannot occur until the off-cathode space-charge-free gradient is *everywhere* zero or negative.

Other variations, within the normal operating range, from the predicted form and from expected values of the exponent, result partly from asymmetries and irregularities in interelectrode geometry and in paths of electron travel, partly from the effects of small initial velocities which the electrons have on leaving the cathode, and partly from other peculiarities of electron-emitting surfaces.   Chaffee gives an interesting discussion of these variations.[E VII]

The analysis of electric circuits containing tubes is commonly based

on a graphical representation of electrical tube characteristics. Figs. 50a and 50b are such graphical representations for an RCA 6C5 power-amplifier metallic-envelope triode manufactured by the RCA Radiotron Company. The curves of Fig. 50b, called *plate characteristics*, show the variation of plate current with plate voltage for a variety of grid voltages; those of Fig. 50a, called *mutual characteristics*, show the variation in plate current in response to grid voltage changes, plate voltage being the parameter.

The prediction of circuit behavior is materially aided by a knowledge of the magnitudes of certain tube " constants " that are defined partly in terms of the slopes of these curves, and described by the names *plate resistance, transconductance,* and *amplification factor.* These quantities are discussed in detail in Chapter XII. In triodes the values of the first two of these " constants " are, like the slopes of the characteristic curves, dependent on I_p, but their product *is* reasonably constant and equal to the amplification factor μ.

The lower end or *cut-off point* of each of the characteristic curves of Fig. 50 is presumably associated with the condition $E_p = -\mu E_g$ in Equation (282). Actually the various theoretical equations do not represent accurately the manner of decrease to the zero axis at small current values. The average temperature energy of the electrons emerging from the cathode is proportional to the filament temperature, being usually between one- and three-tenths of an electron volt. For sufficiently small values of $E_g + \dfrac{E_p}{\mu}$ the initial electron energies have an important part in the determination of the magnitude of the current. The effects of initial electron energies are analyzed in Sections 124 and 125, Chapter XI.

61. Grid Current. Equation (282), from the nature of its development, describes the variations in total current from the cathode, while the curves of Figs. 3 and 50 relate to plate current only. As long as the grid voltage is negative, which is the normal condition for the majority of simple tube applications, there is no grid current. For positive grid voltages and low or negative plate voltages a substantial portion of the total cathode current may pass to the grid. The dotted curves in the lower left portion of Fig. 3b indicate the magnitude of the total current from the cathode there; although in this range the total current follows the space-charge-limited type of variation, it is divided between grid and plate.

Fig. 51a illustrates the general form taken by potential distribution lines for a triode when the grid voltage is positive and the plate voltage relatively low. The grid wires occupy potential peaks, into which the

streaming electrons " fall," so that most of them fail to reach the plate.
Fig. 51b is a potential section through the grid wires in a plane parallel
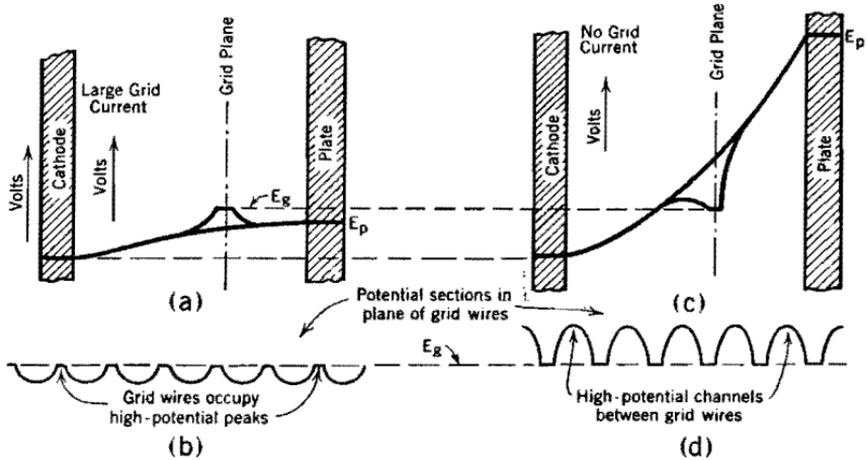
FIG. 51. Space-charge-flow potential distribution in a triode with positive grid,
indicating manner of diversion of current from positive grid to plate, at high plate
voltages. Plate voltage is low in (a) and (b), high in (c) and (d). See also Fig. 52.

to plate and to cathode. The grid wires evidently occupy potential
peaks in both directions, so that there is considerable grid current.
This corresponds to conditions as
at A, Fig. 52.

In contrast with this, the high
plate voltage of Fig. 51c results in
the existence of high-potential
" channels " between grid wires,
illustrated in Fig. 51d. There is a
tendency for the electrons to follow
these channels in much the way
that water follows channels down
a steep hillside; they tend to pass
by on either side of each grid wire.
However, electrons whose accelera-
tion due to the field has aimed
them directly at the grid will strike
and enter it; furthermore, they
arrive with energy enough to knock
other electrons out of the grid.
The field near the grid is such

FIG. 52. Grid current curves for an RCA
Type 203-A (transmitter-type) triode.

as to remove these *secondary electrons* at once to the plate, for the po-
tential line slopes upward away from the grid wires on all sides. Thus

the grid current is small, because (1) most of the electrons pass along the high-potential channels, and (2) those that do enter the grid knock others out, producing a *secondary emission* current. This corresponds to conditions as at point C, Fig. 52.

It will be noted that certain voltage combinations in Fig. 52 make the grid current negative. This happens when each electron that strikes and enters the grid releases on the average more than one secondary electron. With a field like that illustrated in Fig. 51a there can be no secondary emission current, for secondary electrons appear in a field that returns them at once to the grid.

62. Tetrodes or Screen-Grid Tubes. The plate of a *triode* serves two functions: (*a*) in combination with the grid, it determines the electric field configuration within the tube, and therefore the magnitude of the space-charge-limited current that leaves the cathode; (*b*) it receives the electrons at the end of their travel and starts them along a return circuit to the cathode. To realize that these two functions are quite distinct it is only necessary to recall that in case of positive grid and low or negative plate voltages the electrons flow to the grid, yet the two electrodes work together to control the current, just as under more usual conditions. When a triode is used in an amplifier circuit these two functions partly conflict with one another, for reasons pointed out in a brief discussion of circuit behavior that appears near the end of this section.

The construction of the usual form of tetrode, or four-electrode tube, known commonly as a *screen-grid* tube (Fig. 53), is such as to segregate the two functions performed by a triode's plate. In a screen-grid tube the screen and control grid control the current, but the plate receives the electrons. The cathode and control grid perform the same functions as the cathode and grid in a triode.

The plate of a screen-grid tube is normally at a higher potential than the screen. The dotted potential line in Fig. 53b terminates at about the lower end, the solid one at about the upper end, of the normal range of plate voltage variations. As long as the plate voltage is within this range, the few electrons that make direct hits on the screen wires enter the screen and constitute the screen current. The others pass to the plate, because the gradient beyond the screen is positive. Thus the ratio of screen current to plate current is approximately that of projected area of screen wires to the area of the holes in the screen. Since this ratio is small, most of the electrons pass to the plate. Thus the plate to a large degree still has, as in a triode, the function of receiving the electrons.

However, the screen potential takes over the part played by a triode's

plate potential in determining the magnitude of the *total* space-charge-limited current, so that the equivalent voltage is $E_g + \dfrac{E_{\text{screen}}}{\mu_{\text{screen}}}$. The reason for this is that the screen shields the cathode from the plate effectively enough so that changes in plate voltage modify only very slightly, if at all, the space-charge-free off-cathode gradient and *total* tube current (plate current plus screen current) dependent on it.
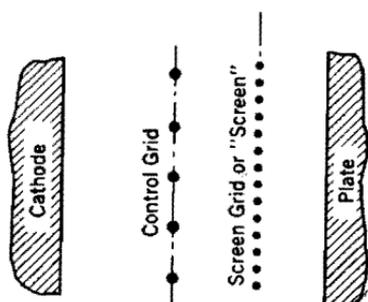
(a) Geometrical arrangement.
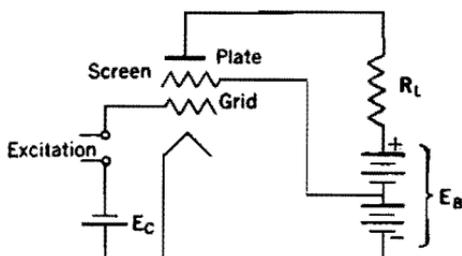
(c) Screen-grid amplifier circuit.
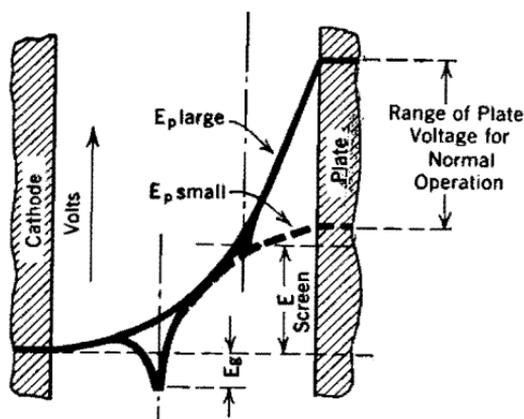
(b) Space-charge-free potential distribution, normal operating voltages.

FIG. 53.   Screen-grid tube.

This *lack* of control of plate voltage over total current is indicated by the horizontal nature of the total-current curves in Fig. 54. There is, it is true, a rather sharp but small rise in total current as the plate voltage changes from small negative to small positive values, but on the whole the *total* current is seen to be rather remarkably unaffected by changes in plate voltage. The small rise just above zero voltage will be explained later.

Within the normal operating range of plate voltages most of the electrons pass on through the screen to the plate, as explained above. Therefore the plate current is nearly equal to the total current and, like it, is very little affected by plate voltage variations. Fig. 55 is a set of plate characteristic curves that correspond to the total-current curves of Fig. 54, all being taken with the screen at 90 volts. The useful part of this diagram is that above about 100 volts, and it will be seen that in that region the curves have only very gentle slopes.

The " voltage gain " of an amplifier is defined in Section 1 as the ratio of the *change* in voltage $I_p R_L$ across the load resistor, Figs. 1 and 53c, to the *change* in grid voltage; see also Chapter XII.   In both triode and screen-grid amplifier circuits a rise in grid voltage causes an *increase* in
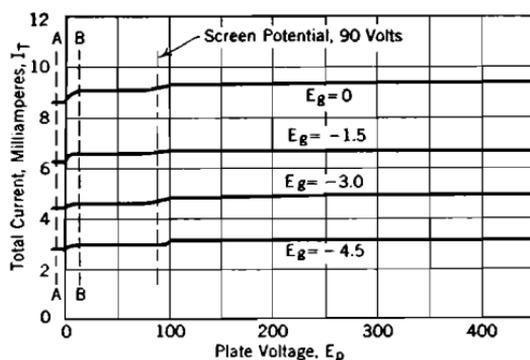

FIG. 54.   Curves of total current (screen current plus plate current) vs. plate voltage, for an RCA Type 24 screen-grid tube.

the equivalent voltage, so in plate current; see Equation (282).   This increase in plate current makes the voltage across R_L grow, so decreasing E_p.   The *plate* voltage E_p is always less than the *plate battery* voltage E_B by the amount of the voltage drop in the load.

In a triode this decrease in E_p tends to *reduce* the equivalent voltage.   The net change in I_p and in $I_p R_L$, so the voltage gain obtainable, is markedly restricted by the fact that E_p falls whenever E_g rises, and both affect the equivalent voltage.   In a screen-grid tube E_{screen} rather than E_p appears in the equivalent voltage, and E_{screen} is always constant, for the screen is connected directly to the battery.   No such restriction to the growth of I_p as that in a triode due to falling E_p exists in a screen-grid circuit, and the gain may be one or two or three hundred instead of between 5 and 15.

For this reason screen-grid tubes are commonly used in *voltage amplifier* circuits, in which a large voltage gain is desirable.   Usually the resistance or alternating-current impedance of the load in such circuits is a matter of hundreds

FIG. 55.   Plate characteristic curves for an RCA Type 24 (heater-type cathode) screen-grid tube.

of thousands of ohms, and plate currents are very small.   Triodes are commonly used in *power amplifier* circuits, in which considerable alternating-current power in the load, rather than large voltage gain, is desirable.   Load resistances in such circuits are usually a few thousands of ohms, and plate currents much larger than in voltage amplifier circuits.

At very high frequencies interelectrode capacitances play an important part in amplifier circuit behavior. In particular, the results of careful shielding of parts external to the tube, such as lead wires, coils, etc., from one another, to avoid troublesome capacity coupling between grid and plate circuits, may be largely nullified by the grid-plate capacitance within a triode. The mesh of the screen of a screen-grid tube is sufficiently fine to shield effectively the two circuits from one another inside the tube; this is a very valuable feature.

63. The Various μ's for a Tetrode. It is possible to carry out an analysis of the electrostatic field of a screen-grid tube in a manner somewhat similar to that used for a triode in Chapter II. Suppose a transformation including one control-grid-wire section is made, and the cathode charge for one section called $-\tau_c$ as before. Then, following the general method of Section 17,

$$-\tau_c = E_g C'_{gc} + E_s C'_{sc} + E_p C'_{pc} \qquad \text{(285 esu)}$$

where the primed C's are capacitance coefficients for the transformed section only, and E_g, E_s, E_p are the grid-to-cathode, screen-to-cathode, and plate-to-cathode potentials. This can be rewritten

$$-\tau_c = C'_{gc} \left[E_g + \frac{E_s}{\dfrac{C'_{gc}}{C'_{sc}}} + \frac{E_p}{\dfrac{C'_{gc}}{C'_{pc}}} \right] \qquad \text{(286)}$$

which is exactly analogous to Equation (96).

An expression for F_0 is obtained by multiplying both sides of this equation by $4\pi n$. The form corresponding to Equation (82) is

$$-F_0 = \frac{E_g + \dfrac{E_s}{\mu_{sg}} + \dfrac{E_p}{\mu_{pg}}}{d_1} \qquad \text{(287 esu)}$$

where

$$\mu_{sg} = \frac{C'_{gc}}{C'_{sc}} \qquad \text{(288)}$$

$$\mu_{pg} = \frac{C'_{gc}}{C'_{pc}} \qquad \text{(289)}$$

$$d_1 = -\frac{1}{4\pi n C'_{gc}} \qquad \text{(290 esu)}$$

Thus a tetrode has two geometric "μ's," a pentode similarly three, and so on.

It has been pointed out that in the usual tetrode the screen is an effective electrostatic shield between cathode and plate. The same idea is expressed in the statement that C'_{sc} is very much greater than C'_{pc},
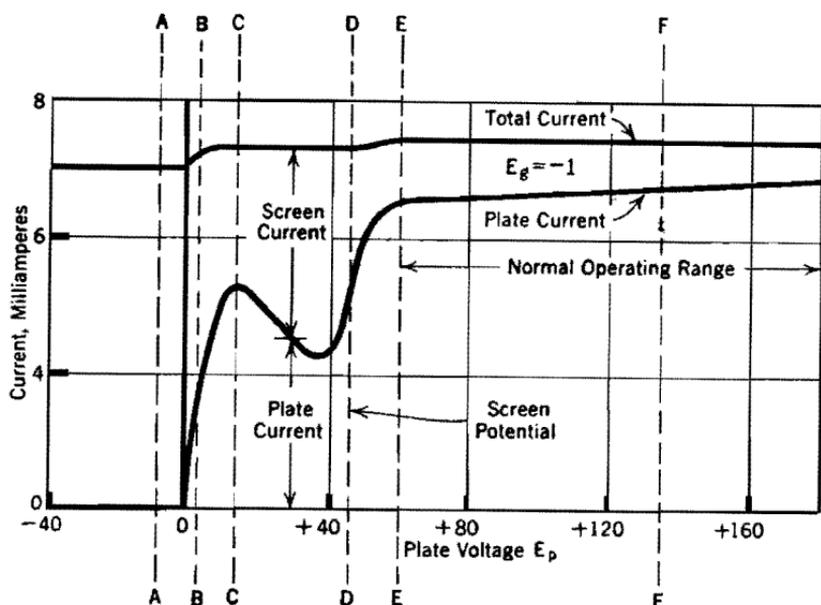
FIG. 56.   Plate characteristics of a tetrode or screen-grid tube; see Fig. 57.   Directly heated cathode.   Screen potential 45 volts.
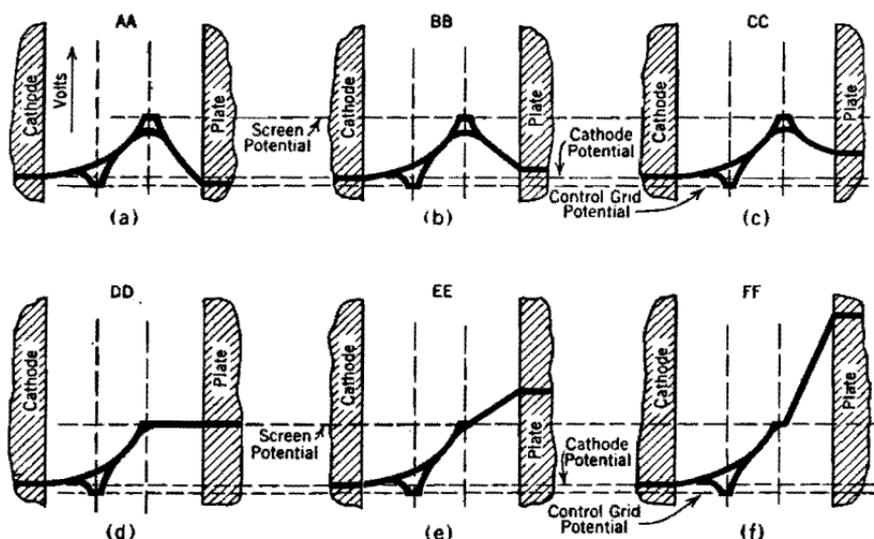

FIG. 57.   Space-charge-free potential distributions in screen-grid tube, corresponding to potentials as at *A-A*, *B-B*,——— *F-F*, in Fig. 56.

so that μ_{pg} is ordinarily very much larger than μ_{sg}.   The lack of response of current to changes in E_p arises from this fact.   The flatness of the total-current curves of Fig. 54 indicates that μ_{pg} for that tube must be very large indeed.   The value of μ_{sg} is usually within the general range of values of μ for triodes.

Neither μ_{sg} nor μ_{pg} are commonly used in screen-grid amplifier circuit calculations, nor are they properly called " amplification factors." They relate to F_0 and so to the total current in the tube; circuit calculations relate to plate current only.   *Only in triodes is the geometric μ the same quantity that is used to describe the " amplifying " properties of an equivalent circuit.*   In Chapter XII it will be shown that the amplification factor may be defined entirely in terms of graphical properties of tube characteristic curves, and when so defined is a constant dependent entirely on tube geometry *only for triodes*.   Thus the amplification factors listed for tetrodes in manufacturer's tube manuals are neither μ_{sg} nor μ_{pg}.

64. Analysis of Screen-Grid Characteristics; Secondary Emission. Although variations in plate voltage have very little effect on the total current through a screen-grid tube, they do affect very strikingly the division of the current between screen and plate.   Figs. 57a, b, c, d, e, f illustrate the space-charge-free potential structure for various positions marked AA, BB, CC, etc. on the one plate characteristic curve of Fig. 56. This curve corresponds to a slightly negative grid voltage, in a tube with properties a little different from that used for Figs. 54 and 55; in particular it has a directly heated cathode, with a drop of several volts along the filament.   The screen and control grid voltages are the same for all conditions illustrated in Figs. 56 and 57, all variations in potential structure and in current being due to changes in plate voltage.

At AA the plate voltage is slightly, but very slightly negative, so that no electrons can have energy enough to reach it; the plate current is zero.   Electrons are emerging from the cathode in considerable volume; of course many of them travel through the screen holes, for by the time they reach the screen their velocity is such as to carry them successfully through the considerable side-pulling field that draws them toward individual screen wires.   But even though most of them pass the screen, none can ever reach the plate, for the energy each one has at any instant is not more than that corresponding to the potential through which it has fallen; it cannot possibly reach a lower potential than that from which it started.

Each electron's situation may be compared to that of a ball rolling down the smooth sloping side of a trough with a rounded bottom; it cannot acquire kinetic energy enough on the way down to carry it up the

opposite slope to a point higher than that from which it started. A ball started down one side of a trough, at the bottom of which there are regularly spaced holes, presents a more nearly complete analogy. If the top of the opposite side of the trough is higher than the starting point, the ball may roll back and forth several times, first down one side, then up and down the other, and so on, before it happens to drop into one of the holes. Similarly at AA most of the electrons oscillate back and forth between the two sides of the screen many times before finally entering the screen wires, as all must eventually do.

At BB the plate current is rapidly rising. There are two distinct reasons why in this portion of the curve a small increase in plate voltage produces a large rise in plate current, as follows:

(a) There is a potential drop of several volts along the filament of this tube; when as at BB the plate potential is near that of the midpoint of the filament, the electrons that start from a filament section that is more positive than the plate cannot possibly reach the plate. Each increment of plate voltage increases the length of the filament section that is lower in potential than the plate and so able to supply electrons that can reach it. This effect is not present in tubes with cathodes that are true equipotential surfaces.

(b) Dispersion in passing the neighborhood of the screen makes it impossible for all electrons that start from a potential lower than that of the plate to reach it. All such have energy enough, if properly directed, to bring them to the plate, but many are deflected in passing between screen wires, as described in connection with Fig. 28, to an extent that makes it impossible for them to get to a plate that is at a potential even several volts above that of the point on the filament from which they start. This effect is present whether the cathode is an equipotential surface or not.

Of course there are great variations in the amount of deflection that takes place. Some electrons fail to reach the plate by one volt, not so many by two volts, still fewer by three volts, and so on; a few may fail by as much as ten volts. If the plate is raised to one volt above the starting-potential, those that formerly failed to reach it by one volt or less can arrive; with the plate up two volts it receives in addition those formerly failing by between one and two volts, and so on. Each increase in plate voltage makes it possible for a considerable number of additional electrons to reach the plate, so increasing the plate current.

Before the plate voltage has risen far enough to permit the plate current to approach at all closely to its maximum share of the total current, an entirely different effect begins to distort the shape of the curve, as at CC. Here the plate voltage has just become high enough for some of the electrons that strike the plate to arrive with energy sufficient to " splash " one or more electrons out of the plate surface. As soon as the plate potential is from four to six volts higher than that of the starting point, incoming electrons begin in this way to produce

secondary emission from the plate. Since the ejected electrons appear in a field which draws them toward the screen, the plate current becomes less, the screen current greater, by just the amount of secondary emission current.   This accounts for the pronounced dip in plate current just beyond *CC*.   In general the secondary emission current increases linearly with the energy possessed by electrons that arrive at the plate, so that the tendency is toward a straight-line decrease of plate current in this region.

It is possible to have more secondary electrons leaving the plate than there are primary electrons arriving, so that the net plate current may become negative.   Each incoming electron must then be releasing, on the average, more than one secondary electron.   This principle has been used to amplify photoelectric currents.

Evidence of secondary emission disappears rather abruptly as the plate potential approaches and rises above screen potential, as beyond *DD*.   Although incoming electrons at the plate continue to be increasingly capable of ejecting secondary ones, the plate potential is higher than that of the screen, so that the electric field adjacent to the plate prevents any ejected electrons from leaving the neighborhood of the plate.   On the contrary, they promptly reenter it.

EE is about the beginning of the normal operating voltage range of the tube used for Fig. 56; and *FF* is about the normal operating value of plate voltage, that is, the plate potential may ordinarily be expected to swing about equally above and below *FF*.

The plate characteristic curves of Fig. 55 are not quite horizontal, nor are they exactly parallel to one another, in the operating range. Since the corresponding *total-current* curves are horizontal and parallel to one another, the gradual growth in plate current in the operating range cannot be attributed primarily to growth in the space-charge-free off-cathode gradient.   The cathode continues, even at these high voltages, to be quite effectively screened from the electrostatic influence of the plate.   But as the plate potential rises, it affects the detailed potential configuration in the neighborhood of the screen in such a way that fewer electrons are stopped there.   High-potential channels between screen wires, like those illustrated in Fig. 51, become more pronounced, making " direct hits " of the screen wires less frequent; as the plate potential rises, a slightly smaller proportion of the total current enters the screen wires in this way, and a slightly larger proportion passes on to the plate.

Since this change is in the nature of an increase in the ratio of plate current to total current, the actual change should be greater along the upper plate characteristic curves of Fig. 55 than along the lower ones;

that is, the higher ones should have steeper slopes than the lower ones. The figure shows that this is to some extent true.

It is rather generally true, as for the variation just described, that each ordinate of any one of these plate characteristic curves is a definite fraction, dependent on plate voltage, of a total current which depends on grid and screen voltages. This behavior is discussed in more detail in Section 66.

Individual screen-grid mutual characteristic curves, obtainable either by direct measurement or by transfer from a set of plate characteristics, are very similar to triode mutual characteristic curves (Figs. 3a and 50a). They lie very much closer to one another than do the triode curves, however, because of the small changes in current that result from plate voltage variations in tetrodes.

65. Shielding of Screen by Oscillating Space Charge. The curves of Fig. 54 bear out for the most part the statement that *total* current is independent of plate voltage. However, there is a small but sharp rise in total current just above zero plate voltage in each curve. As explained in the following paragraphs, this is attributable to a change in space-charge distribution in the neighborhood of the screen.
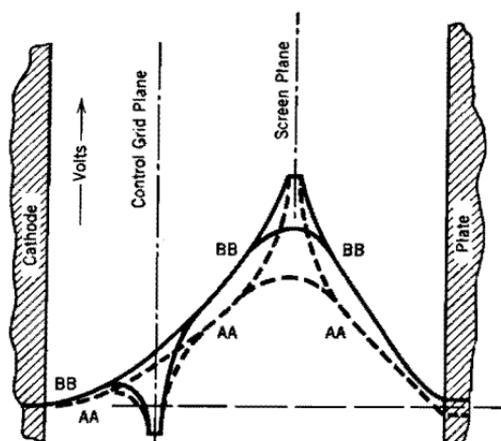

Dotted lines *AA* describe the space-charge-*flow* potential distribution for the negative-plate condition *AA*, in Figs. 54 and 57. All electrons enter the screen after oscillating.

Solid lines *BB* also describe a space-charge-*flow* potential distribution, but for the positive plate condition, *BB*, in Figs. 54 and 57. Nearly all electrons go to the plate.

FIG. 58.   Space-charge shielding of screen wires.

The dotted lines in Fig. 58 describe the potential variation within the tube when the plate is slightly negative and current is flowing. All electrons must enter the screen, but they do so only after considerable oscillation back and forth through it; this corresponds to conditions as at *AA* in Figs. 54 and 57. These oscillating electrons represent a considerable space charge in the neighborhood of the screen, so that the dotted-line space-charge-flow variation in Fig. 58 has a very marked convex-downward flexion immediately adjacent to the screen.

The solid lines in the same figure similarly represent space-charge-flow variations of potential, but for conditions as at BB, Figs. 54 and 57. The plate is now sufficiently positive so that most of the electrons enter the plate after the first flight through the tube; there are few if any oscillating electrons. There is very little space charge near the screen, for the electrons are moving very rapidly as they pass it; therefore the solid lines in Fig. 58 follow, near the screen, practically the space-charge-free pattern.

The essential contrast between the solid and dotted lines is of the type that would result from a reduction of screen potential, for a lowering of the screen potential would drop the BB lines to about where the AA lines now are, at all points except those immediately adjacent to the screen. The sharp dotted-line flexion near the screen requires a steep gradient there; this " uses up " a considerable part of the potential difference between screen and cathode, so that to an observer near the cathode the screen *appears* to have a lower potential in the dotted-line (negative plate) case than in the solid-line (positive plate) case. Therefore the total space-charge-limited current is less for the AA than for the BB condition, Figs. 54 and 57.

The same idea may be expressed differently by saying that the negative space charge around the screen partially neutralizes the effect that the positive surface charge on the screen wires would otherwise produce. For example, in the dotted-line case the screen appears to a remote observer to have less positive charge, therefore less potential, than it actually has.

At the current minimum just to the left of DD, Fig. 56, there is again considerable screen current, especially when the total current is large. However, this current is due to secondary electrons which start from the plate, at a potential only slightly less than that of the screen. Their energy on approaching the screen is moderate, so that only a few oscillations occur, and the space-charge effect is moderate. Its disappearance accounts for the slight rise in total current as the plate rises above screen potential and thereby eliminates the secondary-emission current to the screen.

66. Pentodes, Beam Power Tubes, and Critical Distance Tubes. Five-electrode tubes can be designed for various purposes, but as usually used, the term *pentode* is applied to devices, otherwise similar in arrangement and principles to screen-grid tubes, in which a fifth electrode, screen-like but *at cathode potential,* is introduced between screen grid and plate to " suppress " secondary emission. As illustrated in potential-diagram form in Fig. 59, the essential result of its presence is to produce another potential valley and hill. This makes the gradient

just outside the plate always such as to drive back into it any electrons that may be ejected. This prevents any secondary electrons from leaving the plate and eliminates entirely the dip in plate current that is characteristic of the low-voltage part of a set of screen-grid characteristics. Figs. 60a and 60b are typical pentode mutual and plate characteristics. The detail shapes of the curves at low plate voltages are, as in screen-grid tubes, jointly dependent on (a) electron dispersion in passing through control grid and screen or screens, and on (b) the fact that all electrons may not start from the same potential. At higher plate voltages the division of current between screen and plate depends chiefly on (c) the screening-fraction of the screen, which measures the probability of direct hits on it in the absence of electron deviation by the field, and on (d) the effect of plate potential variations in creating and deepening between-screen-wire channels. In screen-grid tubes, but not in pentodes, there is an additional effect, (e) the flow of secondary electrons from plate to screen.

FIG. 59. Potential diagram for a pentode, showing at the plate a positive gradient which suppresses secondary emission.

(a) Mutual characteristics.

(b) Plate characteristics.
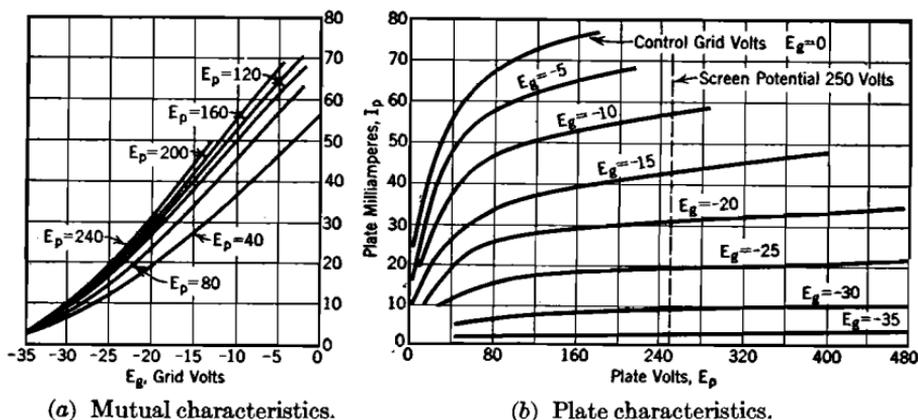
FIG. 60. Average characteristic curves, RCA Type 59 tube, pentode connection, with screen at 250 volts.

All these influences are alike in that they act to modify the fraction of the total current that reaches the plate, regardless of how much or little the total may be. This fraction is thus approximately the same for a given plate voltage regardless of the magnitude of the total current.

In summary, for both screen-grid tubes and pentodes of the usual

type, (1) screen and control grid potentials determine the total space-charge-limited current in just the way plate and control grid voltages do in a triode, and since screen potential is usually held constant, the total current is a simple function of the control grid voltage; (2) the plate voltage specifies what per cent of this total current shall reach the plate.

Hence all plate characteristics for a given tetrode or pentode may be expected to be alike in shape, their ordinates being related by constant ratios. It is interesting to recall that the plate characteristics for a given triode are alike except for a horizontal shift.

Pentodes are often used in *power amplifier circuits,* as in such service they can provide alternating currents of the desired magnitude in load circuits, and at the same time permit a considerably higher voltage gain than is possible with power amplifier triodes. Some fidelity of response may be sacrificed when pentodes rather than triodes are used.

In " beam power tubes "[63] and " critical distance tubes "[64] potential distributions qualitatively similar to that illustrated in Fig. 59 are obtained without the use of fifth electrodes. The negative space charge of the electron stream is depended on to produce the potential minimum between screen and plate that the suppressor grid provides in a pentode. As long as such a minimum exists, no secondary electrons pass from plate to screen. The potential minimum in case of this " space-charge suppression " of secondary emission may not drop to zero voltage. The requirement for the suppression of secondary emission is that at some point between screen and plate the potential must fall below its value at the plate.

The essential features necessary for space-charge suppression of secondary emission, of the type employed in beam power and critical distance tubes, are that

(1) Between screen and plate, electron flow must take place in a beam of high enough concentration so that space charge makes an appreciable contribution to the convex-downward flexion of the potential distribution curve.

(2) The plate must be located far enough beyond the screen to permit the convex-downward flexion to produce a minimum in the potential distribution curve between screen and plate.

The properties and uses of tubes in which secondary emission is suppressed by space charge are essentially similar to the properties and uses of pentodes. However, critical distance and beam power tubes appear to have some distinctly desirable features that are not possessed by pentodes.

In order to understand the underlying principles of operation of beam power tubes and critical distance tubes it should first be noted that the

validity of Equation (232), Section 51, is not confined to the region between cathode and control grid. Equation (232) is in fact a very general differential relation that applies wherever space-charge flow takes place along parallel straight lines (one-dimensional flow). It therefore applies in the region between screen and plate, in tubes having essentially parallel-plane geometry. The results obtained from an analysis based on true one-dimensional space-charge flow apply approximately to flow along only slightly diverging or converging paths.

The boundary conditions that must be used in obtaining solutions of Equation (232) are very different for the different interelectrode regions. The solution for the region between screen and plate, in a tube employing space-charge suppression of secondary electrons, must satisfy the following assumptions:

(a) The current density J is fixed and known, being established by the potentials and locations of control grid and screen, just as in a screen-grid tube. More accurately, for reasons that will appear later, the current density cannot exceed this fixed value J.

(b) Let $x = 0$ at the potential minimum that exists somewhere between screen and plate, and that is responsible for the suppression of secondary emission. Then at the screen, $x = x_1$ (numerically negative) and at the plate $x = x_2$ (positive). The location of the potential minimum is not known to begin with; however, it is known that

$$x_2 - x_1 = d \qquad (290.01)$$

where d is the fixed spacing between screen and plate.

(c) Let E_0 be the potential, relative to the cathode, at the potential minimum. E_0 is unknown to begin with, and must be evaluated to suit the known boundary conditions.

(d) The electrons pass between screen wires at some approximately known potential E_1 that is slightly lower than screen potential (see Fig. 59). Therefore

When $\qquad\qquad x = x_1, \quad E = E_1$

(e) The electrons are at the potential E_p on arrival at the plate. Therefore

When $\qquad\qquad x = x_2, \quad E = E_2 = E_p$

(f) At the potential minimum, the potential gradient is zero, so that

When $\qquad\qquad x = 0, \quad \dfrac{dE}{dx} = 0 \quad \text{(also } E = E_0) \qquad (290.02)$

In the solution of Equation (232) with these new boundary conditions,

the first integration constant does not vanish, because $E = E_0$ when $dE/dx = 0$. Therefore instead of Equation (235) the following expression must be used:

$$\left(\frac{dE}{dx}\right)^2 = 16\pi J \sqrt{\frac{m_e}{e}} \left(\sqrt{E} - \sqrt{E_0}\right) \qquad \text{(290.03 esu)}$$

Other useful forms of Equation (290.03) are

$$\frac{dE}{dx} = \pm \sqrt{16\pi J \sqrt{\frac{m_e}{2e}} \left(\sqrt{E} - \sqrt{E_0}\right)} \qquad \text{(290.04 esu)}$$

$$\frac{dE}{\sqrt{\sqrt{E} - \sqrt{E_0}}} = \pm \sqrt{16\pi J \sqrt{\frac{m_e}{2e}}} \, dx \qquad \text{(290.05 esu)}$$

Integration is effected by the use of the substitution $\sqrt{E} = u$, and yields the following:

$$\left(\sqrt{E} - \sqrt{E_0}\right)^{\frac{3}{2}} + 3\sqrt{E_0}\left(\sqrt{E} - \sqrt{E_0}\right)^{\frac{1}{2}} = x\sqrt{9\pi J \sqrt{\frac{m_e}{2e}}} \quad \text{(290.06 esu)}$$

No integration constant need be added, because $x = 0$ when $E = E_0$. Equation (290.06) can be rearranged into the following form:

$$x = \frac{E^{\frac{3}{4}} f\left(\dfrac{E}{E_0}\right)}{\sqrt{9\pi J \sqrt{\dfrac{m_e}{2e}}}} \qquad \text{(290.07 esu)}$$

Here $f\left(E/E_0\right)$ is a function of E/E_0 that is defined as follows:

$$f\left(\frac{E}{E_0}\right) = \left[\sqrt{1 - \sqrt{\frac{E_0}{E}}}\right]\left(1 + 2\sqrt{\frac{E_0}{E}}\right) \qquad \text{(290.08)}$$

$f\left(E/E_0\right)$ has numerical values as follows:

E/E_0	$f(E/E_0)$	E/E_0	$f(E/E_0)$	E/E_0	$f(E/E_0)$
1.000	0.00000	1.10	0.6271	4.0	1.414 (maximum)
1.001	0.06705	1.20	0.8340	5.0	1.409
1.002	0.09408	1.40	1.058	6.0	1.397
1.004	0.13385	1.60	1.181	8.0	1.372
1.01	0.2107	1.80	1.257	10.0	1.350
1.02	0.2958	2.00	1.307	15.0	1.306
1.04	0.4126	2.50	1.376	20.0	1.275
1.06	0.4984	3.00	1.401	100.0	1.138
1.08	0.5682	3.50	1.412		1.000

For large values o⸍ E/E_0, $f(E/E_0)$ may be evaluated from the following infinite series:

$$f\left(\frac{E}{E_0}\right) = 1 + \frac{3}{2}\sqrt{\frac{E_0}{E}} - \frac{9}{8}\frac{E_0}{E} - \frac{5}{16}\left(\frac{E_0}{E}\right)^{\frac{3}{2}} - \frac{21}{128}\left(\frac{E_0}{E}\right)^2 - \cdots$$

When $x = x_1$, $E = E_1$, so that

$$x_1 = \frac{E_1^{\frac{3}{4}}f\left(\dfrac{E_1}{E_0}\right)}{-\sqrt{9\pi J\sqrt{\dfrac{m_e}{2e}}}} \qquad (290.09 \text{ esu})$$

The negative sign of the radical is employed because at $x = x_1$ the slope dE/dx is negative; see Equation (290.04).

Now let

$$a_2 = \frac{E_2^{\frac{3}{4}}}{\sqrt{9\pi J\sqrt{\dfrac{m_e}{2e}}}} \qquad a_1 = \frac{E_1^{\frac{3}{4}}}{\sqrt{9\pi J\sqrt{\dfrac{m_e}{2e}}}} \qquad (290.10 \text{ esu})$$

By comparison with Equation (236), it will be seen that a space-charge-limited current density J of the usual type would flow between parallel plane electrodes a_2 centimeters apart if at a potential difference E_2, and that a_1 and E_1 are similarly related. Using Equation (290.09)

$$x_2 = a_2 f\left(\frac{E_1}{E_0}\right) \qquad (290.11)$$

Similarly

$$x_1 = -a_1 f\left(\frac{E_1}{E_0}\right) \qquad (290.12)$$

Using Equation (290.1)

$$a_1 f\left(\frac{E_1}{E_0}\right) + a_2 f\left(\frac{E_2}{E_0}\right) = d \qquad (290.13)$$

Equation (290.13) permits determination of E_0 by a trial and error method, if d, J, E_2 and E_1 are known.

Equation (290.13) can be used to determine certain critical relations. Let the spacing d have some fixed value. One critical condition is that for which $E_0 = 0$, for no electron flow can take place past a potential minimum that dips below cathode potential. If $E_0 = 0$, $f(E/E_0) = 1$. Thus if the potential minimum just touches the zero-voltage axis, Equation (290.13) becomes

$$a_1 + a_2 = d \qquad (290.14)$$

Equation (290.14) and (290.10) permit the determination of the value

of E_2 (plate potential) at which this critical condition exists, for given values of J, E_1, and d. In any particular case J depends on the values of control grid and screen potentials.

If the plate potential falls below the critical value, the current density falls very rapidly below its previously constant value, and the screen current rises. The potential minimum remains at zero volts, but its location shifts toward the screen, a_2 increasing, because J falls off much more rapidly than $E_2^{\frac{3}{2}}$, in Equations (290.10). The value of a_1 is no longer determinable from Equations (290.10), because electrons are oscillating back and forth between the screen wires, and the boundary conditions are correspondingly more complicated than when all electrons that pass the screen arr ve at the plate.

The analysis just given explains why, in beam power tubes and critical distance tubes, the plate current drops off very rapidly as the plate voltage falls below a critical value. Other very interesting relations relative to critical distances and potentials can be obtained by further study of the various equations given above.

67. Oscillating Space Charge. A " Barkhausen-Kurtz " oscillator [65, F 309] is a circuit arrangement which produces very high frequency electric currents and radio waves by making use of electron oscillations similar to the movements back and forth through the screen that occur with potentials as at AA, Fig. 57. A circuit of this type employs a triode, with the grid positive and the plate at a negative or zero potential, as illustrated in Fig. 61. The current is ordinarily *tempera-ture-limited*. There is a particular frequency of oscillation for a particular set of dimensions and voltages, just as the motion of a ball rolling back and forth across a trough has a definite periodicity that is dependent on the height of its starting point and the slopes of the sides of the trough.

In a triode with potentials as in Fig. 61, nearly all the individual elec-

FIG. 61.   Space-charge-free potential distribution in a triode used as a Barkhausen-Kurtz oscillator.

trons oscillate back and forth between the two sides of the grid before entering it, even when the current to the grid is perfectly steady. The to-and-fro nature of the movements of the individual electrons can play a part in causing external circuit oscillations only when there is an appreciable periodic *mass movement* of space charge from one side of the grid to the other and back again. Such an oscillating space charge

is an alternating current with a capacitance coupling to the external circuit through the electrodes; it may resonate to produce sustained circuit oscillations if the circuit constants are correctly chosen.

The oscillation frequency for mass movement can be approximated by calculating the periodicity to be expected in the movement of a single electron. The mass movement periodically distorts the potential structure, but reasonably correct frequency estimates can be made by ignoring this distortion. The simplest assumption on which to base mathematical calculations of ‖the oscillation period is that of straight-line variation of potential between cathode and grid and between grid and plate, as indicated by the dashed lines CG and GP in Fig. 61.

Electrons which originate at the cathode and pass between grid wires have in general enough energy to permit them to approach closely the plate, which is at cathode potential. The direction of motion then reverses and each one travels back to the neighborhood of the cathode. If the equation of the space-potential curve is known, the time necessary to complete a cycle of motion from C to P and back again is in general calculable from Equation (154), which expresses velocity in terms of potential. For the relation

$$\frac{dx}{dt} = v = 5.93 \times 10^7 \sqrt{E} \qquad (291\ p)$$

can be written

$$t = \frac{1}{5.93 \times 10^7} \int \frac{dx}{\sqrt{E}} \qquad (292\ p)$$

and if E can be expressed in terms of x or vice versa, integration over a cycle gives the required time.

If the origin is taken at C, the equation of the potential line from C to G is

$$\frac{x}{s_g} = \frac{E}{E_g}, \quad \text{or,} \quad E = E_g \frac{x}{s_g} \qquad (293)$$

Using this in the manner suggested above, the time required to travel from C to G is

$$t_{CG} = \frac{1}{5.93 \times 10^7} \sqrt{\frac{s_g}{E_g}} \int_C^G \frac{dx}{\sqrt{x}} = \frac{2}{5.93 \times 10^7} \sqrt{\frac{s_g}{E_g}} \sqrt{x} \Big]_{x=0}^{x=s_g}$$

$$= \frac{2 s_g}{5.93 \times 10^7 \sqrt{E_g}} \qquad (294\ p)$$

The time for travel from G to P is similarly obtainable, the distance factor being $s_p - s_g$ instead of s_g. The total time t for a complete cycle is twice the sum of t_{CG} and t_{GP}, and is given by the simple

expression

$$t = \frac{4s_p}{5.93 \times 10^7 \sqrt{E_g}}$$ (295 p)

The frequency of oscillation is the reciprocal of this oscillation period; it is

$$f = \frac{5.93 \times 10^7 \sqrt{E_g}}{4s_p}$$ (296 p)

The wave length λ of the electromagnetic radiation that corresponds to this frequency is found by dividing the velocity of light, 3×10^{10} centimeters per second, by the frequency. It is

$$\lambda = 2023 \frac{s_p}{\sqrt{E_g}} \quad \text{(centimeters)}$$ (297 p)

The times for traveling from C to G and from G to P could have been obtained more simply in this particular case by the use of average velocities, for the assumed field is uniform and corresponds to uniformly accelerated motion. The more general method has been used in order to illustrate the manner of obtaining the time of travel in a more general type of field, as for example that corresponding to a concentric arrangement of electrodes.

In real apparatus there is a steady component of current due to electron flow from cathode to grid, which is, as mentioned earlier, usually a temperature-limited rather than a space-charge-limited current. Entrance of this grid current into the grid wires is of course preceded as usual by individual electronic oscillations, so that the effective grid potential is modified by space-charge shielding of the type described in Section 65. For this reason the frequency of oscillation must be expected to decrease somewhat if the direct current to the grid increases.

The analytical method presented above for determination of frequency is incomplete in that it fails to account for the effects of the periodic distortion of the shape of the potential diagram that is essential to coupling into the external circuit, as described in the next section.

68. Coupling between Internal and External Oscillations. The manner in which the internal electronic oscillations can be electrostatically coupled into an external resonating electric circuit is illustrated in Figs. 62a, b, c. The first of these illustrates the potential diagram when the oscillating mass of space charge is adjacent to the cathode; it will be noticed that the gradient just outside the cathode is very small, hence the negative surface charge on the cathode is also small. By contrast, there is a considerable negative surface charge on the plate,

corresponding to the steep off-plate gradient. In Fig. 62b the space-charge mass has moved to the neighborhood of the grid; the gradient at the cathode is greater than before, that at the plate less, so that some negative charge must have been transferred from plate to cathode through the external circuit. (Since the steady component of current in these devices is normally temperature-limited, not space-charge-limited, the gradient at the cathode need not always be near zero.) At Fig. 62c the space-charge mass has moved to a region adjacent to the plate, so that the entire plate charge has been driven away through the outside circuit to the cathode surface.

Thus for the plate to remain at cathode potential there must be an oscillating current flow in the external circuit between plate and cathode. If this external circuit consists of an inductance and capacitance in series resonance at the proper frequency there may be a very large high-
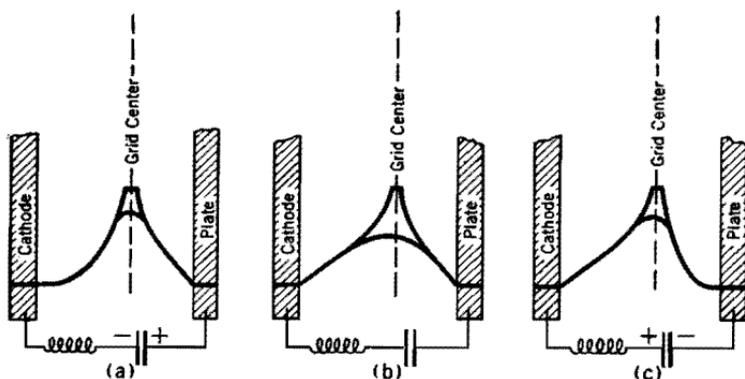
Fig. 62. Coupling between oscillating space charge and external circuit.

frequency voltage across the inductance and across the capacitance, yet no variation in plate potential.

The analysis just given indicates the manner of transfer of *periodicity* from inside to outside the tube, but does not illustrate the manner of transfer of *power* from the direct-current grid circuit to the alternating-current plate-cathode circuit. Any such power transfer must be accompanied by a corresponding variation of plate voltage, and occurs only when the external circuit contains resistance or its equivalent, in addition to resonating inductance and capacity.

Suppose that the electronic space-charge mass has accumulated kinetic energy on its way from cathode to grid and has passed through the grid; as the plate is approached the oncoming charge attempts to drive the negative charge away from the plate surface through the external circuit. Since the external circuit contains some resistance, this can happen only if the plate goes somewhat negative, so the plate is forced to drop below

cathode potential. In order to pull the same negative charge back through the load resistance the plate later swings positive, shortly after the internal space charge reverses its motion. Thus there is a component of alternating plate voltage which is in phase with the current flow in the external circuit. The impetus to the production of this plate-voltage variation is given by the kinetic energy of the electron's motions, and this energy is in every case acquired by acceleration due to the grid potential. Thus the grid-circuit battery is the original source of supply of the alternating-current energy in the oscillating circuit between plate and cathode.

PROBLEMS

CHAPTER VI

1. *Plate resistance R_p* is defined in a footnote near the end of Section 53, and again in Section 133, Chapter XII. Using Equation (282) and the definition of R_p, derive the following expression for the plate resistance of a triode:

$$R_p = \frac{\mu E_g + E_p}{n I_p} \qquad (297.1 \text{ p})$$

2. Show that the manner of occurrence of μ in Equation (282) is consistent with the definition of μ given by Equation (490), Chapter XII, and is also consistent with Equation (100) in Chapter II.

3. Plot a curve of the various values of E_p that are necessary to maintain I_p constant at 8 milliamperes, as E_g is varied, for the triode of Fig. 3. From the resulting graph of E_p vs. E_g at constant plate current, determine the value of μ, as defined by Equation (490), Chapter XII, for the triode of Fig. 3.

4. Same as Problem 3, except that Fig. 50 is to be used, and plate current assumed constant (a) at 8 milliamperes, (b) at 2 milliamperes. Also determine μ for Fig. 50 from each of several different cut-off points, using Equation (100), Chapter II. Which of these various methods of determining μ for a triode is the most satisfactory? (See Section 60.)

5. Suppose that in the triode of Fig. 20 the grid voltage is made +30 and the plate voltage +50.

(a) Sketch the space-charge-free potential distribution curves for paths through and between grid wires.

(b) On the same sketch, locate as nearly as possible the *space-charge-flow* potential distribution curves for the same electrode voltages.

(c) Estimate what per cent of the total current flow through the tube will be grid current, neglecting secondary emission from the grid.

(d) Calculate plate current and grid current, in milliamperes per sq cm.

6. Locate four points each on the mutual characteristic curves for $E_p = 200$ volts, 300 volts, and 400 volts, for the tube of Fig. 55. Draw the portions of the mutual characteristic curves that are located by these points. Determine the grid-plate transconductance s_m for this tube at $E_p = 300$, $E_g = -3$, using Equation (489), Chapter XII, as the definition of s_m. Also determine s_m at $E_g = -8$, $E_p = 130$, for the tube of Fig. 3, that at $E_g = -8$, $E_p = 200$, for the tube of Fig. 50, and that at $E_g = -15$, $E_p = 120$, for the tube of Fig. 60.

7. Determine the values of R_p that correspond to those for s_m asked for in the preceding problem, then calculate μ in each case, using Equation (491), Chapter XII.

8. For $E_p = 150$ in the tube of Figs. 54 and 55, find the ratio of screen current to total current, for $E_g = 0, -1.5, -3, -4.5$. Repeat at $E_p = 50$, and at $E_p = 400$.

9. Reproduce to scale on a sheet of graph paper the plate characteristic curve for $E_g = -15$, Fig. 60b. Mark also on your graph paper the intersections of all the plate characteristic curves with the vertical line for which $E_p = 200$, as read from Fig. 60b. Now construct a complete set of characteristics by using the assumption that total current is independent of plate voltage, but that plate voltage determines the fraction of the total current that reaches the plate, that fraction being the same for a given plate voltage regardless of how much the total current may be. This construction can be accomplished without knowing the total current. Compare the set of characteristics so prepared with those of Fig. 60b.

10. By making a rough comparison between the shapes of the curves in Figs. 55 and 60b, estimate the secondary emission current due to passage of electrons from plate to screen when $E_g = 0$, $E_p = 50$, Fig. 55. Estimate also the screen current due to electrons striking the screen on their first flight from the cathode. What is the ratio of the number of secondary electrons emitted to the number of electrons causing secondary emission, according to your estimates?

11. (a) Sketch a plate characteristic curve similar to that for $E_g = 0$, in Fig. 55, except that it is to pass through the point $I_p = -1, E_p = +50$. That is, secondary emission is presumed to be pronounced enough to make the plate current negative for certain values of plate voltage, screen held at 90 volts.

(b) Sketch the plate characteristic curve for $E_g = -3$ that corresponds to the one for $E_g = 0$ drawn in part (a). Total current is presumed to be as in Fig. 54.

12. Explain why the contrast between the potential distribution curves asked for in parts (b) and (c) of Problem 4, Chapter V, illustrates a part, but only a part, of the explanation for the space-charge shielding of screen wires discussed in connection with Figs. 58 and 54.

13. (a) Redraw Fig. 57c with the plate far enough away from the screen to produce space-charge suppression of secondary emission. Your potential distribution line should represent space-charge-flow conditions.

(b) Same as (a), except that the plate voltage should be low enough so that the critical condition for which $E_0 = 0$ is reached (see Section 66).

14. Consider a parallel-plane tetrode that is similar to a screen-grid tube except that the plate is located at a considerable distance d centimeters from the screen. Plate current density is 6 milliamperes per sq cm, screen potential 150 volts. Electrons pass through between screen wires with an average kinetic energy of 140 electron volts. Plate voltage 50.

(a) Find how large d must be in order to insure space-charge suppression of secondary emission at this current density and this combination of plate and screen voltages.

(b) What will happen if the plate voltage drops below 50 volts, plate current and d still as in part (a).

(c) If the plate voltage rises above 50 volts?

(d) Below what value of plate voltage will this tube be unable to pass 6 milliamperes per sq cm to the plate because of space charge limitations?

15. A certain short-wave Barkhausen-Kurtz oscillator uses a tube in which the grid is 0.2 cm from the cathode, plate 0.3 cm from the cathode. $E_p = 0, E_g = +500$ volts.

(a) Find the frequency of the electron oscillation in cycles per second.

(b) The plate is moved to a location 0.4 cm from the cathode, and plate potential is lowered to 500 volts below that of the cathode.   Grid still 0.2 cm from the cathode and at +500 volts.   Determine the frequency of electron oscillation.

(c) What happens to the frequency of electron oscillation if the direct current to the grid increases?   Why?

16. Imagine a triode in which the potential distribution is like that in Fig. 61 except that the grid is midway between cathode and plate, and the potential distribution curve for the path between grid wires is an inverted parabola with its vertex at the grid plane.   Potential at the vertex is 100 volts, $s_p = 0.8$ cm.   Find the frequency of electron oscillation for this potential distribution, the electrons being assumed to start from rest at the cathode.

17. Derive an expression in the form of Equation (287) in which μ_{sg}, μ_{pg}, and d_1 are expressed in terms of tube dimensions, for a cylindrical screen-grid tube.   Both control and screen grid are assumed to consist of regularly spaced wires parallel to the axis of the tube. r_g, n_g, R_g are respectively the radius of the control grid cylinder, the number of grid wires per centimeter, and the radius of a control grid wire, just as in Fig. 25. r_{sc}, n_{sc}, and R_{sc} are the corresponding quantities for the screen, and r_c and r_p are cathode and plate radii respectively.   In order to do this, use two transformations, one applicable in the neighborhood of the control grid and cathode, the other in the neighborhood of screen and plate, then superpose the potential equations for the two transformations.

CHAPTER VII

THERMIONIC CATHODES

69. Electron-Emitting Efficiency of a Cathode Surface. The cathode surfaces of thermionic devices are made of special materials which when heated to high temperatures release electrons (thermionic emission) in very much the way that hot water releases water molecules (evaporation). The current in most high-vacuum apparatus is space-charge-
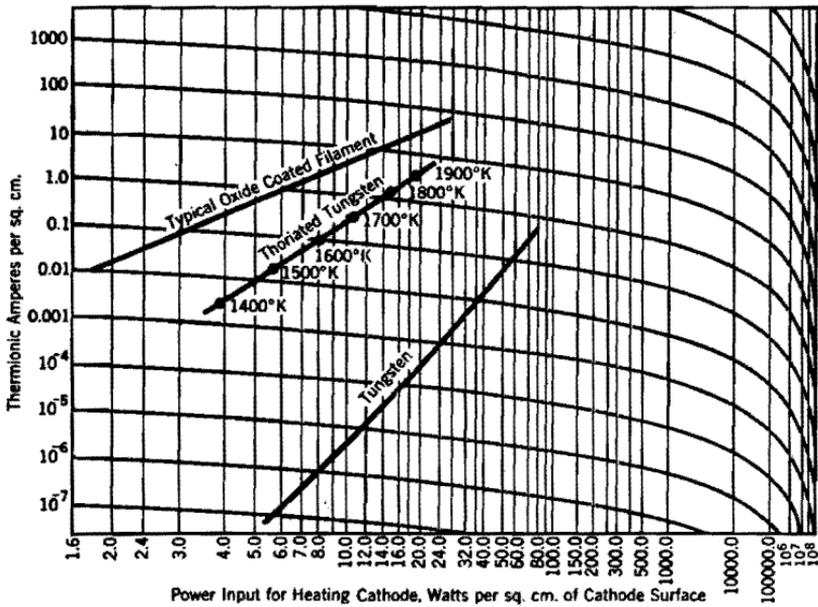
Fig. 63. Power-emission chart, coordinate warping based on Dushman's equation for thermionic current density. (See Section 77.)

limited, so that the basic requirement for a cathode is that it must be able to provide by thermionic emission at least the maximum space-charge-limited current demanded by the normal electrode potentials.

Cathodes are heated electrically; it is important to use materials and arrangements of parts that make the requisite thermionic current available with a minimum of heating power consumption. The more efficient cathode surfaces are somewhat fragile, having a limited life at best, and many of them are more or less easily damaged by the

148

presence of small amounts of gas in the tube.   Hence the element of judgment is important in selecting the proper type of cathode for a given service; power consumption and life requirements must be properly balanced.[E IV, V, S, T]

The overall electron-emitting efficiency of a cathode is measured by the ratio:

$$\frac{\text{Thermionic amperes available}}{\text{Power for heating, in watts}}$$

Both numerator and denominator may be expressed per square centimeter of cathode surface.   Fig. 63 illustrates the variation of thermionic current with power input for three different types of cathode surface, plotted on a specially warped set of coordinates which makes all the curves substantially straight lines.   These curves represent graphically the combination of two basic relationships, one between thermionic current density and temperature, the other between temperature and power consumption.   Section 79 contains a brief description of oxide-coated and thoriated cathodes.

70. Dushman's Equation Relating Thermionic Current Density and Temperature.   The mathematical expression of the first of the two relations just referred to is called Dushman's equation,[66, T24] derived later, in Section 120.   Dushman's equation is expressible in three forms that have identical meanings, as follows:

$$J_{th} = A_0 T^2 \epsilon^{-\frac{E_W e 10^7}{kT}} \tag{298 p}$$

$$J_{th} = A_0 T^2 \epsilon^{-\frac{E_W}{E_T}} \tag{299}$$

$$J_{th} = A_0 T^2 \epsilon^{-\frac{b_0}{T}} \tag{300}$$

in which

J_{th} = thermionic current density in amperes per sq cm.

A_0 = a semi-empirical constant; its units are:

$$\frac{\text{amperes per sq cm}}{(\text{degrees Kelvin})^2}$$

T = temperature of the cathode in degrees Kelvin (degrees Centigrade + 273).

k = Boltzmann's universal gas constant; its value is 1.372×10^{-16} erg per degree K.

e = the electronic charge, in coulombs.

E_W = an empirical constant of the emitting surface, called its *work function*, measured in electron volts, abbreviated to volts.

E_T = the *voltage equivalent of temperature*, defined by the relation

$$E_T e 10^7 = kT \qquad (301\text{ p})$$

which reduces numerically to

$$E_T = \frac{T}{11600} \qquad (302\text{ p})$$

b_0 = the temperature equivalent of work function, thus

$$b_0 = 11600\, E_W \qquad (303\text{ p})$$

The quantity A_0 is called semi-empirical because theoretical considerations indicate that it should have the same value for all homogeneous metal surfaces, a prediction which is reasonably well substantiated by experiment.[66] Theory also correctly predicts its order of magnitude. By a homogeneous metal surface is meant one whose behavior is not complicated by the presence of surface layers of different composition from that of the underlying metal.

The three forms of the exponent contained in Equations (298), (299), and (300) illustrate general types of exponents that are encountered repeatedly in electronic work. In the first form, Equation (298), the numerator of the exponent, $E_W e 10^7$, describes in ergs per electron the energy, in excess of that ordinarily possessed at .room temperature, which a favorably situated electron inside the hot metal must have in order to escape through the metal surface. This energy is roughly analogous to the latent heat of evaporation, per molecule, of a liquid at some temperature below the boiling point. For example, the equation relating the vapor pressure of the mercury vapor inside a mercury rectifier to the temperature of the pool of mercury in the base of the tube (see Table XVI) has an exponential factor similar in form to that appearing in Equations (298)–(300).

The *work function E_W* is a measure of the required energy of escape in electron volts; its value is usually between one and six or seven volts, varying with the material and processing of the surface. The quantity b_0, defined by Equation (303), is the work function measured in temperature units. Tables of emission constants sometimes give values of E_W, sometimes of b_0. See Table III.

71. The Voltage Equivalent of Temperature. The electrons, while within the metal and immediately after emerging from it, behave considerably like the particles of a gas heated to cathode temperature, in that they have randomly directed velocities of various magnitudes, the average velocity depending on the temperature. The quantity kT that appears in the exponent of Equation (298) is a definite amount of energy, in ergs per gas particle or per electron, which is characteristic

of the temperature T in a very fundamental way (see Chapter X). The corresponding characteristic random velocity will be called α, defined by the relation

$$\tfrac{1}{2}m\alpha^2 = kT \tag{304}$$

m being the mass of one particle. The voltage equivalent of temperature, E_T, is here defined as *the characteristic energy kT measured in electron volts,* as indicated in Equations (301) and (302). This meaning applies equally well whether the particles are charged or neutral, the electron volt being, it must be remembered, a unit of *energy.* The extent to which the energy kT, or E_T, characterizes the temperature cannot be fully appreciated without a study of theoretical thermodynamics,[U XXXVI] but it is possible to illustrate its importance in several specific ways, as follows:

(*a*) kT is the kinetic energy, in ergs, that is associated with the *most probable* total translational velocity of any particle in an ordinary gas at temperature T.[V 79, W II] This statement means that if one were to make a record of the magnitudes of the velocities, regardless of direction, of all the gas particles within a given volume at a given instant (a "snapshot observation throughout a volume"), more particles would be found to have velocities at or near α than similarly at or near any other value. It also means that if a random selection of one particle were made from the interior of the volume and its velocity noted, the chance of the velocity's being at or near α would be greater than for any other velocity.

(*b*) The familiar ideal gas law, $PV = RT$, can be written

$$PV = NkT \tag{305}$$

where

    P = pressure, dynes per square centimeter (baryes)*
    V = volume, cubic centimeters
    N = total number of gas particles in the volume
    kT = as above, the characteristic particle-energy for temperature T

(*c*) If the translational thermal motions of the particles of an ordinary gas are segregated into x-directed, y-directed, and z-directed components, the average energy associated with each of these three possible kinds or "degrees of freedom" of motion is $\tfrac{1}{2}kT$ ergs, or $\tfrac{1}{2}E_T$ electron volts. Hence the average total kinetic energy of each gas particle in monatomic gases, such as helium, neon, argon, mercury vapor, and sodium vapor, is $\tfrac{3}{2}E_T$, three times the value for each of the three degrees of freedom.[W II] Molecules of a diatomic gas are dumbbell-like and have two additional degrees of freedom of motion, one vibrational, one rotational, so that the average kinetic energy per particle is $\tfrac{5}{2}E_T$ volts. Thus at 40° C the average air-particle energy is $\dfrac{273 + 40}{11600} \times \dfrac{5}{2} = \dfrac{1}{15}$ electron volts.

* The *barye* is the pressure exerted by a force of one dyne uniformly distributed over one square centimeter. A pressure of one millimeter of mercury is 1333 baryes, and one barye is 0.752 micron of mercury (a micron is a thousandth of a millimeter). The barye is sometimes called a microbar, a bar of pressure being 10^6 baryes. Standard atmospheric pressure is 760 millimeters of mercury, and approximately 1,013,000 baryes, that is, 1.3 per cent more than 10^6 baryes.

(d) If an observer were to record the velocities with which each gas particle that arrives hits the wall of an enclosure for a definite area and period of time ("time exposure over a surface"; see Section 121), the average normal-to-the-surface energy at impact would be found to be just kT ergs, or E_T volts. This is greater than the overall average energy of normal-to-the-surface motion ($\frac{1}{2}E_T$) because the high-velocity particles move faster than the low-velocity ones, therefore hit more often and are counted more times in a second.

(e) If the time-exposure-over-a-surface observer in (d) above were to analyze his records of y- and z-directed impact velocities, both being parallel to the surface, their average energies would be just $\frac{1}{2}E_T$ each, for rapid y- or z-directed motion has no tendency to increase the number of impacts against a y, z surface. Hence the average *total* energy at impact on a surface is

$$E_T + \tfrac{1}{2}E_T + \tfrac{1}{2}E_T = 2E_T \tag{306}$$

All the relations described in (a), (b), (c), (d), (e) above can be shown to be true by means of the Maxwellian velocity distribution equations given in Sections 111–115.

72. Energies of Escaping Electrons. It has been experimentally demonstrated that as electrons emerge from a hot emitting surface their *random velocity distribution* is just that of the particles of a perfect gas. The average energy of emergence, that is

$$\frac{\text{Total energy brought through the surface per second}}{\text{Number of electrons emerging per second}}$$

has been found to be just $2E_T$, and is divided between the three component directions in accordance with Equation (306), and for the same types of reasons. Ordinarily only the normal-to-the-surface energies, whose average value is just E_T, are of importance.

For example, the electrons emerging from a cathode whose temperature is 2320° K have on the average the same *outwardly-directed* kinetic energy that one of them would gain by traversing a potential of $2320/11600 = 0.2$ volt. If the electric field just outside the cathode surface tends to force them back into it, and has a magnitude of 5 volts per centimeter, any electron possessing just the average energy travels outward, in spite of the field, to a distance $0.2/5 = 0.04$ centimeter, that is, until its kinetic energy vanishes. Having reached that point it must fall back into the cathode, acquiring an energy of 0.2 volt due to inwardly directed motion by the time it reaches the surface. Whatever energy of parallel-to-the-surface motion it has on emergence remains unchanged during the outward flight, reversal, and reentrance.

The most important aspect of this discussion of the energies of escaping electrons is the indication it gives that the velocities at emission are in all ordinary devices small relative to those acquired from the electric fields. Cathode temperatures usually range between 1000° and 2000° K, corresponding to energies between about 0.1 and 0.2 volt, whereas poten-

tials inducing flight between electrodes are many volts, and often many hundreds of volts. Initial velocities of emission are ordinarily of little or no consequence, but circumstances sometimes arise that give them considerable importance. They have been entirely neglected in the discussions of space-charge-limited currents in previous chapters.

73. Graphical Evaluation of Emission Constants. The graphical representation of the relation between thermionic current and temperature is usually based on a logarithmic expression. For example, Dushman's equation can be written:

$$\log_\epsilon J_{th} + 2 \log\left(\frac{1}{T}\right) = \log_\epsilon A_0 - b_0\left(\frac{1}{T}\right) \tag{307}$$

and experimental data relating temperature to thermionic current density represented by plotting $1/T$ horizontally and $\log J_{th} + 2 \log 1/T$ vertically. This results in straight-line relationships, of the type illustrated in Fig. 64, for a variety of substances. The corresponding values of work function and of the constant A_0 appear in Table III.

Fig. 65 is a somewhat more easily readable representation of the set of facts that appear in Fig. 64. The straight lines that represent the thermionic properties of individual types of surfaces are exactly the same in the two figures, but Fig. 65 has a direct-reading coordinate system. Its vertical coordinate lines are chosen and marked according to convenient temperature intervals, but are located in accordance with the uniform $1/T$ scale of Fig. 64. The warped, nearly horizontal, coordinate lines of Fig. 65 are similarly chosen and marked for convenient current density intervals, but are bent downward so as to correspond to constant values of current density in the coordinate system of the original figure. Each one corresponds to the graph, on the coordinates of Fig. 64, of an equation of the form

$$\left[\log J_{th} + 2 \log\left(\frac{1}{T}\right)\right] = \text{a constant} + 2 \log\left(\frac{1}{T}\right) \tag{308}$$

where the constant is a selected value of $\log J_{th}$. The tendency of the various straight lines to focus at a common point suggests the universal nature of the constant A_0, for $\log_\epsilon A_0$ for any line is the intercept at $1/T = 0$, or $T = \infty$, in Fig. 64. This intercept cannot be given an intelligible significance on the coordinate system of Fig. 65 because of the fact that $\log 1/T = -\infty$ when $1/T = 0$.

An early and still widely used alternative for Dushman's equation is known as Richardson's equation ^{T22}; its form, explained later in Section 130, is

$$J_{th} = aT^{\frac{1}{2}}\epsilon^{-\frac{b}{T}} \tag{309}$$
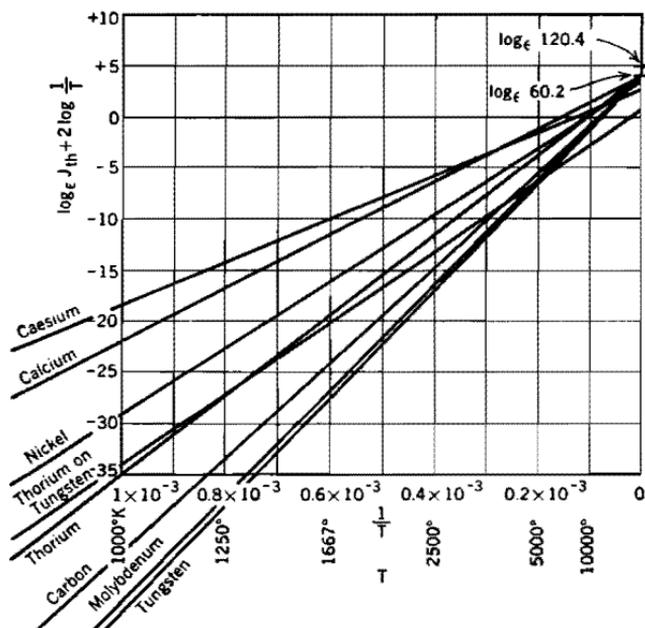
Fig. 64. Logarithmic temperature-emission chart, illustrating method of determining thermionic emission constants. For each line, the slope is b_0, and the infinite-temperature intercept is $\log_\epsilon A_0$. Plotted according to Equation (307).
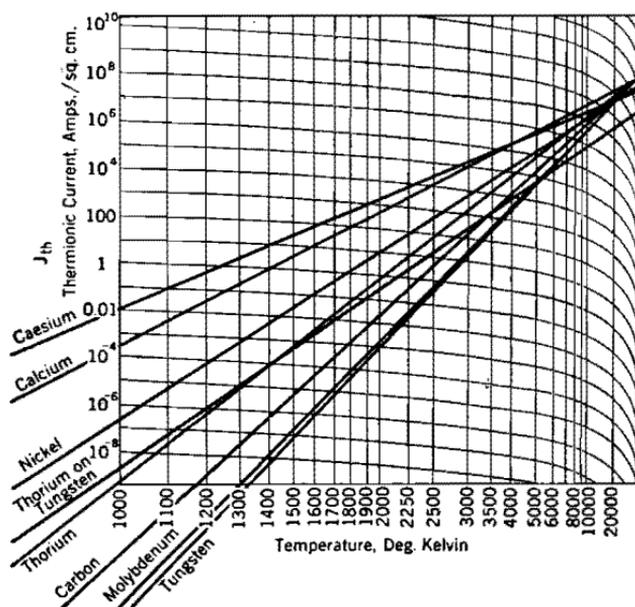
Fig. 65. Direct-reading temperature-emission chart with warped coordinate lines (compare with Fig. 64).

a and b being empirical constants. The exponent has the same general significance as in Dushman's equation, and b is, like b_0, the work function measured in temperature units. A graphical representation of Richardson's equation, similar to Fig. 64, would employ along the vertical scale the quantity $\log J_{th} + \frac{1}{2} \log (1/T)$.

Every set of experimental data so far obtained gives an equally satisfactory straight line on either type of graph, or for that matter on one in which the vertical scale is simply $\log J_{th}$, because the effect of the exponential factor completely overshadows that of T^2 or $T^{\frac{1}{2}}$. The thermionic properties of any emitter can therefore be completely described by a set of empirical constants corresponding to Dushman's equation, as in Table III, or by a corresponding set based on Richardson's equation, or by a set based on any reasonable value of the exponent of the T factor, including the zero exponent which removes the T factor altogether.

The present-day preference for Dushman's equation is based on rather definite theoretical indications that it is the correct form,[T] strongly supported by the convergence of the various straight lines of Figs. 64 and 65 toward a common point whose location is closely predicted by the theory. Present-day theory predicts a value of 120.4 for A_0, but experimental data favor a value of 60.

74. Cathode Power Dissipation. The second of the two relations that contribute to the power-emission relationship described by Fig. 63 is that between temperature and cathode power dissipation.[S, C XIV, X VII] The temperature of an electron-emitting cathode is of course that at which equilibrium exists between rate of heat removal and electrical heat input.

The power directly expended in causing electron emission is the product of emission current in amperes by the work function in volts. This amount of power describes the rate of removal of heat by the cooling action of electron emission. The mechanism is the same as that of cooling by evaporation; the particles that leave are high-energy ones, so that after the departure of any one of them the average kinetic energy of the body as a whole is less than it was before.

Although the heat input requirement necessary to make up this "electron evaporation" heat loss is measurable with refined apparatus, and measurements of it have been made for determining work function, it is in most high-vacuum thermionic devices negligible by comparison with the radiant heat loss. In an ideal emitter it would be a major part of the total heat-power input, yet would itself be small. Most present-day devices are far from the ideal, which can, however, be much more closely approached in gaseous-conducting than in high-vacuum devices.

Corresponding to the cooling action of electron departure from the cathode there is a heating action on entrance to the plate. Each electron delivers to the plate, when it strikes, energy corresponding to the plate voltage plus the plate-surface work function.

In actual high-vacuum apparatus nearly all the heat is removed by radiation, so that the relation between heat power input and temperature is controlled almost entirely by the heat-radiating and radiant-heat-absorbing properties of the electrodes. Convection is entirely absent, and the rate of heat removal by conduction through the supports and terminals, though measurable, is small relative to the power radiated.

75. Heat Transfer by Radiation; Emissivity Coefficients. The rate of radiant heat dissipation from a cathode is the difference between the radiation rate from the cathode to its surroundings and that from surroundings to the cathode; the latter is usually very small. The quantitative relation between the temperature of a radiating body and the power radiated away from it is, for a " gray " surface [X 206, X 199, C 106]

$$P_R = 5.73 \times 10^{-12}\gamma T^4 = 5.73\gamma\left(\frac{T}{1000}\right)^4 \qquad (310 \text{ p})$$

in watts per square centimeter, T being temperature in degrees Kelvin. The quantity 5.73×10^{-12} is a universal radiation constant (the Stefan-Boltzmann constant) and γ is the *emissivity coefficient* * of the surface.
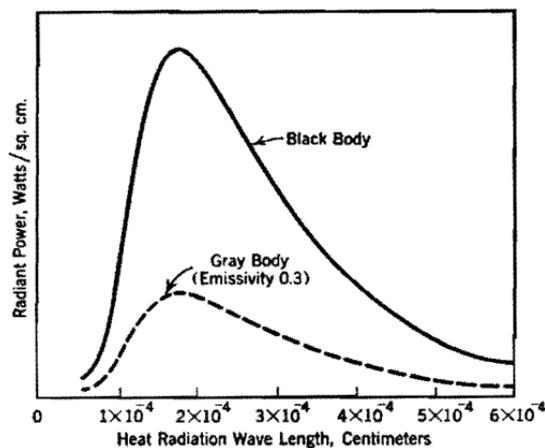
FIG. 66. Spectral distribution of black and gray body radiation, at 1600° K.

Radiation literature uses a great variety of symbols to describe this coefficient, which has a double significance; it is (*a*) the ratio of power *radiated* by a given gray surface to that radiated by a perfectly black surface at the same temperature, and (*b*) the fraction of incident radiation that is *absorbed* by a gray surface, this fraction being unity for a perfectly black surface. For a given temperature a truly black surface radiates more heat than any other kind, and absorbs all the radiation incident to it. Lampblack (soot) is the blackest substance known; its emissivity coefficient is about 0.98.

* The quantity here called the *emissivity coefficient* is called the *emission coefficient* by Moore,[C 106] and is called the *emissivity* by Richtmyer.[X 199]

Heat radiation, like light and radio broadcast radiation, takes place at a variety of wave lengths. The solid line in Fig. 66 illustrates the distribution of the energy radiated by a perfectly black body at one particular temperature. [X 258] This distribution curve has the same general shape for all temperatures, but the wave length corresponding to the crest varies inversely as the temperature.

The dotted line is the corresponding radiation distribution curve, at the same temperature, for a *gray body* whose emissivity coefficient is 0.3. Its ordinates are at every point 0.3 of the height of those for the black body. A gray surface is one for which this fraction is the same over the entire curve; if, for example, the dotted curve were at the left 0.1, and at the right 0.3 of the height of the solid one, the surface whose properties it described could not be called gray. The overall emissivity coefficient for a not-gray body has only a limited significance. It is reasonable to assume that all vacuum-tube electrodes have gray surfaces, but such an assumption in regard to the enclosing glass would be open to question.

Fig. 67 illustrates graphically, on logarithmic coordinates, the variation of P_R with temperature for several different surfaces.[S VI] The fact that the various lines are not all parallel to the straight black-body line indicates that the emissivity coefficients must vary slightly with the temperature. The value of the emissivity coefficient γ corresponding to any point on these curves is the ratio of P_R at that point to P_R for the black-body line at the same temperature.

As applied to the study of the incidence of radiant energy on a surface, the emissivity coefficient describes the fraction of such energy that is not reflected; that is, $(1 - \gamma)$ may be called a *reflectivity* coefficient. The energy that is not reflected may be absorbed within the body of the receiving substance, or it may pass on through it to remote regions.

In most ordinary electronic devices the distances to the heat-receiving surfaces are large relative to the dimensions of the radiating filament. Under such circumstances successive reflections between radiator and receiver make the overall radiant-energy transfer P follow the general law that is known to apply for concentric spheres or cylinders if the radius of the hot inner cylinder is much smaller than that of the outer one. This relation is

$$P = 5.73 \times 10^{-12} S\gamma(T_c^4 - T_p^4) \tag{311 p}$$

where S, γ, T_c, are respectively the radiator's area in square centimeters, the emissivity coefficient of both radiator and receiver, and the temperature of the radiator. T_p is the temperature of the receiver. This

can usually be simplified, for T_p^4 is ordinarily very much smaller than T_c^4, even when the receiving surface is at a moderately high temperature. Ordinarily the expression

$$P = 5.73 \times 10^{-12}\, S\gamma\, T_c^4 \qquad\qquad (312\ \text{p})$$

gives a satisfactory approximation to the true radiant power dissipation from the cathode.
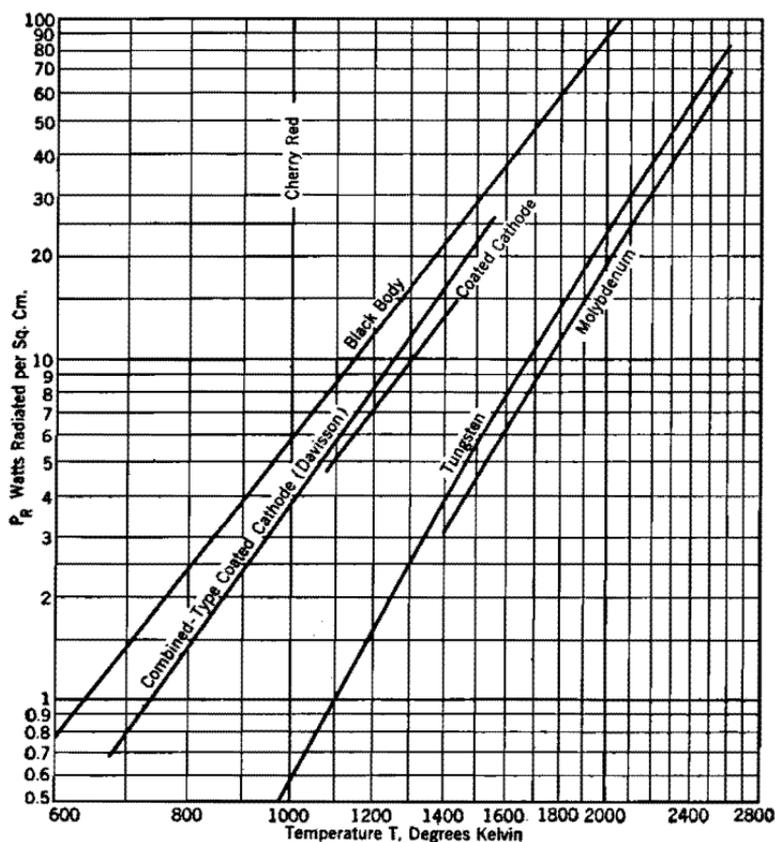
FIG. 67.   Radiation from thermionic surfaces.

Curves for platinum and tantalum would lie for the most part between those for tungsten and molybdenum.

For Davisson's combined-type coated cathode,

$$P_R = 5.735 \times 10^{-12}\,(0.4 + 2.5 \times 10^{-4}\, T)\, T^4$$

For tungsten (Worthing and Forsythe),

$$\log P_R = 3.680\,(\log_{10} T - 3.3) - \frac{1040}{T} + 1.900$$

If the cathode dimensions are not small relative to distances to the receiving surfaces, the analysis of the rate of radiant heat transfer from the cathode requires a study of multiple reflections for the particular

geometry involved. The approach to such problems is described by Moore.[C XIV]

76. Temperature Measurements; Lead Losses. Basic measurements of the temperatures of hot filaments are made by means of optical pyrometers, which compare the optical brilliance of the surface to be measured with that of a standard one.

In any attempt to use Equation (312) for accurate calculations the rather pronounced effect of "lead losses," that is, the cooling of the ends of the filament by conduction to the supports or lead-in wires, must be taken into account. Within this limitation, and subject to accurate knowledge of cathode dimensions and of the emissivity coefficient, Equation (312) can be used for the estimation of cathode temperature at a given power input. The emissivity coefficients of oxide-coated surfaces vary somewhat with the details of surface preparation, but that for thoriated tungsten is the same as for ordinary tungsten; in general, monatomic layers on a surface do not modify the emissivity. Numerical values of emissivity coefficients can be estimated from Fig. 67. A very smooth surface has a definitely lower emissivity than one that is even slightly roughened.

Temperature can be determined, if the resistance properties of the filament are known, by measurement of the ratio of hot to cold resistance; see Table IV. Koller [S VI] gives a good general discussion of methods of temperature measurement, and of the handling of lead losses.

77. Overall Relationship between Thermionic Current Density and Power. Since nearly all the heat supplied to the cathode of a high-vacuum thermionic device is removed by radiation, it is permissible, if lead losses are small, to equate electrical power input to cathode radiation, either total or per unit area. If E_f and I_f are root mean square values of heating voltage and current, and S the total heat-radiating surface, P_R, the radiant power dissipation *per square centimeter* of radiating surface is, in watts:

$$P_R = \frac{E_f I_f}{S} \tag{313 p}$$

This permits evaluation of P_R from experimental data.

Equation (312) can be written

$$\frac{1}{T_c} = 1.55 \times 10^{-3} \gamma^{\frac{1}{4}} \left(\frac{1}{P_R}\right)^{\frac{1}{4}} \tag{314 p}$$

If this expression for $1/T_c$ is used in Equation (307), the thermionic current density is expressible directly as a function of P_R. A convenient form of this relationship is

$$\log J_{th} + 2 \log\left(\frac{1}{P_R}\right)^{\frac{1}{4}} = \log \frac{A_0 10^6}{2.4 \sqrt{\gamma}} - 18 E_W \gamma^{\frac{1}{4}} \left(\frac{1}{P_R}\right)^{\frac{1}{4}} \tag{315 p}$$

Equation (315) suggests the use of experimental data in a manner similar to that employed in connection with Figs. 64 and 65. Numerical values of P_R and J_{th} are available from laboratory tests, so that corresponding values of $\log J_{th} + 2 \log \left(\dfrac{1}{P_R}\right)^{\frac{1}{4}}$ and of $\left(\dfrac{1}{P_R}\right)^{\frac{1}{4}}$ can be calculated for each individual observation. If the former are plotted vertically, the latter horizontally, an approximately straight line can be drawn through the points. The slant lines that appear in Fig. 63 were obtained in exactly this way. Their slight curvatures result from small variations in the emissivity coefficients with rise in temperature. The effect of these changes in γ on linearity is minimized by its occurrence to the $\frac{1}{4}$ power in the expression $18\ E_W \gamma^{\frac{1}{4}}$ for the slope. The intercepts with the infinite power axis on such a graph depend on γ as well as on A_0.

It is convenient here, as with Figs. 64 and 65, to have a direct-reading graph. Fig. 63 has a direct-reading coordinate system, the vertical lines being spread according to an inverse fourth power law, to agree with the fact that the points must, on rectangular graph paper, be plotted in terms of $1/P_R^{\frac{1}{4}}$. The nearly horizontal coordinate lines are bent so that each one represents, as in Fig. 65, a constant value of thermionic current density J_{th}.

The warping in Fig. 63 is based on Dushman's equation, which was used in building up to Equation (315). A useful kind of special graph paper, called "power emission paper" with coordinates similar to those of Fig. 63 but based instead on Richardson's $T^{\frac{1}{2}}$ equation, has been designed by the engineers of the Western Electric Company[67] and can be purchased from the Kueffel and Esser Company.

78. Inward-Radiating Cathodes. High emission efficiency is favored by a small rate of heat radiation from the cathode, for this permits the maintenance of a high cathode temperature with small energy consumption. The emissivity coefficient depends on the material and smoothness of the cathode; total heat radiation is dependent, not only on emissivity, but also on the geometrical shape of the cathode.

A cathode so shaped as to have some similarity to a furnace, with electron-emitting interior walls radiating heat toward one another, and having the nonemitting outer surfaces heat-insulated by two or three reflecting baffles, can be made many times more efficient than the ordinary outward-radiating kind.[S VI] The radiation from each part of the inner surface helps to maintain the temperature of the other parts, and very little radiation escapes to the outside.

Such heat-conserving cathodes can be used in gas-filled electronic devices in which the space charge in the interior region is neutralized

by positive ions of the gas.   Electric field effects in this interior region are much the same as those inside a metallic conductor, and the electrons can be drawn out through an open end of the cathode enclosure.   The radiating and emitting areas are not the same for this type of cathode, so that power-emission charts lose some of their significance.   Consider-able time (often a matter of minutes) must be allowed, after filament power is turned on, for them to reach temperature equilibrium before, plate current is allowed to flow.

Electrons that are emitted start out from a surface, in general, in the same direction as radiation.   In any high-vacuum space-charge-limited-current device the electrons must continue to travel in approximately parallel or diverging paths toward the receiving electrode.   If the receiving electrode is placed so as to be a satisfactory target for electrons, it is likely also to be a good target for heat radiation.   Consequently the inward-radiating heat-conserving cathode construction has not been applied successfully to high-vacuum thermionic devices.   Although electrons can be made to run around corners in a vacuum by suitable combinations of electric and magnetic fields, attempts to concentrate their emergence in a region out of the straight-line path of radiant-energy flow introduce current-restricting space charges.

79. Low-Work-Function Surfaces.   High emission efficiency is favored by low work function, and a great deal of ingenuity, scientific thought, and expense has been devoted to the search for materials and for process-ing methods which will produce very low work functions, with consider-able success.[EV, T, s]   Most present-day commercial electronic devices are equipped with *oxide-coated* or *thoriated* cathodes, prepared in accordance with the results of these investigations.   The surface structure and behavior of these special types of cathodes are described in Sections 103–107.   For some purposes cathodes of pure or nearly pure refractory metals, such as tungsten or molybdenum, are used.

Structural stability, including resistance to damage by bombardment of ions of residual gas, and other considerations (such as evaporation of surface material) that relate to useful life, must be considered along with work function and emissivity in comparing merits of various materials and processes.   Of course the melting point must be higher than the temperature at which appreciable emission takes place; hence the higher the work function, the more refractory a material must be to serve at all, regardless of efficiency.   This rules out copper, nickel, aluminum, and similar metals entirely.

PROBLEMS

Chapter VII

1. Find the values of E_T, the voltage equivalent of temperature as defined in Section 70, that correspond to 0° C, to 100° C, and to 2500° K.

2. (a) Use Dushman's equation to find what the *temperature-limited* current to the plate of the diode of Problem 3, Chapter V, would be with a filament temperature 2500° K. (Use Table III.)

(b) Above what plate voltage will the current be temperature-limited, if the radius of the plate is 30 times that of the cathode? (Use Fig. 47 or Table I.)

(c) Sketch approximately to scale, with scales marked, curves of plate current against plate voltage for this tube at a filament temperature of 2500° K.

(d) Same as (c), except at a filament temperature of 2000° K.

(e) Estimate the power, in watts, that must be supplied to this filament to maintain its temperature 2500° K; also, to maintain it at 2000° K.

3. The cathode of a certain cylindrical triode (Fig. 25) is being operated at a temperature such that the temperature-limited current density J_{th} is 60 milliamperes per sq cm. The actual space-charge-limited current density J_p is one-third of the temperature-limited current density. Cathode radius is 0.05 cm, plate radius 0.5 cm. Axial length of tube is 4.0 cm. $E_p = 130$ volts, $E_g = -10$ volts. The average *total* energy of the electrons as they leave the cathode surface is 0.2 electron volt. The work function of the cathode surface is 2.10 volts.
Find:

(a) The average *outwardly directed* energy of the escaping electrons.

(b) The temperature of the cathode.

(c) The value of A_0 in Dushman's equation, for the material of the cathode surface.

(d) The power required to heat the cathode if its emissivity is 0.4. (Assume that all the heating power input to the cathode is radiated to the plate.)

(e) The rate (in watts) at which *electron emission* removes heat from the cathode, in just the way evaporation cools a pool of water. Note that this rate is insignificant relative to that found in (d).

(f) The power, in total watts and watts per square centimeter, that must be radiated by the plate. (Note that this must include the power radiated to it by the cathode.)

(g) The temperature reached by the plate, if its emissivity is 0.9 (a graphite plate).

4. Consider a parallel-plane triode in which the cathode material, cathode temperature, ratio of J_p to J_{th}, and plate and grid voltages are the same as in the previous problem. The cathode radiates only from the side toward the plate. Answer questions similar to (d), (f), (g) of the previous problem, dealing now entirely in watts per sq cm. What do the results of the calculations indicate to you?

5. A plane tungsten cathode, area ten sq cm, is operating at 2500° K. One-half cm distant from it is an anode, also plane, at a potential 200 volts positive with respect to the cathode.

(a) Is the current space-charge-limited or temperature-limited?

(b) As the plate voltage rises from zero, at what value of plate voltage does the current first become temperature-limited?

CHAPTER VIII

WORK FUNCTIONS OF HOMOGENEOUS SURFACES

80. Ionizing Potentials of Atoms. An atom of any element consists of a central positively charged nucleus surrounded by a cloud or " atmosphere " of electrons, normally of just the right number to neutralize the positive nuclear charge.[W I, X 361] These electrons place themselves systematically in a series of " shells " located progressively farther out from the nucleus. Each shell can accommodate no more than a definite number of electrons, though it may contain less than the maximum, even when a more remote shell is partially occupied. The maximum number of electrons in each shell is $2n^2$ for the nth shell.[X 425]

Large atomic weight is usually associated with large nuclear charge, so that the electronic atmospheres extend into more remote shells for the heavy than for the light elements. None of the elements completely fill either the fifth or sixth shells, but some radio-active elements have one or two seventh-shell electrons in each atom. Table V gives what is believed to be the correct distribution of electrons among the shells for the various elements. The *atomic number* is simply the nuclear charge, in electron-charge units; hence also the normal number of electrons in each atom's atmosphere. Fig. 68 illustrates in a general way how the electrons in a sodium atom's atmosphere are distributed. There are two in the first shell, eight in the second, and a lone one in the third. The placement, in the figure, of the eight second-shell electrons at the corners of a cube is merely suggestive of a high degree of symmetry in their average positions.

An atom from which one of the outer-shell electrons has been removed is said to be *ionized*. The restraining force that tends to prevent such removal is the electrostatic attraction between the electron about to leave and the

FIG. 68. Locations of the eleven electrons of a sodium atom's "atmosphere."

remainder of the atom. Such attraction exists because, as an electron starts away, its departure leaves the atom positively charged by an amount equal to the electron's own negative charge. A definite minimum amount of energy, usually measured in electron volts and called

163

the *ionizing potential*, must be given to an outer-shell electron to enable it to escape in spite of the loss of kinetic energy experienced in overcoming this electric restraining force.   Thus the ionizing potential is a measure of the strength of attachment of an outer-shell electron to the atom of which it is a part.

For a given type of electron arrangement, the attachment of a remote-shell outer electron is less than that of a close-in one.   Thus the ionizing potential of sodium is 5.12 volts, of rubidium 4.16 volts, and of caesium only 3.87 volts, although in all three elements the last shell contains one, the next-to-the-last eight, electrons.

81. Free Electrons in Metals.   An electron that, like the eleventh one of sodium, has the entire outer shell to itself, is in general much less strongly attached to its atom than one, like a third-shell argon electron, which has many companions in its own shell.   Thus although the outermost electrons of both sodium and argon are in the third shell, the
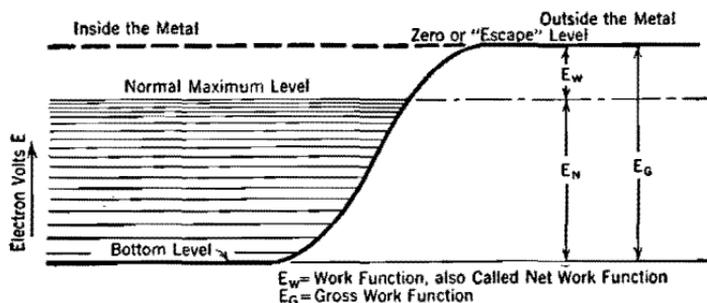
FIG. 69.   Energy levels for the electron gas in a cold metal (normal state).

ionizing potential of sodium is 5.12 volts, that of argon 15.69 volts. Good electrical conductors are made from the elements, scattered all the way through the periodic table, in which the farthest-out shell that is occupied contains only one, two, or — in a few elements — three electrons.

Atoms of these elements are located so close together when in solid metallic assembly that the outer-shell electrons are no more closely associated with one atom than with another.   The electron attachment to individual atoms is weak to begin with, and vanishes altogether when in this very close spacing.   At least one, sometimes two, in a few cases three, electrons per atom are not bound at all to any *one* atom, but are free to rove throughout the interior of the metal.   They can therefore move very rapidly in response to electric fields that result from the introduction of the metal into electric circuits; they are called *free electrons*.[TV, U536]

Since these roving electrons continually interchange their kinetic energies of motion with vibrational kinetic energies of the atoms of the

metal, the freedom of electron motion aids in the transfer of heat energy from one point to another; hence good electrical conductors are also good heat conductors.

82. Work Function. The free electrons possess varying amounts of kinetic energy, which for reasons detailed in a later section range upward from zero to a definite normal maximum, usually several electron volts. There are always more electrons with energies near the normal maximum than with lesser amounts. The highest-energy free electrons in a metal correspond roughly to the outer-shell electrons in a single atom, in that they are more nearly in position to escape from the metal than are any of the others.

Although there is no attachment of consequence between the free electrons and individual atoms in a metal, there is a strong attachment between the free electrons and the metal as a whole. An electron escaping from the surface of a conductor, like one on the way out from a single atom, leaves on the parent body a positive charge, called the *image* charge, which exerts an electric *image force* tending to prevent departure.

A definite minimum amount of energy, usually measured in electron volts and called the *work function*, already made use of in Chapter VII, must be given to any one of the highest-energy free electrons to enable it to escape in spite of the loss of kinetic energy experienced in overcoming the image force. The addition to low-energy free electrons of an amount of energy equivalent to the work function cannot release them, for it is an electron's total energy to which the possibility of escape is primarily related. Work functions of metals range between 1.5 and 7 volts (see Table III) as compared with a range of 3.5 to 25 volts for the ionizing potentials of atoms and molecules in gases and vapors.

83. Energy-Level Diagrams; Gross and Net Work Function. As an escaping electron moves away from the metal its kinetic energy is reduced to just the extent that its *potential energy* is increased; similarly the kinetic energy of a baseball in an upward flight is decreased to the same extent that its potential energy, measured by its height, grows larger. The electric-force barrier offered to an electron's remotely similar outward flight may be represented diagrammatically in an *energy-level diagram*, Fig. 69,[U 540] in which vertical distances describe energies.* It is convenient to use the *electron volt* as a unit of energy in such diagrams, both potential energy and kinetic energy being convertible between ergs and electron volts E by the familiar relation

$$Ee10^7 = \text{ergs} \tag{316 p}$$

---

* Herzfeld[63] gives an extensive bibliography of articles relating to the origin and uses of the energy-level concept as applied to the electrons in metals.

The upward reverse curve in the heavy solid line in Fig. 69 represents the change in an outwardly moving electron's potential energy from a uniform " bottom-level " value everywhere inside the metal at the left to a uniform " zero-level " value everywhere outside at the right. Thus the *potential energy* is *zero outside, negative inside.* The difference between these two extreme zero and bottom *energy levels* is just the increase in potential energy, hence also the loss in kinetic energy, experienced by an escaping electron during its outward flight. This difference describes the kinetic energy that must be imparted to a *stationary* electron inside the metal in order to permit escape, and will be called the *gross work function* in this text. It is considerably larger than the quantity heretofore called work function. Sections 92–95 contain discussions of the factors that determine gross work function magnitudes.

Since the electrons in the metal are not stationary, the addition of a considerably smaller amount of energy than the gross work function may result in one's escape. The many horizontal lines in the left, interior-of-the-metal, part of Fig. 69 represent *kinetic-energy levels.* They symbolize the fact that the free electrons in the metal possess a great variety of kinetic energies ranging from none at all (for the " bottom-level electrons ") up to that described by the normal maximum level. The quantity described in the preceding section and in Chapter VII as work function, and universally referred to in that way, is the additional kinetic energy that must be given to an electron *already in the normal maximum level* to enable it to escape. This is of course just the difference between gross work function and the kinetic energy of the electrons in the normal maximum level.

Work function will in this text sometimes be called net work function to emphasize the contrast with gross work function, but there need be no ambiguity ever as to the meaning of the simple unmodified phrase " work function."

The height of the normal maximum level is dependent entirely on the concentration of free electrons within the metal, in a manner that is fully discussed in Sections 84–91. Available evidence indicates that it ranges from about 2 volts in some of the lighter metals to 10 or more in some of the heavier ones. It is interesting to note that the *average kinetic energy* of the " gas " made up of the free electrons in a metal is several volts ($\frac{3}{5}$ of the normal maximum) even at very low temperatures. This is in sharp contrast to the average molecular kinetic energy in ordinary gases, which rarely exceeds $\frac{1}{2}$ to $\frac{3}{4}$ of a volt, even at high temperatures.

84. Normal (Low-Temperature) Distribution of Kinetic Energy. It has been stated that the many horizontal lines in the left, interior-of-

the-metal, part of Fig. 69 suggest the many different values of kinetic energy possessed by the free electrons. The actual number of electrons that have a given amount of kinetic energy is controlled, for reasons outlined in a later section, by the following two general sets of facts: [69, 70, 71, T V]

(a) The values of kinetic energy possible for an electron to have differ from one another by finite though extremely small amounts. Each permissible value can be represented diagrammatically as a *kinetic-energy level;* hence the use of discrete horizontal lines in the diagram.

(b) Each kinetic-energy level can accommodate only a limited number of electrons, the number increasing approximately as the square root of the energy. This is true for the same kind of reason that limits to a definite maximum the number of electrons in any one shell of an atom's electronic atmosphere. Each kinetic-energy level may be empty, partly filled, or completely filled, just as each of the shells around an atomic nucleus may be empty, partly filled, or completely filled.

Absolute zero ($0°$ K) is the temperature at which all particles, including the free electrons, have the least possible energy. The electrons cannot all have bottom-level kinetic energy, however, for at that level there is no to-and-fro motion, and it happens that there can be only two such electrons. The levels immediately above the bottom can accommodate only a few electrons each, and the number available per level grows rather slowly. *Most of the electrons must " occupy " levels that are several volts above the bottom.*

Least-possible energy does require, however, that no electron shall be able to find a vacancy in a level lower than the one it occupies. This means that at absolute zero all levels must be completely filled as far up as is necessary to accommodate all the electrons, and that all levels above that height must be completely empty. The " normal maximum " level in Fig. 69 is just the level so described; all the electrons, but no more, can be accommodated below it.

The discussion just given has been based on conditions at absolute zero temperature. It happens, however, that because of the very small mass and very high concentrations of the free electrons, very few of them indeed occupy levels above the normal maximum in metals heated even considerably above ordinary room temperature. Hence at " normal " room temperatures the distribution of electrons among the levels (" kinetic-energy distribution ") is practically the same as that at absolute zero, and the adjective " normal " is appropriate for absolute-zero conditions. To the electron gas in a metal the change from $0°$ K to $300°$ K appears a trifling variation.

85. Thermionic Emission. It is sometimes helpful to think of the bottom level in Fig. 69 as the bed of a lake, the normal maximum indicating the normal water level, and the zero or escape level the height of the

land around the lake.[T 22]   As long as there is no wind, all the space below the normal water level is full of water, and that above it empty, for exactly the same reason that at absolute zero all the kinetic-energy levels below the normal maximum are filled and those above it empty: that is, in the absence of agitation the water must be so distributed that no water particle shall be able to find a vacant space at an elevation lower than the one it now occupies.

Heating the metal has an effect similar to that of agitating the water. Wind or some other disturbing factor may produce waves or spray; but waves or spray can exist only if some of the normally filled space is vacated in order to provide water for the spray drops or wave crests. If the agitation is severe enough some of the spray rises above the escape level and passes out over the land (thermionic emission) and can be drawn away to other lakes (other electrodes) if there is a slope (electric field) in the terrain.   If there is no slope in the terrain, the water carried out onto the perfectly level shore piles up there (space charge) with the result that a water-surface gradient (electric field set up in accordance with Poisson's equation) makes some of the water flow back into the lake.   This analogy is imperfect and must not be carried too far, but it can be very helpful.

Thermionic emission takes place only when thermal agitation throws an appreciable proportion of the high-kinetic-energy free electrons up into normally vacant levels that are *above the escape level.*   It is evident from this requirement why the important energy element in thermionic emission is the net rather than the gross work function.   Dushman's equation grows out of a mathematical statement of the relation between temperature and the number of electrons that occupy high-up normally vacant levels.

Fig. 71 on page 175 illustrates the distribution of free electrons among the kinetic-energy levels in tungsten,[69] assuming that each tungsten atom contributes two free electrons, for (*a*) 0° K ("normal") and (*b*) 2500° K, the temperature at which pure tungsten filaments are commonly operated.   Ordinates describe the relative number of free electrons occupying levels located according to the various kinetic-energy values.   The zero-temperature curve is a true horizontal-axis parabola that terminates abruptly at the normal maximum energy value.   It will be noted that the *average energy* of the free electrons changes very little between 0° and 2500°.

86. Why Kinetic-Energy Levels Have Finite Spacings; the "Quantum" of Action.   It is not difficult to state the reasons why the available kinetic-energy levels within a metal are separated by finite amounts of energy and can each accommodate only a limited number of free electrons.

Such a statement has the merit of introducing principles of electron behavior that are important in studies of photoelectric action and of the properties of gaseous electrical discharges.

Electric charge occurs only in integral multiples of the electronic charge; mass likewise only in integral multiples of minute units. Still another physical quantity known as "*action*," which is of importance in periodic motion, occurs only as integral multiples of a very definite small amount.

The *action* of a particle in periodic motion is the line integral of the momentum taken over one cycle of movement. Thus the action associated with the periodicity of a particle of mass m moving with uniform velocity v in a circular path of radius a is [X 364, W 22]

$$\int_{s=0}^{s=2\pi a} mv\, ds = 2\pi mav \qquad (317)$$

The symbol h will, in accordance with universal practice, be used to represent the indivisible unit of action. $h = 6.542 \times 10^{-27}$ erg-second; h is called "Planck's constant," or sometimes "Planck's quantum of action."

Action is described in erg-second units because its dimensional formula (ML^2T^{-2}), happens to be obtainable by multiplying that for energy (ML^2T^{-3}) by a time-factor T. It is, however, often desirable to think of action as the product of momentum by distance traveled in the direction of the momentum, rather than as the product of energy by time.

The fact that the action of a circularly moving particle must be an integral multiple of h is expressed by the relation

$$2\pi mav = nh \qquad (318)$$

n being any whole number.

An entirely different type of periodic motion in which the action must be "quantized" is that of a particle shuttling back and forth between two parallel-faced walls of an enclosing rectangular box, from whose inside surfaces the particle rebounds without loss of energy. For the sake of simplicity the box may be supposed to be a hollow cube. The particle might be one of the free electrons within, and the walls the work-function barrier around, a piece of metal or a metallic crystal. Such a particle's velocity component parallel to one edge of the cube is constant in magnitude but reverses periodically. The action limitation requires that the momentum times cyclic length of path must be an integral multiple of h, thus

$$mu \cdot 2a = n_u h \qquad (319)$$

where u is the velocity parallel to one edge of the cube, m is the mass of the

particle, a the length of one edge of the cube, and n_u any whole number.

With a cube of given size this limits the velocities u to integral multiples of $h/2ma$; for convenience this quantity will be called $1/M$, so that

$$M = \frac{2ma}{h} \tag{320}$$

There are of course three expressions like Equation (319) as follows:

$$2mau = n_u h \tag{321}$$

$$2mav = n_v h \tag{322}$$

$$2maw = n_w h \tag{323}$$

v and w being velocities in the two directions normal to the u motion and to each other. Thus three distinct *quantum numbers*, n_u, n_v, n_w, are required to specify any particular type of motion within the cube. The total or resultant velocity c of course satisfies the relation

$$c^2 = u^2 + v^2 + w^2 \tag{324}$$

This suggests the use of a quantity n, here called the *resultant quantum number*, related to these other quantities as follows:

$$n = Mc \tag{325}$$

and

$$n^2 = n_u^2 + n_v^2 + n_w^2 \tag{326}$$

The total translational kinetic energy is of course $\frac{1}{2} mc^2$; this can also be expressed as follows in terms of n^2:

$$\frac{m}{2M^2} n^2 = \text{total translational kinetic energy of the particle, in ergs} \tag{327}$$

Since n_u, n_v, n_w must be whole numbers, n_u^2, n_v^2, n_w^2, and their sum n^2 must also be whole numbers. One or more sets of values of n_u, n_v, n_w can in fact be found to fit most whole-number values of n^2. Free electrons within a metal may therefore possess kinetic energy values corresponding to almost any whole-number multiple of the quantity $m/2M^2$; this amount of energy when expressed in electron volts is just the " finite though small " difference, referred to in (a), Section 84, between the heights of adjacent kinetic-energy levels of Fig. 69.

87. Why Each Kinetic-Energy Level Can Accommodate Only a Limited Number of Electrons; the "Exclusion Principle." It was stated in Section 84 that " each kinetic-energy level can accommodate only a limited number of electrons." It has just been pointed out that the movements of the particles in the cube, or of electrons in the metal, must be " quantized," so indicating discrete, separate values of transla-

tional kinetic energy.   In addition, what physicists call the " exclusion principle " must be accepted;[69, T 242] this requires that *no two particles within the cube can have an identical set of quantum numbers.*   The exclusion principle and the limitation of cyclic action to whole-number multiples of *h* are adopted, not as outgrowths of any abstruse philosophical concepts of matter, but *because if consistently used they give solutions, that are verifiable by experiment, to an immense number of complicated problems in remotely scattered fields of scientific work.*

The energy corresponding to any given type of motion depends on n^2, according to Equation (327); and the number of different particles that can be accommodated in a given kinetic-energy level depends, according to the exclusion principle just stated, only on the number of
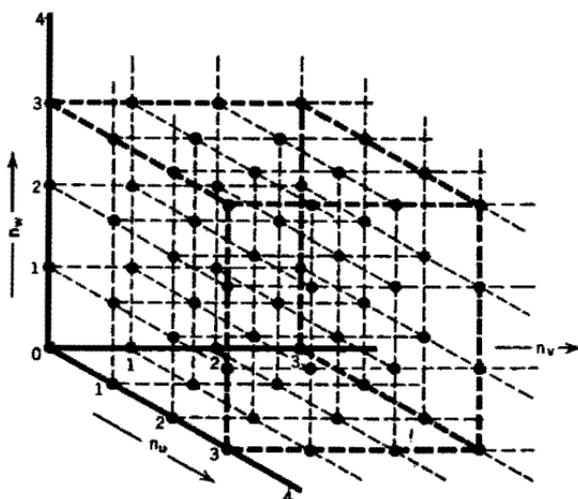
Fig. 70.   Quantum-number lattice.

different ways in which n_u, n_v, n_w can combine to give a particular value of n^2, as listed in the column entitled "number of possible arrangements " in Table VI. For example, $n^2 = 2$ can correspond to (*a*) $n_u = 1$, $n_v = 1$, $n_w = 0$, (*b*) $n_u = 1$, $n_v = 0$, $n_w = 1$ and (*c*) $n_u = 0$, $n_v = 1$, $n_w = 1$. Thus the kinetic-energy level corresponding to $n^2 = 2$ can accommodate just three different kinds of translational motion simultaneously.

Fig. 70 is a three-dimensional diagram of the various possible quantum-number combinations.   There is a scale of n_u-values along the forward axis, of n_v-values along the horizontal axis to the right, and of n_w-values upward along the vertical axis.   The octant bounded by the positive portions of these three axes contains a set of many points, one point being located at each possible combination of values of n_u, n_v, n_w. These quantum numbers have all possible whole-number values, and the points represent what is called a " cubical lattice."   Each point corre-

sponds to a particular kind of motion; the energy for each kind is measured by the square of the resultant quantum number n, which is the radius vector from the origin to the lattice point that identifies that particular motion; see Equations (326) and (327). The sixth column in Table VI states how many lattice points are located at the radius n given in the fifth column.

The exclusion principle is a statement of this fundamental fact of nature: it is as impossible for any two of these particles in a box to have simultaneously identical velocities as for them to have identical positions. Human sense perceptions make the positional restriction obvious enough, but do not permit a ready, off-hand grasp of the velocity restriction.

Properly to understand this principle the concept of individuality among electrons or gas particles must be completely abandoned. The box contains, not so many particles, but merely a certain number of identifiable and unduplicated positions, an equal number of identifiable and unduplicated velocities (more strictly, momentum values), and a certain total kinetic energy. Coordinates and momenta have definite, differing values described by numbers that cannot be mistaken for one another; not so with the particles. Any one board in a lumber yard can be permanently identified by sufficiently close examination, but there is no such thing as identifying an individual electron or atom.

88. Electron Spin. As indicated by the last column in Table VI, there can in fact be *two* electrons in the cube, not just one, with each possible type of translational motion. Each electron has *four* types of periodic motion, as follows: (a) x-directed shuttling, (b) y-directed shuttling, (c) z-directed shuttling, and (d) an *electron spin*,[U 443, W 7] similar to the spin of the earth on its own polar axis. Only two kinds of spin-motion are possible; each electron may spin clockwise, or it may spin counterclockwise. The angular velocity of either type of spin is just such that a change to the other results in a change of one action-unit h. The spin quantum number is always $\frac{1}{2}$, thus

$$\frac{1}{2}h \text{ clockwise} + h = \frac{1}{2}h \text{ counterclockwise}$$
and
$$\frac{1}{2}h \text{ counterclockwise} + h = \frac{1}{2}h \text{ clockwise} \tag{328}$$

Since there are two possible spin quantum numbers, $\frac{1}{2}$ clockwise, and $\frac{1}{2}$ counterclockwise, two electrons, one with each kind of spin, can have each possible type of translational motion without violating the requirement that no two particles can have an identical set of quantum numbers. The number of electrons that can be accommodated in each kinetic-energy level is therefore just twice the number of lattice points at the corresponding radius in Fig. 70.

It is probable that the mental picture of a box full of shuttling, spin-

ning particles is as remote from the true mechanism inside a metal or metallic crystal as is the picture of individual flux lines threading through a coil from the true nature of a magnetic field.   Yet both concepts are useful because, if properly employed, they point the way to correct solutions of scientific and engineering problems.

89. Actual Energy of the Normal Maximum Level.   Since the lattice points of Fig. 70 are everywhere spaced one n-unit apart, each one is located at the center of a cube whose volume is one cubical n-unit; hence the number of lattice points within any restricted part of the figure is simply the volume of that region in cubical n-units.   This is especially true when the region considered is so very large that any question as to whether its boundary passes through the middle or the edge of a particular row of cubes has an inappreciable effect on the determination of the total volume.   This happens to be true for all interesting cases.

Invariably the values of n considered in any specific problem are tremendously large, and the energy-unit spacing $m/2M^2$ between individual kinetic-energy levels is almost unbelievably small, relative to the total range of energy values.   So it is convenient, as in many somewhat similar scientific and technical analyses, to ignore the discontinuity between levels in setting up a mathematical treatment.   This is just as permissible as it is to work with space-charge density ρ.   All space charge is of course made up of individual electrons or ions, yet it is not necessary to locate them individually in order to predict their effect on potential by means of Poisson's equation.

The symbol E_N will be used to describe the normal maximum energy in electron volts, and n_N to describe the corresponding resultant quantum number;   they are related in accordance with Equation (327), that is:

$$\frac{m}{2M^2} n_N^2 = E_N e\, 10^7 \tag{329 p}$$

This expression is made valid for any values of E and n simply by dropping the subscripts. n_N is invariably a tremendously large number. The absolute zero or " normal " condition of the free electrons within a metal is that at which all kinetic-energy levels below the maximum are completely filled.   As applied to the lattice-point picture, Fig. 70, this means that all of the lattice points, out to a very large radius n_N, are " occupied " by their full quota of electrons, that is, two for each lattice point.   The number of points within the n-space octant that is thus completely filled is of course one-eighth of the volume, in cubical n-units, of a sphere of radius n_N.

Let N symbolize the number of free electrons per cubic centimeter of real volume of the metal;   of course the real volume of the cubical piece

of metal or metallic crystal being considered is a^3 cubic centimeters. Hence the total number of electrons to be accommodated below the normal maximum kinetic-energy level, so within the n-space spherical octant of radius n_N, is Na^3. Therefore

$$Na^3 = 2\frac{1}{8}\left(\frac{4}{3}\pi n_N^3\right) = \frac{2\pi n_N^3}{6} \tag{330}$$

The electron spin accounts for the factor 2. This expression can be rearranged into the form

$$n_N = a\left(\frac{6N}{2\pi}\right)^{\frac{1}{3}} \tag{331}$$

If this expression for n_N is introduced into Equation (329), and Equation (320) used to eliminate M, a solution for the normal maximum kinetic energy E_N is obtained, as follows: [T 246]

$$E_N = \frac{h^2}{8em10^7}\left(\frac{6N}{2\pi}\right)^{\frac{2}{3}} = 5.73 \times 10^{-15}\left(\frac{N}{2}\right)^{\frac{2}{3}} \tag{332 p}$$

The factor 2 under the N represents the effect of electron spin.

Thus the normal maximum energy is found to depend only on N, the number of free electrons per cubic centimeter, aside from the universal constants h, e, m, π. For tungsten, E_N is 9.17 volts, assuming that tungsten has two outer-shell, therefore free, electrons per atom (see Table V). For caesium, with one free electron per atom, E_N is 2.02 volts. The number of atoms per cubic centimeter is found by dividing the density, in grams per cubic centimeter, by the mass per atom, as determined from the atomic weight.

90. Normal Average Energy. The average kinetic energy of the electrons within a metal under normal (absolute zero) conditions is obtainable by the usual averaging procedure. Any solid sphere may be thought of as consisting of a very large number of very thin concentric spherical shells; similarly in the n-space region the octant of the large sphere that normally accommodates all the electrons may be thought of as made up of octants of spherical shells, each of radius n and thickness dn. Each shell is presumably thin enough so that dn is minute by comparison with n; yet thick enough so that the number of electrons to be found in a small part of it is reckoned in millions. A similar situation existed in regard to Fig. 6. The space-charge layer there illustrated is so thin as to occupy an almost infinitesimal part of the total region between the electrodes; yet in order to permit use of the space-charge-density concept the layer must be thick enough so that the number of electrons within a small portion of it is reckoned in millions.

The kinetic energy of each electron is $n^2 \dfrac{m}{2M^2}$, from Equation (327).

The number of electrons within a given shell, symbol dN_S, is twice the volume of the shell in cubical n-units (twice because spin permits each lattice point to accommodate two electrons); that is

$$dN_S = 2 \times \tfrac{1}{8} \times 4\pi n^2\, dn = \pi n^2\, dn \tag{333}$$

The average energy is determined by multiplying the energy per electron for each shell by the number of electrons within the shell, adding all such products by a process of integration between $n = 0$ and $n = n_N$, and dividing by the total number of electrons as given by Equation (330). Mathematically:

$$\text{Average Energy} = \dfrac{\displaystyle\int_{n=0}^{n=n_N} \dfrac{m}{2M^2} n^2 \cdot \pi n^2\, dn}{\dfrac{\pi n_N^3}{3}} = \dfrac{3}{5}\dfrac{m}{2M^2} n_N^2 \tag{334}$$

Thus the normal average energy is three-fifths of that of the normal maximum level, that is, $\tfrac{3}{5}E_N$, in electron volts.

91. Normal Energy Distribution. Fig. 71a illustrates the normal *distribution of energy* among the various levels. The first step in determining the shape of this distribution curve is to express the number
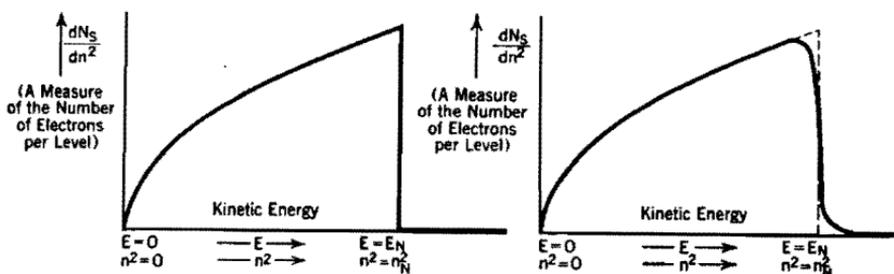
(a) Normal distribution, 0° K.   (b) 2500° K.

FIG. 71. Energy distribution curves for the electrons within tungsten.

dN_S of electrons in each shell in terms of the energy for the shell, as measured by n^2, rather than in terms of the total velocity as measured by n; see Equation (333). The mathematical process is as follows:

$$dN_S = \pi n^2\, dn = \pi n \cdot n\, dn = \pi \sqrt{n^2}\, \tfrac{1}{2}\, dn^2 \tag{335}$$

The distribution curve in Fig. 71a describes the variation with energy of the ratio

$$\dfrac{\text{Number of electrons in a shell}}{\text{The corresponding differential increment of energy}}$$

If the energy is measured by n^2, the differential increment of energy is measured by dn^2, so that the equation of the distribution curve is:

$$\frac{dN_S}{dn^2} = \frac{\pi}{2}\sqrt{n^2} \qquad (336)$$

A graph of dN_S/dn^2 plotted vertically, n^2 horizontally, in accordance with this equation is a parabola with a horizontal axis, vertex at the origin, as in Fig. 71a. It drops abruptly to zero at $n^2 = n_N^2$, for the shells beyond $n = n_N$ are all vacant; dN_S for them is zero. Since the ordinates in such a curve may usually be purely relative, they can be laid out to an electron-volt energy scale by selecting an arbitrary ordinate at $E = E_N$, corresponding to $n^2 = n_N^2$ then making all other ordinates vary according to the square roots of the corresponding electron volt values on the energy scale. Formal conversion of Equation (336) to electron-volt energy units is accomplished by substituting for n^2 and dn^2 their equivalents in terms of E and dE, as obtained from Equation (329).

A curve of *distribution of velocities* describes the variation with *velocity* of the ratio of the number of electrons in a shell to the corresponding differential increment of velocity. The equation for a velocity distribution curve, using n as the measure of total velocity, is directly obtainable from Equation (333); it is

$$\frac{dN_S}{dn} = \pi n^2 \qquad (337)$$

The unit of measurement along the horizontal axis is in this case n, so that the distribution curve is a parabola with vertex at the origin, but with the axis *vertical;* that is, at right angles to the velocity scale. Such a curve appears in a later figure on page 225 in connection with a general discussion of velocity distributions in gases.

92. The Outward Flight of an Electron; the Image Force. The atoms of a metal will here be referred to as ions, for each one has contributed one or more electrons to the make-up of the gas of free electrons that pervades the interior. The structural properties of a metal are of course due to attachments between the ions, which are arranged according to some regular crystalline pattern. The bottom energy level in Fig. 69 is made horizontal in order to indicate the equipotential nature of the interior, which contains no space charge because free electrons and fixed ions are present in equal concentrations.

On the way out from the interior an escaping electron passes through three more or less distinct regions, distinguished by the nature of the predominant force action in each, as follows: [66, 72, 73, T 21]

(1) It must pass between the adjacent members of a last layer of ions. While passing through this layer, and for a very short distance beyond it, the important forces are those due to the near-by ions.

(2) At a distance from the last layer that is not much greater, possibly less, than the spacing between ions within the metal, the forces due to individual particles become indistinguishable and the metal takes on the aspect of an equipotential surface. The electric field between an equipotential surface and an electron x centimeters in front of it is, as illustrated in Fig. 72, exactly the same as that in the front half of the region between an electron in free space and an equal and opposite *image charge* located $2x$ centimeters from the electron. For this reason the restraining force is often called the *image force*. It has the value

$$f = -\frac{e^2}{4x^2} \quad \text{(dynes per electron)} \quad \text{(338 esu)}$$

within this second region in which the electron reacts to the metal as to an equipotential surface. The negative sign indicates that the force is inwardly directed, that is, it tends toward a decrease in x.

(a) True Field

(b) Image Field

Fig. 72. The electric field around an escaping electron.

(3) There may be an external electric field aiding escape, that is, tending to draw electrons away from the metal to other electrodes (anodes). Such a field, of strength F statvolts per centimeter, exerts an outwardly directed force eF dynes. At a definite value of x, which will be called x_m, this withdrawal force is just equal and opposite to the restraining force, the net force being zero. If F is uniform, the "escape distance" x_m is determinable from the force equation

$$\frac{e^2}{4x_m{}^2} = eF \qquad \text{(339 esu)}$$

Any electron whose kinetic energy of outwardly directed motion vanishes before it reaches the distance x_m from the surface must fall back into it. Any electron that still possesses some outwardly directed velocity after coming that far, enters a third region in which the external field force predominates, and is accelerated toward a more positive electrode. If the external field force is zero or negative, electrons that escape do so without passing through such a third region.

Fig. 73a illustrates the variation between the restraining force in

dynes and distance from the surface.   This is the entire force experienced by an outwardly moving electron in the absence of any external field (no third region present).

It has been experimentally demonstrated that the force curve follows the image-force law within the second region.   The shape of the force
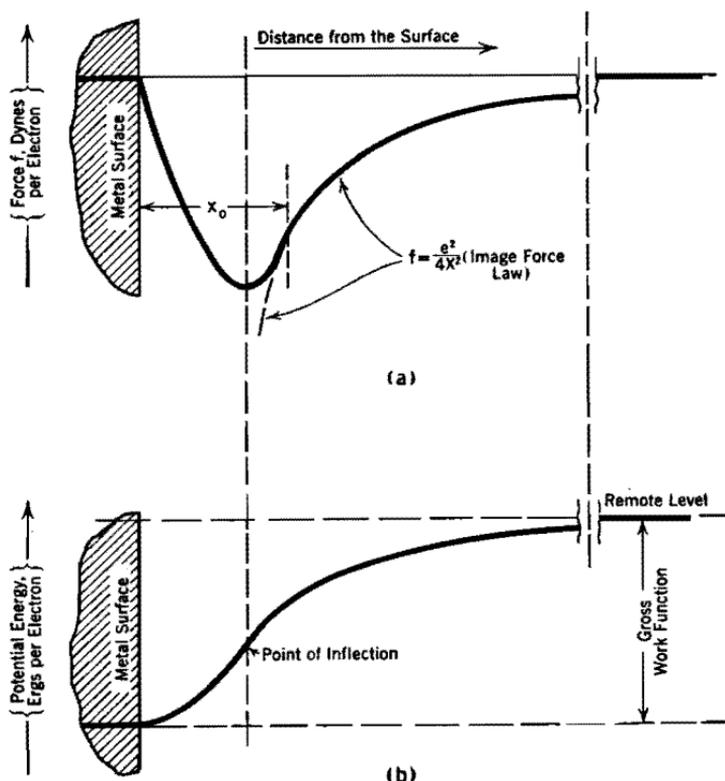
FIG. 73.   Force relations during electron escape from a metal.
(a) Restraining force f on an escaping electron.
(b) Potential energy barrier at a metal surface.

$$Ee = \int f \, dx$$

curve in the first region close to the surface, where force actions of individual ions predominate, can only be estimated, from general principles that are discussed in a later section.   It presumably approaches zero for small values of x.

93. Relation between Potential-Energy Curve and Force Curve. Fig. 73b is a diagram very similar to Fig. 69; however, it indicates measurement of energy in ergs rather than in electron volts.   In both figures the gross work function is measured by the difference in height between the inside (left) and outside (right) levels of the potential-

energy curve.   The gross work function in ergs is the total amount of work done against the restraining force f of the previous section during an electron's escape.   But this is simply the total area, considered positive, between the restraining-force curve, Fig. 73a, and the x-axis. For the familiar relation

$$\text{Work} = \text{Force} \times \text{Distance} \tag{340}$$

has the following mathematical form as applied to this particular situation:

$$\left.\begin{array}{l}\text{Total change in ergs of} \\ \text{potential energy during escape}\end{array}\right\} = -\int_{x=0}^{x=\infty} f\, dx \tag{341}$$

This integral is just the area between the force curve and the x-axis, if the force is in dynes.

Since the force and the ordinates of the force curve are negative, the integral and the corresponding area are also negative.   The minus sign in Equation (341) is necessary to indicate that the change in potential energy is positive, that is, an increase.

A similar equation can be used to evaluate increase in height *along* the potential-energy curve, Fig. 73b.   Hence the decrease in kinetic energy, in going from the surface to a point x_1 centimeters from it, is as follows:

$$\left.\begin{array}{l}\text{Increase in height of potential-energy} \\ \text{curve in going from } x = 0 \text{ to } x = x_1\end{array}\right\} = -\int_{x=0}^{x=x_1} f\, dx \tag{342}$$

The integral in this expression is of course the area between a selected portion of the force curve and the x-axis.

Since ordinates of the potential-energy curve represent, negatively, *areas* under the force curve, obtained by integration, ordinates of the force curve must conversely represent, also negatively, *slopes* along the potential-energy curve, obtained by differentiation.   And since it is possible to describe the potential-energy curve in terms of electron-volt units, it must also be possible to describe the force curve in terms of electron-volt-per-centimeter units, as indeed it is.   Dynes are related to electron volts per centimeter, symbol F, by the relation

$$eF10^7 = \text{dynes per electron} \tag{343}$$

which is obtained by dividing both sides of Equation (316) by distance in centimeters.

The symbol x_m (escape-distance) was used in the preceding section to describe the distance from the surface at which the external field exerts a force just equal and opposite to the image force, so that the

actual force on an escaping electron is zero. The force being zero, the slope of the potential-energy curve is also zero; therefore the escape-distance describes the position of a maximum or "crest" in the potential-energy curve or "hill." Two potential-energy curves that have such crests are illustrated in Fig. 78, page 190. Any electron that starts outward, but comes to a stop before reaching the crest, must of course fall back into the metal surface. Any electron that reaches the crest with ever so little velocity passes over it and "rolls" down the other side of the hill to the opposite electrode. The other side of the hill is the third type of region described in Section 92.

94. Potential-Energy Diagrams vs. Potential Distribution Diagrams. The relation described above between potential energy and restraining force is exactly the same as that between electric potential and electric field intensity. Whether the symbol F is used to describe electric field intensity in volts per centimeter, or to describe the restraining force in electron volts per centimeter, the force in dynes per electron is given by Equation (343).

So far the only apparent difference in nature between potential distribution diagrams, like those in Figs. 4, 7, 20, and 44, and potential-energy diagrams that employ the electron volt as the unit of energy, is the simple one of being upside down relative to one another. That is, as an electron moves to the right in a region like that of Fig. 44 its potential curve *rises*, yet its potential energy *becomes smaller*, for it moves in the direction of the electric force. Fig. 44 is of course a potential distribution diagram; the corresponding potential-energy diagram would be exactly the same figure turned upside down, that is, with all ordinates negative instead of positive. The underlying reason for this reversal is that electric potential is defined in terms of positive charges, while the particles for which these potential-energy diagrams are needed carry negative charges.

There exists a temptation to jump to the false conclusion that this graphical inversion is the only distinction of consequence between potential distribution curves and potential-energy curves in energy-level diagrams. The nature of a much more fundamental distinction can be clarified by a study of conditions in the image-force region of Fig. 73.

Imagine the electron to be momentarily at some specified distance x_1 from the metal surface. The *solid* line in Fig. 74 is a curve of the negative of the *electric field intensity* due to the image charge, which is, for the particular moment represented, located at $-x_1$. The equation for the solid line is

$$F = -\frac{e^2}{(x_1 + x)^2} \qquad \text{(344 esu)}$$

The *dotted* line in Fig. 74 is the *restraining-force curve*, which near x_1 follows the image-force law, Equation (338). The contrast is apparent. The two curves must of course have a common value at $x = x_1$, for the image force, whether measured in dynes per electron or in electron volts per centimeter, is always just that due to the electric field; *but the electric field is continually changing as a result of the electron's motion.*

The solid-line curve of electric field intensity, and so of course the corresponding potential distribution in the image-charge field, is different for each new momentary position of the electron.

Thus there can be no such thing as a single curve either of electric field intensity or electric potential that is valid for the entire duration of the electron's outward flight. Yet it is perfectly possible to draw, as in Figs.
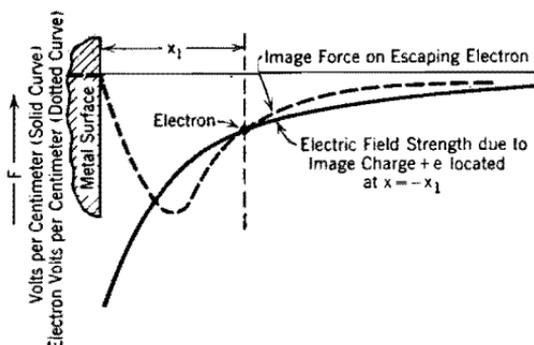
FIG. 74. Contrast between the curve of electric force and that of electric field strength in the continually changing field set up by an escaping electron.

69 and 73*b*, curves which correctly describe the variations with location of an electron's *potential energy* as a result of outward flight through the continually changing field.

A study of conditions close to the surface, in the first type region described in Section 92, leads to very different results. Here the important forces on an escaping electron are those due to near-by ions, whose locations are permanently fixed. Hence in this region the movement of the electron does not affect the locations of the charges that produce the field, and a single electric intensity curve satisfactorily describes the field and the restraining force, which are related to one another in accordance with Equation (343). Similarly, in such a region as is illustrated in Fig. 44 no distinction except that of inversion due to change of sign can be made between the electric potential distribution curve and a potential-energy curve that employs the electron volt as the unit of energy.

A potential-energy diagram for electron movement is thus a very general type of graphical representation that is identical with the negative of a potential distribution diagram *only* for those cases in which the distribution of the charges producing the field remains constant. In tubes carrying ordinary space-charge-limited current the charge distribution is constant, even though individual electrons move, so that

potential diagrams are satisfactory.   In a Barkhausen-Kurtz oscillator tube, Sections 67–68, the charge distribution changes rapidly, requiring a potential-energy point of view for complete analysis.[131]

95. Gross Work Function Inversely Proportional to Atomic Spacing. It has been pointed out that the gross work function is the area under the restraining-force curve, Fig. 73a.   An analysis of the factors determining work-function magnitudes starts with a consideration of the curve itself. Experiments have shown that at the outer extreme it follows the image-force relation.   It seems reasonable to suppose that the shape at the other extreme, just outside the surface, is dependent on the arrangement of the atoms of the metal, for the near-to-the-surface force must be due to attraction between electrons [and individual atoms (more properly called ions).
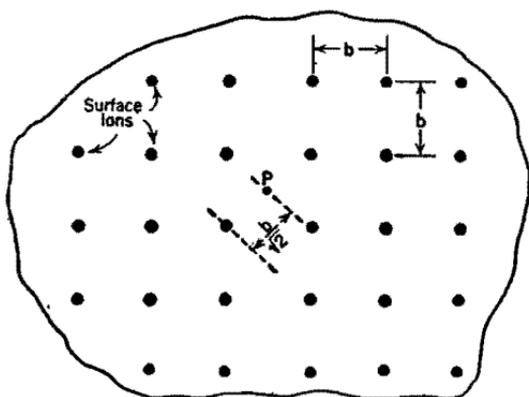
Fig. 75.   Array of surface ions of a metal.

Suppose that the metal surface exposes an ionic arrangement like that shown in Fig. 75, the surface ions being b centimeters apart.   The direction of electron escape is toward the reader.   The most favorable avenue for electron emergence is presumably a path always equally distant from the four nearest ions, that is, along a forwardly directed line ⁑emerging, as at P, from the center of one of the squares.

Calculation of the force on an electron just after it passes the last row of ions ($x = 0$) on outward flight along this path can easily be made, if it is assumed that the forces due to the four adjacent ions are the only important ones for values of x small relative to b.   The distance from each of these four ions to an electron moving along this path is

$$\sqrt{x^2+\left(\frac{b}{\sqrt{2}}\right)^2}$$

that is,

$$\sqrt{x^2 + \tfrac{1}{2}b^2}$$

The electric force *normal to the surface* exerted by the four ions is therefore, in dynes

$$f = -4\frac{e^2}{(x^2 + \tfrac{1}{2}b^2)}\frac{x}{\sqrt{x^2 + \tfrac{1}{2}b^2}} = -\frac{4e^2x}{(x^2 + \tfrac{1}{2}b^2)^{\frac{3}{2}}} \quad (345 \text{ esu})$$

The restraining-force curve of Fig. 73a should start out from the

origin in accordance with this equation. Its slope at the origin should be

$$\left(\frac{df}{dx}\right)_{x=0} = -\frac{8\sqrt{2}e^2}{b^3} \qquad \text{(346 esu)}$$

obtained by differentiating Equation (345), then letting $x = 0$.

The nature of the relation to be expected between the gross work function and the atomic lattice-constant b is important, and can be derived by using a parabolic approximation to the force curve, as suggested by Langmuir.[72] Assume that the force curve starts out from the surface as a parabola that passes through the origin and is tangent to the image-force curve at a distance x_0 from the surface, and that from the point of tangency outward the image-force curve is followed. Fig. 73a is evidently in general agreement with such an approximation. Estimate of the corresponding gross work function is made easy by the fact that the area under the parabolic portion of such a composite force curve is just equal to that under the image-force portion. The effect of the lattice-constant is introduced by making the slope of the parabola at the origin have the value specified by Equation (346); this carries the parabolic approximation a step farther than Langmuir did.

The equation of the parabola is found as follows: For the parabolic portion

$$f = Ax^2 + Bx \qquad \frac{df}{dx} = 2Ax + B \qquad \text{(347 esu)}$$

For the image-force portion

$$f = -\frac{e^2}{4x^2} \qquad \frac{df}{dx} = +\frac{e^2}{2x^3} \qquad \text{(348 esu)}$$

At $x = x_0$ the two force curves are equal and tangent, so that both the forces and their derivatives must be equal. The use of this fact leads to the following pair of equations

$$e^2 = -4Ax_0^4 - 4Bx_0^3 \qquad \text{(349 esu)}$$

$$e^2 = +4Ax_0^4 + 2Bx_0^3 \qquad \text{(350 esu)}$$

Equations (349) and (350) can be solved for A and B, giving

$$A = +\frac{3}{4}\frac{e^2}{x_0^4} \qquad B = -\frac{e^2}{x_0^3} \qquad \text{(351 esu)}$$

The equation of the parabolic portion of the curve is then

$$f = \frac{e^2}{x_0^2}\left[\frac{3}{4}\left(\frac{x}{x_0}\right)^2 - \left(\frac{x}{x_0}\right)\right] \qquad \text{(352 esu)}$$

That of the image-force portion, in similar form, is

$$f = \frac{e^2}{x_0{}^2} \frac{1}{4\left(\dfrac{x}{x_0}\right)^2} \qquad \text{(353 esu)}$$

Integration to determine the area under the parabolic portion between $x = 0$ and $x = x_0$ yields $e^2/4x_0$. Similar integration between $x = x_0$ and $x = \infty$ for the image-force portion gives the same result. The total gross-work-function area including both portions is therefore $e^2/2x_0$. To express this in electron statvolts, E_G, both sides of the equation

$$E_G e = \frac{e^2}{2x_0} \quad \text{(ergs)} \qquad \text{(354 esu)}$$

can be divided by e. Subsequent conversion to practical units gives

$$E_G = \frac{9}{2} \frac{e}{x_0} 10^{11} \quad \text{(electron volts)} \qquad \text{(355 p)}$$

This is the gross work function (see Fig. 69) that corresponds to the parabolic approximation to the near-the-surface part of the restraining-force curve.

In order to relate gross work function to the lattice-constant b the parabolic curve must be made to come out of the origin with the slope, given by Equation (346), that the true curve must be expected to have. The slope of the parabola at the origin is B, from Equation (347), so that

$$B = \frac{8\sqrt{2}e^2}{b^3} \qquad \text{(356 esu)}$$

Combination of this with Equation (350) gives

$$\frac{1}{x_0} = \frac{2}{b} 2^{1/6} \qquad (357)$$

Substitution into Equation (354) results in the following final expression for gross work function calculated in accordance with the parabolic approximation and matched with true force expectation at $x = 0$:

$$E_G = 9 \times 2^{1/6} \times 10^{11} \frac{e}{b} = \frac{16.1}{b} \times 10^{-8} \quad \text{(electron volts)} \quad \text{(358 p)}$$

Thus the gross work function may be expected to vary, among similar metals, inversely with the atomic spacing. The parabolic approximation used for demonstrating this is of course rather an arbitrary one, but it has the merits of being easy to state, having a reasonable general form, and giving work functions of the right order of magnitude. Other syste-

matic mergings between the fixed-ion force at small values and the image force at large values of x result in a similar inverse relationship between gross work function and lattice-constant.

Actual gross work functions for the alkali metals are plotted against values of $1/b$ in Fig. 76. The solid line through the points indicates the actual nature of the dependence of gross work function on $1/b$, while the dotted line through the origin indicates the dependence predicted by Equation (358). Evidently the true force curve encompasses a
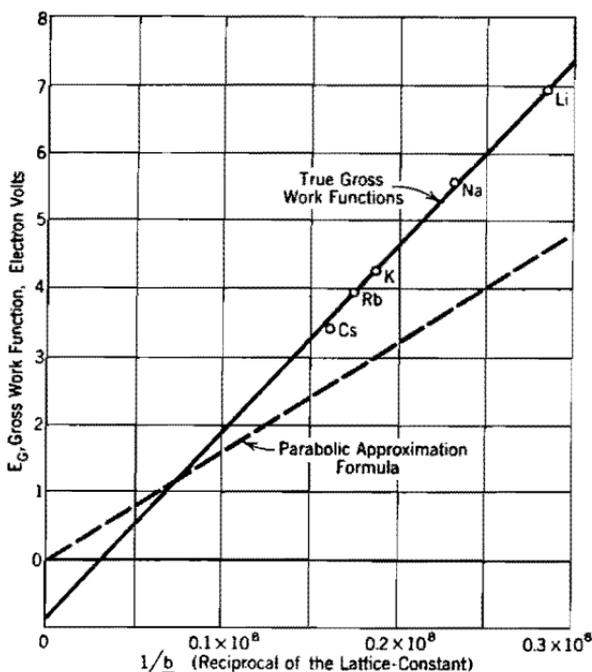
FIG. 76. Relation between gross work function and atomic spacing, for the alkali metals.

somewhat larger area than the parabolic approximation, by some 30 to 50 per cent. True gross work functions for this group of metals are seen to vary nearly in proportion to $1/b$.

The first step in obtaining both the lattice-constant and the gross work function for use as in Fig. 76 is to determine the number of ions and electrons per cubic centimeter of metal. This is done by dividing the specific gravity (See Table VII) by the mass per atom, determinable from the atomic weight. The alkali metal atoms release one electron each, so the number of electrons per cubic centimeter is the same as the number of ions. The height E_N of the normal maximum level is determined from Equation (332). The net work functions E_W of the

alkali metals are known from photoelectric measurements, as given in Table III. And of course

$$E_G = E_N + E_W \tag{359}$$

This relation determines E_G for use in Fig. 76.

The lattice-constant b depends on the number of ions per cubic centimeter and on the nature of the crystal structure. The ions of most metals whose work functions are interesting are arranged in a cubical crystal structure. In some of them the arrangement of ions is like that of the points in the quantum-number lattice, Fig. 70; such metals are said to crystallize in space-centered cubes. In others crystallization is body-centered, that is, each cube contains an ion at its center. The alkali metals are believed to form body-centered crystals.[Y] No force action due to the centered ion was considered in the fixed-ion force analysis, because its effect was presumed to be neutralized by the electron left behind in the cube (each body-centered-cube normally contains two electrons).

There are just as many cubes as ions in a *space-centered* crystal structure, because each ion may be thought of as located at the center of a cube that contains no other ions. Obviously a *body-centered* structure contains one more ion per cube than does a space-centered one, so that there are two ions and electrons per cube in the alkali metals.   If the number of ions per cubic centimeter and the number of ions per cube are known, the number of cubes per cubic centimeter, also the size of each, can be calculated. Results of such calculations are given in Table VII.

PROBLEMS

CHAPTER VIII

1. For how many free electrons within a rectangular enclosure can n^2 have the value 7? the value 8? (See Table VI.)

2. The specific gravity of aluminum is 2.70, and its atomic weight is given in Table V. Find the value of E_N (kinetic energy at the normal maximum level) for aluminum. (The shell distribution of electrons, Table V, gives an indication as to how many free electrons per atom the various metals may be expected to have.)

3. A rectangular enclosure whose volume is 10^{-21} cc contains an electron gas that consists of 30 electrons. Find the normal maximum energy (*a*) by the use of Table VI and Equation (329), also (*b*) by the use of Equation (332). Which answer is correct, and why? (*c*) Find the least possible average energy, in electron volts, using Table VI and Equation (329).

4. (*a*) Find the normal maximum energy, in electron volts, for argon at 0° C and 760 mm pressure, and (*b*) that for helium at a pressure of ten atmospheres and a temperature of 10° K (−263° C).

5. Derive an expression for the work function, in volts, of a homogeneous metal surface from the following assumptions: the restraining-force curve is presumed to start as a straight line passing through zero force at the surface and intersecting the

image-force curve at b centimeters from the surface.   Beyond this intersection, the image-force law is presumed to hold.   Determine the location on Fig. 76 of a straight line that represents your formula.

6. Suppose that a certain metal has a spacing between atoms of 2 angstrom units (1 angstrom unit $= 10^{-8}$ cm).

(a) If the curve of restraining force on an electron escaping from this metal merges with the image-force curve at 30 per cent of one atomic spacing from the surface, and the area under the curve to the left of this point is the same as that to the right, what is the gross work function?

(b) Find the net work function of this metal, assuming that it has a space-centered cubical crystal structure, and that there are two free electrons per atom of the metal.

CHAPTER IX

ENERGY-LEVEL DIAGRAMS OF METALS

96. Purposes for Which Energy-Level Diagrams are Useful. Illustrations similar to Figs. 69 and 73b (pages 165 and 178), called *energy-level diagrams of metals*, are useful in the study of numerous kinds of electronic behavior, as suggested by the following incomplete list:

1. Valve action outside a cathode surface when current is space-charge-limited.
2. Reduction of work function of homogeneous surfaces in the presence of strong electric fields.
3. Contact difference of potential between metals.
4. Properties of oxide-coated, thoriated, and other special thermionic surfaces.
5. Photoelectric emission of electrons.
6. Properties of semi-conductors used in " sandwich " photocells, photoconductive cells, and rectifiers.
7. Thermoelectric and related phenomena.
8. Properties of the cathode spot of an electric arc.

The first four of these subjects will be discussed in this chapter and the fifth and sixth in a later chapter.

97. Valve Action Outside a Thermionic Cathode Surface When Current is Space-Charge-Limited. Dushman's equation for thermionic current density, Equation (299), page 149, describes the maximum electron current that can ordinarily flow from one square centimeter of a heated thermionic cathode at a given temperature. The corresponding maximum total current at any definite temperature is the cathode surface area multiplied by this maximum current density, and is often called the *temperature-limited current*.

When electrode geometry and potentials are so proportioned as to limit the actual current density to a *space-charge-limited* value that is less than the temperature-limited value, some mechanism must exist for turning back into the cathode the electrons that represent the excess of temperature-limited over space-charge-limited current. This mechanism is illustrated in Fig. 77, which represents an evacuated region between two parallel plane electrodes, the cathode being an emitter subject to Dushman's equation. The four curves represent different potential distributions that may exist between the electrodes, all corresponding to the same cathode temperature, but to different plate potentials.

Curves (1) and (2) represent potential distributions when the plate

potential is high enough for the possible space-charge-limited current to be greater than the temperature-limited current of the cathode at the existing temperature; the actual current flowing has, therefore, the temperature-limited value. All electrons emitted enter an electric field that causes acceleration toward the anode, for the slope of the potential line at $x = 0$ is definitely positive. All of them of course reach the plate. The magnitude of the current corresponding to Curves (1) and (2) is therefore determined by the nature and temperature of the cathode surface, not by electrode spacing or potential, hence the name "temperature-limited current."
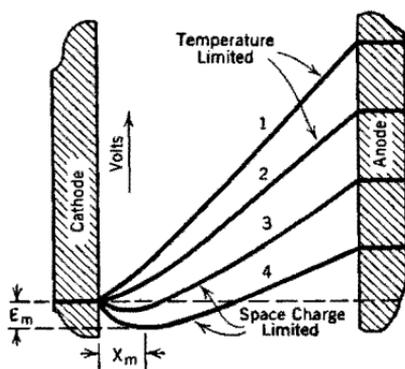
FIG. 77. Potential distribution curves illustrating valve action at emitter surface when current is space-charge-limited.

Curves (3) and (4) represent a much more usual type of potential distribution in electronic devices, corresponding to plate potential (or grid potential in a triode) low enough so that the space-charge-limited current allowed to flow is less than the available thermionic current. Under these conditions electrons are emitted from the cathode more rapidly than geometry and potentials permit their passage to the anode, so that negative space charge accumulates just outside the cathode. This produces a "negative dip" in the potential line just as it leaves the cathode, with consequent negative electric gradient at the cathode surface. All the electrons emitted thermionically under these typical conditions enter a field that tends to send them back into the cathode. Only those escape whose x-directed kinetic energies *after overcoming the work-function barrier* are sufficient to permit them to overcome the additional obstacle presented by the negative dip of E_m volts.[74, 62]

Equilibrium exists when the negative dip is just low enough to send back to the cathode all the electrons that represent the excess of the temperature-limited over the space-charge-limited current permitted by geometry and potentials. If the negative dip is momentarily too low, more than the proper fraction of emitted electrons are sent back to the cathode, so that the space charge causing the negative dip becomes less, and the bottom of the dip rises. If it rises too high, more electrons emerge from the cathode than can pass to the anode, more space charge accumulates, and the bottom of the dip drops back down again. Thus a local mechanism exists which automatically maintains the equilibrium condition.

98. Magnitude of the Negative Potential Dip Outside the Cathode.
The potential distribution curves of Fig. 77 of course take no account
of work-function details at the cathode surface.   Fig. 78 illustrates the
two potential-energy curves, similar to those used in the previous chap-
ter, which correspond to Curves (1) and (3) of Fig. 77.

The *right-hand* part of Fig. 78 is, in accordance with the principles
outlined in Section 94, simply a large-scale potential distribution
diagram inverted.   Although electrons move very rapidly through the
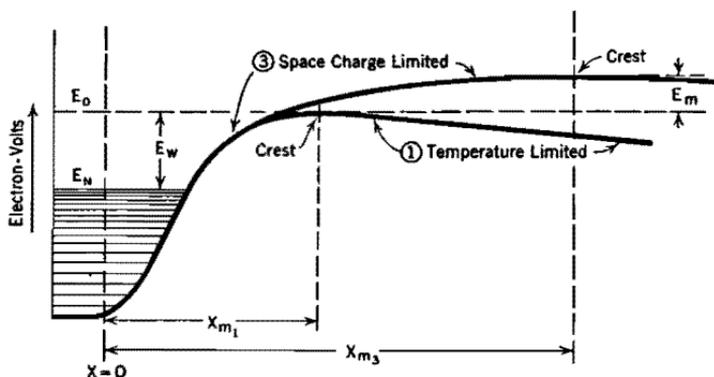
Fɪɢ. 78.   Valve action at emitter surface; potential-energy details as related to work
function.   Compare with Fig. 77.

interelectrode region, the average space-charge density in each locality
remains constant; consequently the flow of space charge does not
result in changes in the distribution of potential.   Electric potential and
electronic potential energy are the negatives of one another.

The *left-hand* part of Fig. 78 resembles Figs. 69 and 73b, since it
represents variations in the potential energy of each individual electron
on the way out from the metal.   The distance scale at the left of Fig. 78
is greatly exaggerated relative to that at the right.   The *zero level* E_0 in
Fig. 78 is as usual the top of the gross-work-function barrier, and
corresponds to the horizontal cathode-potential line in Fig. 77.

The crest or escape-point on temperature-limited Curve (1), Fig. 78,
is located where the image force and the electric force due to the external
field gradient are just equal and opposite, as discussed in Sections 92 and
93.   The escape-distance is under such conditions very small indeed,
as can be determined by a numerical solution of Equation (339) for x_m,
using for F any value of electric gradient that might reasonably be ex-
pected to exist in commercial devices.   Just as Curve (1), Fig. 77,
reaches but does not drop below cathode potential, so Curve (1), Fig. 78,
may for all ordinary external gradients be assumed to reach but not rise
above the zero level.   Only when the external gradient is extremely

large does the crest of such a curve drop appreciably below the zero level, as described later in Section 100.

Just as the accumulation of space charge outside the cathode drops the bottom point on potential Curve (3), Fig. 77, to E_m volts below cathode potential, so it raises the crest of the corresponding potential-energy Curve (3), Fig. 78, to E_m electron volts above the zero level. Thus the crest of the potential-energy curve with current space-charge-limited is E_m electron volts higher than it is with current temperature-limited.

The effect of this rise in the crest on the rate of escape of electrons is exactly the same that would result from an increase of work function by the amount E_m. Hence Dushman's equation can be used to express J_p, the actual space-charge-limited current density, by using $E_W + E_m$ instead of E_W in the exponent, as follows:

$$J_p = A_0 T^2 \epsilon^{-\frac{E_W + E_m}{E T}}$$

$$= A_0 T^2 \epsilon^{-\frac{E_W}{E T}} \epsilon^{-\frac{E_m}{E T}} \tag{360 p}$$

or, calling the temperature-limited current density J_{th}, as in Equation (299),

$$J_p = J_{th} \epsilon^{-\frac{E_m}{E T}} \tag{361}$$

A negative dip of less than a volt is sufficient to reduce the plate current to a small fraction of the thermionic emission, because the exponential factor shrinks very rapidly as E_m grows. For example, with a filament temperature of 2500° K, customary for tungsten filaments, a negative dip of $\frac{1}{2}$ volt makes the plate current about 10 per cent, and of 1 volt about 1 per cent, of the emission. At 1000° K, which is in the temperature range used for coated filaments, the corresponding negative-dip voltages are 0.2 and 0.4. The distance x_m to the bottom of the negative dip is similarly small in most cases; it is usually only a minute fraction of a millimeter. Magnitudes of E_m and x_m are fully discussed later in Section 124. In most cases they are negligibly small relative to E_p (or, in a triode, to the equivalent voltage) and to tube spacings respectively.

99. Conditions in a Triode at and Near Cut-Off. It was shown in Section 60 how to adapt the results of Chapter V to permit description of the characteristic curves of a triode, as illustrated in Figs. 3 and 50, by an equation containing an *equivalent voltage* $E_g + \dfrac{E_p}{\mu}$. Each plate characteristic curve is obtained by using in the general current-voltage relation some definite value of grid voltage. The lower end or " cut-off "

of each curve should then be the point at which the current is just zero because the equivalent voltage is zero.   Actually only an approximate cut-off value can be experimentally determined, for reasons that appear below.

Equation (282), used to describe the characteristic curves, is based on the assumption of a space-charge-limited current for which the voltage E_m and distance x_m to the bottom of the negative dip can be neglected relative to the equivalent voltage and interelectrode spacings.   Reduction of the current toward zero is, of course, accomplished by lowering the grid or plate voltage, or both, so that the equivalent voltage approaches zero.   But the mechanism by which the current is reduced requires that *the negative-dip voltage E_m must increase.*   Thus, as the current becomes very small, E_m increases, and the equivalent voltage decreases.   The former cannot continue to be neglected; in fact, E_m ultimately becomes *larger* than the equivalent voltage, and x_m becomes a considerable fraction of the cathode-to-grid spacing.   Under these conditions the space-charge-limited current equations of Section 60, expressing current in terms of equivalent voltage and tube dimensions, no longer apply; analysis in accord with the method of Section 124 is necessary.

Thus the lower ends of the plate and mutual characteristic curves do not follow the same relations that govern them in the normal operating portions, and the current does not vanish exactly at the point corresponding to zero equivalent voltage, only somewhere near that point.

Reduction of the equivalent voltage to and below zero produces a growth of the negative-dip voltage E_m sufficient to make the current shrink very rapidly to unmeasurably small values.   If the equivalent voltage is reduced to the point where a current meter, useful in the normal working range of the characteristic curve, reads zero, the use of a more sensitive meter will show some current flow.   If that is then reduced to zero a still more sensitive meter will indicate current flow. Mathematically defined cut-off does not exist; but there does exist a narrow voltage range at the lower end of each characteristic curve within which the current becomes minute relative to normal operating currents.

100. Reduction of Work Function by Strong External Fields.[73, 75, T 61] A cathode supplying temperature-limited current is exposed to a positive external field, as in Curves (1) and (2), Fig. 77.   In any ordinary electronic device such an external field is so small that its effect on the height of the crest of the potential-energy curve is negligible, as for Curve (1), Fig. 78, if the cathode surface is a homogeneous one like that of a tungsten filament.   However, very large but experimentally

obtainable fields lower the crest enough to produce measurable effects on the current.

Fig. 79 is a detail of the potential-energy curve in the presence of a very strong external field. The curve is, at every point, below its usual (dotted) position by just the amount of encouragement contributed to escape by the external field. This is the product of field force by distance traveled, that is, eFx ergs or Fx electron statvolts. Hence the potential energy is represented at remote points by a straight line with slope $-F$; this line is, of course, an inverted image of the potential distribution line, as it should be.

The potential-energy curve has a crest at the escape-distance x_m, where, for this temperature-limited current condition,

Fig. 79. Reduction in work function due to presence of strong external field; amount of change is $\sqrt{eF}$ (esu); homogeneous metal surface.

the image force and field force are equal and opposite, as in Equation (339). Solution of this equation for x_m gives

$$x_m = \frac{1}{2}\sqrt{\frac{e}{F}} \qquad \text{(362 esu)}$$

The magnitude of the reduction in work function due to the external field is the amount by which the crest lies below the zero level, which is the escape level in the absence of the field.

It is apparent from a little study of Fig. 79 that there are two contributions to the distance from the zero level downward to the crest, as follows:

(1) The normal potential-energy curve is, at the distance x_m, below the zero level by a contribution obtained by integration of the image force from $x = x_m$ to infinity, as follows:

$$\int_{z=z_m}^{z=\infty} \frac{e^2}{4x^2} = \frac{e^2}{4x_m} \quad \text{(ergs)} \qquad \text{(363 esu)}$$

By substituting for x_m from Equation (362), and dividing by e to convert to electron statvolts, this contribution to the reduction in work function is found to be $\frac{1}{2}\sqrt{eF}$ electron statvolts.

(2) The crest is below the normal potential-energy curve at x_m by an

additional amount Fx_m. Again substituting for x_m from Equation (362), this second contribution is found to be exactly the same as the first.

Thus the total change ΔE_W in work function due to the external field is, in electron statvolts,

$$\Delta E_W = \tfrac{1}{2}\sqrt{eF} + \tfrac{1}{2}\sqrt{eF} = \sqrt{eF} \qquad (364 \text{ esu})$$

or, in practical units

$$\Delta E_W = 300\sqrt{eF10^7} \qquad (365 \text{ p})$$

This can be checked experimentally by varying F up to very large values and measuring the resultant thermionic current; in general, using Dushman's equation

$$J_{th} = A_0 T^2 \epsilon^{-\frac{E_W - 300\sqrt{eF10^7}}{ET}} \qquad (366 \text{ p})$$

so that

$$\log J_{th} = \log\left(A_0 T^2 \epsilon^{-\frac{E_W}{ET}}\right) + \frac{300\sqrt{e10^7}}{ET}\sqrt{F} \qquad (367 \text{ p})$$

A plot of $\log J_{th}$ against $\sqrt{F}$ should then give a straight line of slope $300\sqrt{e10^7}/E_T$ in practical units, which is $\sqrt{e}/E_T$ in electrostatic units. This has been verified experimentally;[73, 75, 76, 77] Fig. 80 illustrates the type of line obtained in the laboratory. The fact that these curves do actually have the predicted slopes is the experimental verification of the image-force law for points not too close to the surface. No experiments so far performed have made x_m small enough to permit discovery of the nature of the close-in departure from the image-force law.

The reduction of work function by external fields that has been described here is often called the "Schottky effect," because the theory just outlined was first stated by Schottky. Straight lines (Fig. 80) that describe the relation between the logarithm of the emission current and the square root of the external field are sometimes called "Schottky lines." Electron emission that occurs chiefly as a result of the presence of a strong external field's effect on work function is often called "field emission."
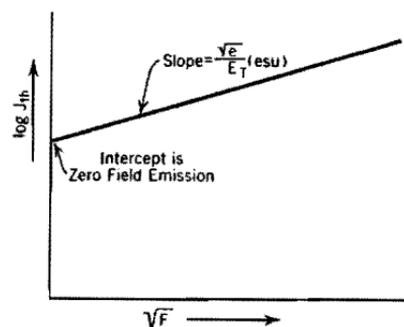
FIG. 80. Variation in thermionic current due to external field, homogeneous metal surface (Schottky line).

101. Contact Difference of Potential. The significance of the term "contact difference of potential" may be understood by reference to Fig. 81. If pieces of two different metals are moderately close to one

another but not touching, both being electrically uncharged, there is, of course, no electric field between them. The potential-energy curve between them is, as illustrated in Fig. 81a, horizontal except very near each surface.

If these two pieces of metal are now brought *very* close together, still however not touching, the two image-force regions overlap. Calculation of the force in the narrow gap between the surfaces would require imaging in *both* conductors, and there would in general be some force
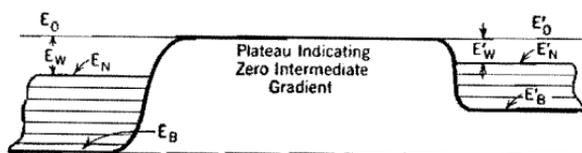
(a) Remote unlike surfaces both uncharged.
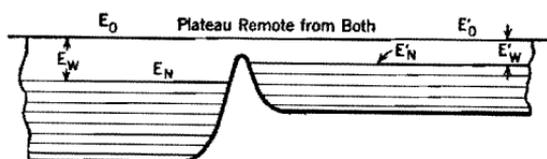
(b) Same surfaces approaching closely, but not touching. No charge interchange; no shift of bottom levels.
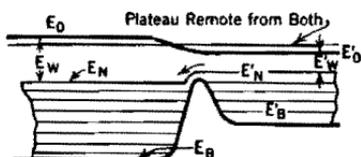
(c) Electrical contact just made, transferring electrons to the left; E_B, E_N, E_0 all raised; E'_B, E'_N, E'_0 all lowered.
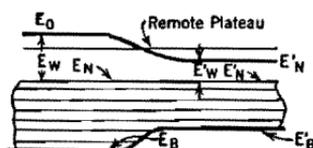
(d) Complete electrical contact; E_N and E'_N levels coincide. Contact difference of potential is $E_0 - E'_0 = E_W - E'_W$.

FIG. 81. Contact difference of potential.

each way. The crest of the potential-energy curve through the gap is then below the zero level, as illustrated in Fig. 81b. As yet there has been no charge transfer, so that both pieces of metal are uncharged; a potential-energy line drawn between two *nonadjacent* (perhaps backside) boundaries of these two metals must still be flat-topped, and the vertical positions of the top and bottom energy levels are unchanged.

As soon as the crest of the potential-energy barrier in the narrowing gap drops low enough so that electrons can spill over from the low-work-function metal to the high-work-function one, electrons move

away from the right-hand piece, giving it a positive charge and the left-hand piece a negative charge. Thus, the right-hand one acquires a more positive potential, the left-hand one a more negative potential than before; the two zero levels shift relative to one another, the right-hand one moving downward, the left-hand one upward, as illustrated in Fig. 81c. This shift continues until the two normal maximum levels line up horizontally, as in Fig. 81d. Then there is no further tendency for electrons to flow either way, because no electron in either metal can by crossing the boundary find a vacant level lower than its present one. This is an equilibrium condition.

The two zero levels are now displaced vertically by just the difference between the two net work functions. This displacement is measurable, and is called the *contact difference of potential;* thus *the contact difference of potential between two metals is simply the difference between their net work functions.*[73, 78, 79, 80, 81] Low-work-function metals are called *electropositive* because they become positively charged and acquire positive potentials when in contact with high-work-function metals.

Two pieces of metal are in contact only along one surface of each, and there are other sides to both. A potential or potential-energy curve between non-contacting sides must, of course, have a slope, to accommodate the change in height in passing from one zero level to the other, and this slope corresponds to an external electric field. Measurements of contact differences of potential are basically measurements of this electric field, and are always made by electrostatic voltage-measuring devices.

Ordinary electromagnetic meters respond only to continuous current flow. The vertical shift of energy levels that sets up the external field between metals in contact establishes an equilibrium in which there is no tendency toward transfer of electrons from one metal to another; therefore electromagnetic meters cannot be used to discover or measure contact differences of potential. Also, all the contact differences of potential around a closed completely metallic circuit must add up to zero, as a little consideration of their origin will make evident; this again indicates that no current flow, even for metering purposes, can result from them in such a circuit.

102. Effect of Contact Difference of Potential on Triode Plate Current. Previous chapters have shown that the plate current in triodes, tetrodes, etc., is dependent on the space-charge-free electric field just outside the cathode surface. The various electrodes are, of course, connected to one another through external circuits, and may be of unlike metals, so that the electrodes expose within the tube just such noncontacting surfaces as are referred to in the previous paragraph. Hence the space-

charge-free off-cathode gradient must include a contribution due to the contact difference of potential between the metals of the electrodes.

Contact differences of potential can be taken into account, in calculations of plate current magnitude, by adding, with proper algebraic sign, to plate and grid voltages respectively, the corresponding values of contact differences of potential relative to the cathode. If the work function of the cathode is considerably less than that of either grid or plate surface, contact difference of potential will have the same effect as lowering both grid and plate potentials, that is, making a negative grid more negative, and a positive plate less positive.

Since the cathode-to-grid and cathode-to-plate contact differences of potential, symbols E_g' and E_p', are constant for any given tube, it is possible to combine them into a constant term

$$E' = E_g' + \frac{E_p'}{\mu} \qquad (368 \text{ p})$$

which can be included in the equivalent voltage, giving to Equation (282) for triode plate current the revised form ^{Q}

$$I_p = B\left(E' + E_g + \frac{E_p}{\mu}\right)^n \qquad (369 \text{ p})$$

It is apparent from Equation (368) that the grid-cathode contact difference of potential is more important than that between plate and cathode, because μ is always greater than unity. The order of magnitude of E' is evident from its relation to work functions of the respective surfaces, discussed in the preceding section.

103. Thoriated Tungsten Cathodes. It is, in general, commercially desirable to obtain the thermionic emission required for a tube with as little cathode heating power as possible. This points to the need for low-work-function cathode surfaces, for small power loss is favored by low cathode temperature, and only materials with low work functions can release appreciable thermionic currents at low temperatures (see Chapter VII). But unfortunately materials such as caesium, rubidium, barium, strontium, etc., that have low work functions, melt, and in some cases boil, at temperatures even lower than those required for appreciable thermionic emission from them.

Special methods of preparation make it possible, however, to provide stable thin layers of low-work-function substances on filaments of tungsten, nickel, platinum, or various alloys. These *base metals* maintain structural stability at temperatures high enough to produce substantial emission from outer layers of low-work-function materials. There is a crude analogy here with the fact that, although water has no struc-

tural stability above 0° C, it is not difficult to maintain wet surfaces at much higher temperatures.

Cathodes consisting of tungsten filaments with surfaces covered by a single layer of thorium atoms have been rather widely used commercially.   The major steps in the process of preparation of such a *thoriated filament* are as follows: [66, T 104, S II, E 115]

(*a*) The filament is drawn from tungsten which contains a small percentage of thorium oxide, called *thoria.*

(*b*) It is "flashed" in a high vacuum; this means simply that it is heated for a short time to a temperature of about 2800° K.   Flashing reduces some of the thoria within the tungsten to metallic thorium.

(*c*) The filament is then "activated" by being held for some time, again in a high vacuum, at a temperature in the neighborhood of 2100° K. During this process the monatomic surface layer of thorium is formed, as a result of the following circumstances:

(1) The thorium atoms formed during flashing diffuse from the interior of the tungsten to its surface.   Tungsten, like all metals, has a granular structure; the thorium atoms are believed to diffuse along boundary surfaces between grains, so that fine grain structure aids diffusion.   This movement toward the surface takes place more rapidly at high than at low temperatures.

(2) At the activation temperature the first thorium atoms to arrive at the surface evaporate from it very much *less* rapidly than new ones arrive from the interior, because of the strong attachment energy that exists between a thorium atom and the tungsten surface; hence thorium atoms accumulate on the surface.

(3) As soon as any small portion of the surface becomes covered with a monatomic layer of thorium atoms, any additional ones that arrive from the interior must form a second layer.   But the attachment energy between thorium atoms of the second layer and those of the first layer is very small.   At the activation temperature evaporation of the second layer occurs much *more* rapidly than arrival of additional thorium atoms from the interior.   Only the atoms of the first layer remain on the surface for any considerable length of time.

(*d*) The filament with its monatomic thorium layer is now ready for use at a temperature several hundred degrees lower than that used for activation, at which neither diffusion of thorium atoms to the surface nor evaporation of them from it are appreciable.   Such filaments may have work functions lower than that of pure thorium, and are satisfactory as thermionic emitters for many purposes.

The monatomic layer of thorium on such a filament can be quickly destroyed by contact with gases, or by heating to above the activation temperature.   There is some evaporation of thorium atoms even at the operating temperature, so that a thoriated surface has a limited life even if properly used.   It can, however, be rejuvenated by "reactivation," providing previous activations or misuse have not driven all the thorium atoms to the surface.   In that case more thorium atoms may be

produced internally by flashing again, for at each flashing operation only a small fraction of the available thoria is reduced to thorium. Further details regarding properties of thoriated surfaces will be found in the reference books listed in the next section.

104. Oxide-Coated Cathodes. Processes wholly different from that just described can be used to produce a surfacing of barium, strontium, or a mixture of the two, over a base metal which is often nickel or a nickel alloy. Such surfaces include a layer of oxygen or of oxides of barium of strontium between the base metal and the active emitting surface. This intermediate layer may be from one, two, or a few, to a few score thousand atomic layers in thickness. The reader is referred to the books by Reimann,[T IV] Chaffee,[E 102] and Koller[S III] and to a paper by Dushman,[66] for detailed discussions of the processes of preparation, properties, and theories of such *oxide-coated* surfaces. Only a very brief review of some of the essential facts and principles will be attempted here.

Each tube manufacturer has preferred formulas for the preparation of oxide-coated cathodes for specific purposes, selected as a result of experimental investigations and from considerations of manufacturing expediency. A commercially used surfacing is likely to have a relatively thick layer (that is, one that is many hundred or thousand, rather than a few, molecular layers thick) of oxides of barium and strontium in chosen proportions, firmly " sintered " (a heat-treatment process) to the base metal. It is presumed, on the basis of excellent but indirect evidence, that an outer layer of barium, probably monatomic, covers the oxide.

The various emission properties of an oxide-coated surface depend on the base metal used, on the proportions of the contributing oxides, and on many details of the complicated processing used. Only by painstaking control of manufacturing operations can any given set of thermionic properties be consistently reproduced. In all cases the emission is markedly affected by relatively weak external fields; in this respect the behavior of oxide-coated surfaces exhibits a sharp contrast with that of homogeneous surfaces; see Section 107.

Oxide-coated surfaces are complex in structure, subject to the effects of many variables in composition, processing, and history during use. No thoroughly satisfactory theory of their behavior has yet been advanced, although considerable progress toward a better understanding of them has been made in recent years. A few fairly definite general statements can be culled out of the mass of conflicting evidence, hypothetical statements, and theories that are to be found regarding them in scientific literature, as follows:

(*a*) Oxide-coated surfaces are, on the whole, more efficient thermionically and at the same time more generally satisfactory than any other type of commercial emitting surface, in spite of the fact that they have relatively high radiation emissivity coefficients.

(*b*) The electron emission properties are primarily dependent on the existence at the extreme outer surface of a very thin layer, quite possibly monatomic, of barium, strontium, or both; probably barium predominates at the surface even when both oxides are used in nearly equal proportions.

(*c*) The composition of the base metal affects the emission properties of the surface, but probably in some indirect way.

(*d*) Any satisfactory theory of emission from oxide-coated cathodes must account for the conduction of the emission current through the intermediate layer of oxide, which is a relatively poor conductor.

(*e*) It seems likely, from considerations growing partly out of analyses of substances (chiefly oxygen, barium, strontium) which can be driven from the surface during activation or during use at excessive temperatures, that a migration of barium and strontium particles through the intermediate layer takes place and accounts for part of the conduction current, *though at high temperatures probably only a very small part.* In general, neutral barium particles are believed to migrate toward the outer surface of the coating, there acquire positive charges by the loss of two outer-shell electrons each, and then travel back as barium ions to the interface between base metal and oxide. Here they are neutralized by stealing two electrons each from the base metal surface, and are ready to migrate forward again. This process is sometimes called *barium circulation.* Strontium particles may circulate similarly.

(*f*) Probably these migrating ions carry a relatively large part of the total thermionic current at the lower operating temperatures but only an insignificant part of it at the higher, more usual, operating temperatures, *the major part of the current then being carried by electron movement through the oxide.* Barium circulation may still be necessary for the preservation of conditions favorable to outward passage of electrons through the interfaces and intermediate layer of oxide.

(*g*) The most generally satisfactory analyses of the properties of oxide-coated thermionic surfaces use an energy-level-diagram approach. The intermediate layer of oxide is treated as a *semiconductor* (see Section 223), whose conducting properties are the result of impurities within it. The impurities are probably barium and strontium atoms; their effects can be analyzed in the general manner used in the study of other semiconductors in a later chapter dealing with photoelectric surfaces. Potential-energy changes in passing through the interfaces between

the various layers can be analyzed in terms of layers of polarized atoms or molecules.

The reader is referred particularly to Reimann's book,$^{T IV, T V}$ for an application of the energy-level point of view to the analysis of the properties of oxide coatings. The next two sections contain discussions that will be helpful preliminaries to Reimann's treatment.

105. Potential-Energy Diagrams for Polarized Atomic Layers. Presentations of energy-level studies of emission problems, such as that given in Reimann's book,[66, T] usually assume an understanding on the part of the reader of the effect on an electron's potential energy of passage through a *polarized layer of atoms*.

Each atom of a polarized layer, though electrically neutral, exposes a positive charge in one direction and a negative charge in the other direction. For example, it is not unreasonable to suppose that the layer of barium on the outer surface of an oxide-coated cathode is monatomic and polarized, as suggested by Fig. 82a. Each barium atom contains two outer-shell electrons, and it is at least conceivable that there may be sufficient attraction between the oxide of the intermediate layer and these outer-shell electrons to keep all of them on the inward sides of the barium atoms of which they are a part. In that case each atom of the outer barium layer exposes two negative electronic charges inwardly. Since each atom is as a whole electrically neutral, the layer must have a positive outward aspect. This situation may or may not actually exist in an oxide-coated surface, but it satisfactorily illustrates a frequently used manner of approach to the analysis of surface-layer problems.
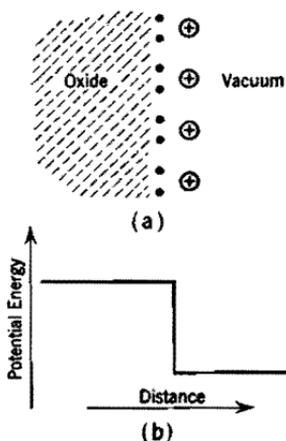
FIG. 82. Polarized layer of barium atoms at outer face of oxide coating.

The gross aspect of a polarized layer of atoms is that of an *electrical double layer*. This means simply that to an observer unable to distinguish between individual particles it would appear to be a sheet of uniform positive surface-charge density just outside a sheet of uniform and equal negative surface-charge density. Such a double layer is, of course, similar in effect to a sheet of uniformly distributed, closely packed electric dipoles, and is also the electrical equivalent of the magnetic double layer or magnetic cap concept sometimes used in analyzing the relation between electric currents and magnetic fields.^{A} [163]

The essential property of an electrical double layer is that passage

through it entails an abrupt change in potential,[4][163] as illustrated in Fig. 82*b*. This is entirely reasonable, for the electric field set up by it is exactly like that due to parallel metal plates infinite in extent, as in Fig. 5. The distance between the charge sheets of an electrical double layer is very small, and the electric field there large. The *electric moment* of a double layer is the product of the surface-charge density of each sheet by the distance between them, and is therefore directly proportional to the potential change experienced in traversing it.
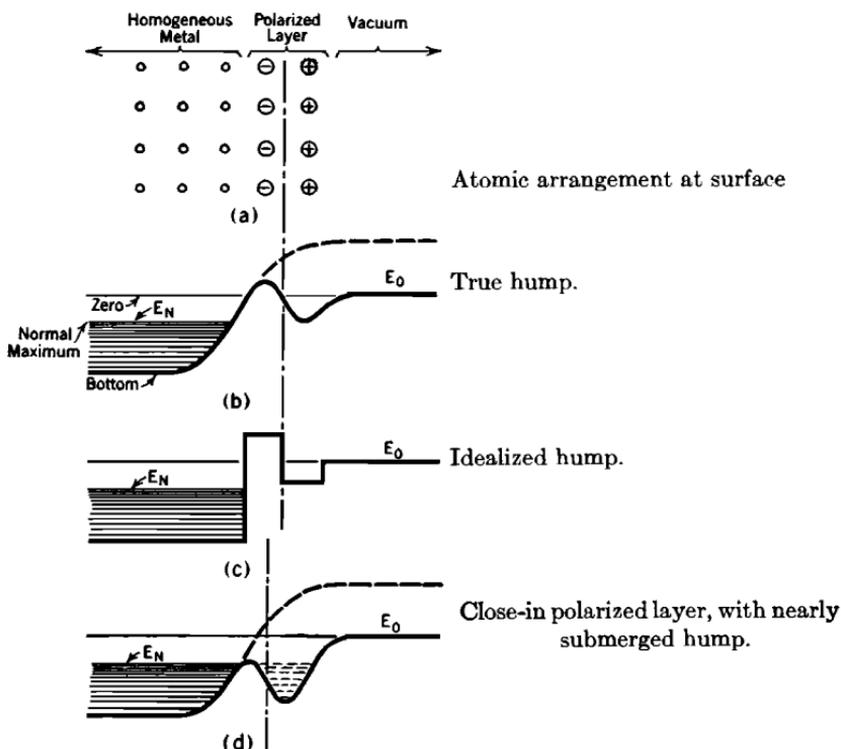
Fig. 83.   "Hump" of potential energy caused by a polarized layer.

A polarized atomic layer with a positive outward exposure (*a*) invariably reduces the work function, and (*b*) sometimes produces a " hump " in the potential-energy curve. These effects can be illustrated most clearly by imagining a polarized layer of atoms, with a single outer-shell electron each, laid over a homogeneous metal surface, as illustrated by Fig. 83*a*. This polarized layer might be a layer of caesium atoms on tungsten. Without the polarized layer the potential-energy curve outward through the surface would have the general shape, familiar from Figs. 69, 73, 78, and 79, that is indicated by the dotted line in Fig. 83*b*. The presence of the polarized layer produces an intermediate decline

of several volts in the potential-energy curve, as illustrated by the solid line in the same figure.

The intermediate decline results in a substantial drop in the height of the ultimate outside or zero level. If the decline begins at some little distance out from the last homogeneous layer of metal ions it may, as shown in the figure, produce an intermediate "hump" whose crest is above that of the outside zero level. Such a hump considerably complicates the problem of predicting probabilities of electron escape. It is sometimes convenient in attempting to make such predictions to idealize the shape of the potential-energy curve into a "block" form, illustrated for this relatively simple composite surface by Fig. 83c.

It must not be assumed offhand that *every* polarized layer results in a hump whose crest is above the outside zero level. If the decline due to the polarized layer begins very close to the last homogeneous layer of ions of the metal, the crest of the hump may be only very little above the bottom level, therefore well below the outside zero level. This is illustrated in Fig. 83d. Here the hump is nearly "submerged," and does not rise above the ultimate outside level.

Fig. 84 is a diagram, adapted from Reimann's book, which illustrates a conception of the potential-energy variation through oxide coatings which has led to reason-
ably satisfactory explana-
tions of many of the observed properties of such surfaces. It assumes two humps, therefore two po-
larized layers, one at the interface between base metal and oxide, the other at the outside of the oxide.
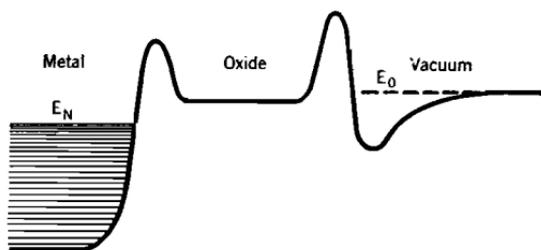
FIG. 84. Potential-energy variation through an oxide-coated surface.

Presumably both are due to polarized layers of barium atoms with positive poles pointed outward. It also assumes a sharp rise in potential energy in passing from base metal into oxide, and a less sharp rise, almost if not quite nullified by the subsequent decline due to the polarized layer of barium, at the outside surface of the oxide.

A study of the details of particle placement in polarized layers makes it apparent that in order to experience the "decline" an electron must emerge *between* polarized atoms; in other words, the potential-energy diagrams shown indicate energy variations through the easiest paths. It seems reasonable to suppose that the atoms of a polarized layer on a metal surface line up with the atoms of the metal itself, so that emergence between metal atoms favors passage between atoms of the outer layer.

106. Transmission of Electron Waves through Potential-Energy Humps. There is serious temptation to jump to the conclusion that the difference between the normal maximum level and the height of the hump's crest is the proper value of net work function to use in calculating thermionic emission from a surface whose potential-energy diagram resembles Fig. 83*b*. The most generally satisfactory theories of electron emission through oxide-coated surfaces have assumed, however, that the *wave-like* properties of electrons permit them to pass *through* quite high potential-energy humps resembling that in Fig. 83*b provided the humps are sufficiently thin*. The thinner a hump is, the greater is its " transmission coefficient," which states the fraction of approaching electrons that can penetrate through the hill in spite of their kinetic-energy levels being below the crest. In general, the transmission coefficient is very small unless the hump is very thin.$^{T\,252}$

Electron waves and light waves are composed of different wave-stuff, the wave equations for electrons being very different from Maxwell's equations that describe the properties of electromagnetic radiations.$^{U.\text{XXVIII},\ U\text{XXIX}}$ Yet a few analogies can be drawn between the two kinds of waves; and one in particular, regarding wave reflection, is helpful here.

A hump tends to reflect back into the metal the electron waves that approach it, as a mirror reflects light waves. A *sufficiently thin* mirror reflects light only partially, so also a sufficiently thin hump reflects electron waves only partially. The wave length of an incident wave is the yardstick by which is determined how thick is " sufficiently thin." Thus a layer of light-reflecting material, or a hump, that is very thin relative to the incident wave length permits a considerable amount of light or electron-wave intensity, respectively, to appear beyond it.

The hump is exposed to as many different wave lengths of electron waves as there are values of the quantum number n_u (see Section 86) represented among the interior electronic motions, because the x-directed wave lengths are inversely proportional to the n_u's.

107. " Saturation "; Failure of Composite Surfaces to Saturate; Grid Control of Temperature-Limited Currents. It will be recalled that when a *temperature-limited current* is flowing through a tube the potential gradient just outside the cathode is definitely positive, that is, every escaping electron emerges into an electric field that drives it toward the anode. Section 100 and Fig. 79 describe and illustrate the variation of electron emission with the strength of this external field *for homogeneous metal surfaces*. The essential point of the discussion there presented is that this variation of temperature-limited currents with external field, though real, is for such surfaces so very small as to be

insignificant within the normal working range of most commercial devices. That this is true is illustrated by Fig. 85a, which shows the *saturation* of the plate characteristic curves of a triode having a pure tungsten filament. All of the curves ultimately level off at a definite temperature-limited or saturation value of plate current. The satura-
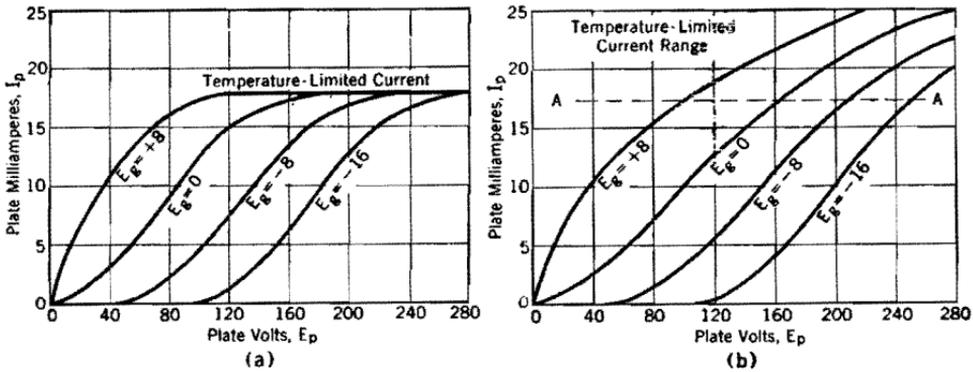
(a)	(b)
"Saturation" with a tungsten filament.	Failure to "saturate" with an oxide-coated filament. Thoriated filaments behave similarly.

FIG. 85. Temperature limitation of triode plate currents.

tion current is independent of plate and grid voltages, because it is for all practical purposes independent of the gradient just outside the cathode.

Tubes containing oxide-coated, thoriated, and many other types of composite cathode surfaces exhibit a marked failure to saturate with increasing plate voltage, as illustrated by Fig. 85b. Above values of plate current indicated by the line AA the current is truly *temperature-limited*, in that the gradient at the cathode surface is definitely positive and the effect of space charge on the determination of current magnitude insignificant. Yet the different plate characteristic curves do not level off at all, and it is obviously possible, even in the temperature-limited range, to control the magnitude of the current by variations in grid voltage. This has not proved to be a useful method of current control, however. There is often an appreciable time lag in the response of current to grid voltage changes, and the individual curves are not easily reproducible. Prolonged passage of a temperature-limited current at a high off-cathode gradient permanently modifies the thermionic properties of a composite emitting surface.

This failure to saturate indicates, of course, that one or both of the constants A_0 and E_W in Dushman's equation have for oxide-coated and thoriated surfaces different values for different off-cathode gradients. The curved solid line in Fig. 86 is the same kind of curve for an oxide-coated surface that the straight line in Fig. 80 is for a homogeneous

one.[73, T125]   At low values of field strength the available emission current is immensely more sensitive to variations in the field strength than is to be expected from the analysis presented in Section 100, as indicated by the steepness of the solid line of Fig. 86 near zero field strength.   At very large field strengths the emission approaches a straight Schottky line, the dotted line whose slope is as demanded by Equation (367).   Thus small fields produce striking and progressive modifications in the emission constants, but the variation for a given increment of gradient becomes less as the applied field is increased, and settles down to the normal effect at very large field strengths.

For comparing different surfaces and the results of different experimenters the values of A_0 and E_W corresponding to *zero-field emission* (vertical-axis intercept, Fig. 86) are often used.   In order to determine these, the zero-field emission current for each of several temperatures must be found by extrapolation from higher points, on a curve like the solid-line one in Fig. 86.   Zero-field emission current cannot be directly measured, because with zero field the current is space-charge-limited, not temperature-limited.
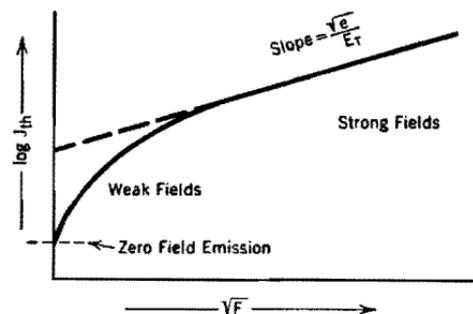
FIG. 86.   Variation of emission with field strength, for oxide-coated and [thoriated surfaces.   The straight portion is a Schottky line (see Fig. 80).

The reason for the marked dependence of the emission from composite surfaces on small and moderate values of off-cathode gradient are not well understood.   Reimann's book [TIV, TV] contains a good résumé and criticism of the various explanations that have been advanced. Presumably failure to saturate has something to do with the response to weak external fields of polarized layers and their " humps."

It has seemed to the author reasonable to suppose that an external field may, by force action on the exposed positive charges, push a polarized layer closer to the underlying substance.   This must reduce the height of the hump, as in Fig. 83*d*, so increasing the emission.   No further change in transmission or work function could result from still stronger fields after the crest of the hump is pushed below the E_0 level; this would satisfactorily explain the observed straight Schottky lines at very strong fields (Fig. 86).

PROBLEMS

Chapter IX

1. The temperature-limited current from a certain cathode is 50 milliamperes at 1700° K and 0.5 milliamperes at 1400° K. Assuming that Dushman's equation applies, find the work function of this cathode.

2. The cathode of Problem 1 has a circular cross-section, of 0.008 cm radius, and is 2 cm long. Outside and concentric with it is a cylindrical anode ("plate") that has a radius of 0.032 cm. Cathode temperature is 1700° K. Find the current flowing

(a) When $E_p = 30$ volts.

(b) When $E_p = 5$ volts.

(c) In case (b), how many volts below the cathode's potential is the lowest point on the potential distribution curve between cathode and plate.

3. Electrodes are as in Problems 1 and 2, temperature 1400° K, current therefore small enough so that the effect of space charge on potential distribution can be neglected. $E_p = 1200$ volts. What is the temperature-limited current? (See Section 100.)

4. The temperature of a certain tungsten filament is 2500° K.

′ (a) If the current is space-charge-limited, and the " negative dip " in the potential distribution curve just outside the cathode extends to 0.05 volt below the cathode potential, what is the ratio of the space-charge-limited current to the temperature-limited emission?

(b) Explain clearly just what factors control the depth of this dip in actual operation.

5. Suppose that, in a certain triode, the work function of the cathode surface is 2 volts, and that the grid is made of nickel, the plate of graphite (a form of carbon). Write an expression for the equivalent voltage for this triode which takes into account the contact differences of potential between the various electrodes.

6. Plot a curve, similar to the curved " Schottky line " of Fig. 86, from the portion of the $E_g = 0$ plate characteristic curve of Fig. 85b that lies above AA, also from the corresponding portions of the $E_g = +8$ and $E_g = -8$ plate characteristic curves. Plot log I_{th} vertically, as in Fig. 86, but plot the square root of the equivalent voltage horizontally. Use $\mu = 6.25$. Why does the use of the square root of the equivalent voltage as the abscissa produce a curve similar to a portion of the Schottky curve?

7. To what value must the work function of a tungsten surface be reduced, by an external field, to raise the current density to 20,000 amperes per sq cm at a temperature of 2500° K? How strong an external field would be required? How near must the adjacent parallel plane anode be if its potential is 10 volts above that of the tungsten?

CHAPTER X

DISTRIBUTIONS OF RANDOM VELOCITIES OF GAS PARTICLES

108. Dependence of Average Energy on Temperature, for Ordinary Gases and Electronic Gases within Metals. The particles of ordinary gases, and the electrons of the electronic gases within metals, are at all times in rapid motion, individual velocities varying greatly from particle to particle and from moment to moment. It is necessary to have some organized knowledge of the magnitudes of these *random velocities* in order satisfactorily to understand various electron emission phenomena and the behavior of conducting gases.

The dependence of random velocities and of the corresponding kinetic energies on various externally imposed conditions can be analyzed, both for electronic gases within metals and for ordinary gases, by means of the shuttling-particle mechanism described in Section 86. Such analysis shows that the average kinetic energy per particle depends not only on the *temperature*, as is to be expected, but also, especially at low temperatures, on the *mass per particle* and the *number of particles per cubic centimeter*.[69, T 5]

Everyday scientific and technical experience indicates that the average energy per particle is proportional to the temperature of a gas, but does not indicate any dependence on the concentration or on the mass per particle. This negative indication is correct for any gas, electronic or otherwise, at sufficiently high temperatures.[69, W II, V 80] In an ordinary monatomic gas, the average kinetic energy in electron volts is $\frac{3}{2} E_T$ for this everyday expectation, E_T being the electron-volt equivalent of temperature, defined in Section 71.

On the other hand, the theory presented in Sections 85–91 predicts that there must be for any gas, as for the electron gas within a metal, a definite minimum below which the average energy cannot go, even if the temperature falls to absolute zero.[69, T IV] This zero-temperature average energy is $\frac{3}{5} E_N$ (see Section 90) where E_N is dependent on mass per particle and concentration, in general accordance with Equation (332). The 2 under the N should be included only for an electron gas, and for any but an electron gas a mass-ratio factor must be used in the denominator of the numerical form.

During *approach* to absolute zero temperature the average energy depends on the *three* quantities above mentioned: temperature, mass per particle, and concentration. Experimental check of this low-temperature

expectation for monatomic gases such as helium, neon, and argon has been attempted. Complete verification of the theory is impossible because condensation to liquid form occurs before the minimum energy as a gas is very closely approached, but a trend that is very definitely in accord with the theoretical prediction is observable before gaseous properties disappear.

The nature of the transition between the high-temperature and low-temperature conditions is illustrated in Fig. 87.[69] The solid line there describes the variation with temperature of the average-particle energy for a monatomic gas for which $\frac{3}{5}E_N$ is 0.01 electron volt. It will be observed that at and above ordinary room temperatures ($68°$ F $= 20°$ C $= 293°$ K $= 0.0253$ electron volt) the average energy for such a gas is proportional to the temperature. The ideal gas law, Equation (305), holds in this upper range.

A gas with properties in accord with Fig. 87 would be called *nondegenerate* at and above room temperatures, whereas it would be called *degenerate* in the range below about $25°$ K ($248°$ below zero Centigrade). In this degenerate range the average energy is for all practical purposes down to the minimum, and varies only to a trifling extent with temperature. The shape of the curve shows that degeneracy is approached gradually with falling temperature. *Complete degeneracy* exists only at absolute zero temperature, and corresponds to the energy distribution shown in Fig. 71a, page 175.

Fig. 87. Variation of average monatomic gas-particle energies with temperature. Gas concentration held constant as temperature changes.

All that is necessary to make Fig. 87 descriptive of the behavior of the electron gas in a metal is to multiply both horizontal and vertical scales by a constant factor, for example, 100. The least average energy is then 1.0 electron volt, corresponding to a normal maximum kinetic energy of 1.667 electron volts, which is a little less than that for metallic caesium. The curve then indicates that the electronic gas within caesium *is degenerate at all temperatures below about* 2500° K, and, therefore, for all temperatures at which caesium can exist in solid or liquid form.

The use of a scale factor 550 makes the figure describe the properties of the electron gas in tungsten, assuming two free electrons per atom,

for it makes the minimum average energy 5.50 volts, which is just three-fifths of the proper value of normal maximum energy, 9.17 volts. The curve then indicates that the electron gas within metallic tungsten must be degenerate at all temperatures below about $25° \times 550 = 13,750° K$. But the melting point of tungsten is a little above $3000° K$. *The electron gases within all metals approach closely to complete degeneracy even at the highest temperatures employed for thermionic emission.*

The average *translational kinetic energy* due to u, v, and w velocities is $\frac{3}{2}E_T$ for all nondegenerate gases. However, diatomic gases, such as hydrogen, oxygen, and nitrogen, have two more " degrees of freedom " than monatomic gases.[W 81] That is, two additional kinds of motion are possible for each particle, and both must be quantized in considering total kinetic energy. They are: vibration between near-together and far-apart positions of the two atoms of each molecule, and rotation of each dumbbell-like molecule about an axis perpendicular to its greatest length. As a result, five instead of three quantum numbers are required to describe the entire instantaneous motion of any particle, and the quantum-number lattice, Fig. 70, must be five-dimensional. Also the average-particle kinetic energy for a diatomic gas in the high-temperature, nondegenerate straight-line part of Fig. 87 is $\frac{5}{2}E_T$ instead of $\frac{3}{2}E_T$, and the slope of the line $\frac{5}{2}$ instead of $\frac{3}{2}$. Each degree of freedom contributes $\frac{1}{2}E_T$ in this range. At low temperatures, the rotational and vibrational degrees of freedom become " suppressed," for reasons similar to those that produce ultimate degeneracy. Below about $75° K$, hydrogen behaves like a monatomic gas.

109. The " Most Probable " Energy Distribution. Of course individual velocities and energies of gas particles vary over a considerable range above and below the average values. The *distribution of energies* among the various particles of a gas can be described mathematically; it can also be represented graphically by an *energy distribution curve.* Such a curve is shown in Fig. 71b, page 175, for the electron gas in tungsten at $2500° K$. A similar diagram, adapted to the use of a velocity instead of an energy scale along the horizontal axis, can be used to describe the corresponding *distribution of velocities*; see Fig. 91b, page 225. The derivation, by means of the shuttling-particle mechanism, of the mathematical forms of velocity and energy distribution laws which such diagrams represent is rather tedious. The principles involved and the results are, however, not difficult to state.

It is obvious that an increase of average energy above the least possible value requires that some of the levels below the normal maximum (see Fig. 69, page 165) must be partially vacated, the particles that formerly occupied the vacated positions moving into levels above the normal

maximum.   The placement of lifted particles among the higher levels, and of vacated positions among the lower ones, takes place subject to the following conditions: [69]

(a) The total number of particles must remain unchanged.

(b) The average energy must be as called for by the temperature, as discussed in the preceding section.   The average energy for any given distribution of particles among the levels can be determined by the process used in Section 90, except that account must be taken of the fact that, at other than minimum-energy conditions, many levels may be partially vacant, partially filled, not completely one or the other.

(c) Requirements (a) and (b) can be satisfied for any given temperature by a very large number of widely differing energy distributions. For example, many different curves could be drawn in Fig. 71b which would call for the same number of total particles and the same average energy as the one shown there, and not require in any one level more particles than there are available positions.   All such curves would represent distributions that not only might, but actually do, exist at one time or another.   The existence of one or another at any particular moment is purely a matter of chance, and there is continual and inconceivably rapid shifting from one to another.   But there is one particular distribution whose chance of occurrence is very much greater than that for any of the others.   It may be called the *most probable distribution*, and is the one shown in Fig. 71b.   The chance of the existence, at any moment at all, of this distribution or one of the many that are almost identical with it, is very much greater than the combined chances for existence of all the others.   *Hence, it is a very close approximation to the true distribution almost all the time.*   The most probable distribution is, therefore, the one used in all scientific and engineering calculations. Most of this chapter is devoted to descriptions of the most probable distributions for ordinary (nondegenerate) and degenerate gases.

The reason why one energy distribution is more probable than any other is that more different arrangements of vacant and filled energy-level positions will produce it than will produce any other, subject to the (a) and (b) requirements.   Arrangements of vacant and filled positions, not arrangements of particles, are dealt with because the particles carry no tags.   Interchanging two particles does not produce a new arrangement, but moving one from a vacant to a filled position, or vice versa, does (see Section 87).

The existence of a most probable energy distribution is analogous to the existence of a "most probable number of spots " for a throw of dice. This number happens to be 7, for 7 spots can result from more number combinations than anything else the dice can show.   7 therefore ap-

pears more often than any other number. 6, 7, and 8, all "in the neighborhood of 7," are to be expected in $\frac{16}{36}$, or 44.4 per cent of all throws. In a gas the similarly combined expectations for distributions almost identical with the most probable add up to very nearly 100 per cent. The combined chances of existence, at any given moment, for all the distributions appreciably different from the most probable one total up to an extremely minute fraction of a per cent.

110. Symbols and Terminology for Distribution Curves and Equations. It is necessary to use a number of mathematical symbols, some new, some already familiar from previous discussions, in the presentation of the facts regarding velocity distributions among gas particles. Statements of the meanings of the various symbols are gathered together in this section in convenient groups to permit ready reference and make evident the system employed in their selection. Velocities are all in centimeters per second, masses in grams, and energies in ergs or electron volts.

Group I. Temperature (see Section 71).

kT, $E_T e 10^7$, $\frac{1}{2}m\alpha^2$, are all equal to one another; each describes in a particular way the kinetic energy, in ergs per particle, that is "characteristic" of the temperature T for any gas, whether ordinary or degenerate.

k = Boltzmann's gas constant, 1.371×10^{-16} erg per particle per degree Kelvin.

e = the electronic charge, 1.590×10^{-19} coulomb per electron.

m = the mass per gas particle. For gas molecules, m can be determined by multiplying the electronic mass, m_e, by the *mass ratio*.

m_e = the mass of an electron, 9.038×10^{-28} gram, or approximately 9×10^{-28} gram.

m/m_e = the *mass ratio*; numerically, as in Equation (149):

$$m/m_e = 1824 \times \text{molecular weight of gas particle.} \qquad (370)$$

T = temperature of the gas in degrees Kelvin, that is, absolute temperature on the Centigrade scale.

E_T = the characteristic temperature-energy in electron volts, as explained in Sections 70 and 71, that is,

$$T = 11600 E_T \qquad (371\ p)$$

α = the velocity corresponding to the characteristic temperature-energy; for purposes of calculation, using Equation (154),

$$\alpha = \frac{5.93 \times 10^7 \sqrt{E_T}}{\sqrt{m/m_e}} \qquad (372\ p)$$

also, from the equality between kT and $\frac{1}{2}m\alpha^2$

$$\alpha = \sqrt{\frac{2kT}{m}} \qquad (373)$$

Group II. Individual Velocities and Energies.

u, v, w are respectively x-, y-, and z-directed velocities of a gas particle.

c is the total velocity of a gas particle regardless of its direction, that is,

$$c^2 = u^2 + v^2 + w^2 \qquad (374)$$

s is the resultant of v and w, that is,

$$s^2 = v^2 + w^2 \qquad (375)$$

E_u, E_v, E_w, E_c, are respectively the energies, in electron volts, associated with the u, v, w, and c velocities. A typical relation is

$$u = \frac{5.93 \times 10^7 \sqrt{E_u}}{\sqrt{m/m_e}} \qquad (376 \text{ p})$$

du, dc, dE_u, dE_c, etc., are small increments of the corresponding velocities and energies. See under dc in Group VI for a comment on their magnitudes.

Group III. Least Energy and Work Function (see Section 83).

E_N is the kinetic energy, in electron volts, of a gas particle in the normal maximum level (Fig. 69), for any gas, electronic or otherwise.

E_W is the net work function of a metallic surface, in electron volts.

E_G is the kinetic energy, in electron volts, necessary to permit an electron to escape by surmounting the gross work function; from Section 83

$$E_G - E_N = E_W \qquad (377)$$

E_G must be the energy of x-*directed* velocity if escape is actually to occur.

c_N is the total velocity corresponding to the energy E_N; c_N and E_N are related as suggested by Equation (376).

u_G is the x-directed escape velocity that corresponds to E_G.

Group IV. Velocity and Energy Ratios. (Introduced to simplify the forms of mathematical expressions.)

$R^2 = \dfrac{c^2}{\alpha^2} = \dfrac{E_c}{E_T}$; R is the ratio of an individual particle's total velocity to the characteristic velocity α.

$r^2 = \dfrac{u^2}{\alpha^2} = \dfrac{E_u}{E_T}$; r is the ratio of a particular particle's x-directed velocity to the characteristic velocity.

$R_N^2 = r_N^2 = \dfrac{E_N}{E_T}$; R_N^2 and r_N^2 both describe the ratio of normal maximum energy to characteristic energy.

$r_G^2 = \dfrac{u_G^2}{\alpha^2} = \dfrac{E_G}{\overline{\overline{E}}_T}$; r_G^2 is the ratio of the x-directed energy that permits escape, to the characteristic energy.

$dR = \dfrac{dc}{\alpha} = \dfrac{d\,(E_c^{\frac{1}{2}})}{\sqrt{E_T}}$, from which

$$dR = \dfrac{1}{2\,\sqrt{E_c E_T}}\,dE_c \qquad (378)$$

and

$$dE_c = 2E_T R\,dR \qquad (379)$$

$dr = \dfrac{du}{\alpha} = \dfrac{d(E_u^{\frac{1}{2}})}{\sqrt{E_T}}$; relations similar to Equations (378), (379) exist between dr and dE_u.

Group V. Average and Root-Mean-Square Velocities and Energies.

$\bar{c}$, $\bar{R}$ are average values of c, R.

$\bar{u}$, r are the averages *of the positive values only* of u, r; overall averages are zero.

$\bar{\bar{c}}$, u, $\bar{\bar{R}}$, $\bar{\bar{r}}$ are root-mean-square values of c, u, R, r.

$\bar{\bar{c}}^2$, $\bar{\bar{u}}^2$, $\bar{\bar{R}}^2$, $\bar{\bar{r}}^2$ are, therefore, the *averages of the squares* of c, u, R, r.

$\bar{E}_c$, $\bar{E}_u$ are average total and average x-directed kinetic energies, in electron volts.

Since R^2, c^2, and E_c are, from their definitions in Group IV, directly proportional to one another, their averages, $\bar{\bar{R}}^2$, $\bar{\bar{c}}^2$, and $\bar{E}_c$ are also proportional, and by the same proportionality factors. $\bar{\bar{r}}^2$, $\bar{\bar{u}}^2$, and $\bar{E}_u$ are similarly proportional.   Therefore

$$\bar{\bar{R}}^2 = \dfrac{\bar{\bar{c}}^2}{\alpha^2} = \dfrac{\bar{E}_c}{E_T}$$
$$\bar{\bar{r}}^2 = \dfrac{\bar{\bar{u}}^2}{\alpha^2} = \dfrac{\bar{E}_u}{E_T} \qquad (380)$$

It will be shown later that for ordinary (that is, nondegenerate) gases, the following numerical relations hold regarding these averages

$$\bar{R} = \dfrac{\bar{c}}{\alpha} = \dfrac{2}{\sqrt{\pi}} \qquad\qquad \bar{\bar{R}}^2 = \dfrac{\bar{\bar{c}}^2}{\alpha^2} = \dfrac{3}{2}$$
$$\bar{r} = \dfrac{\bar{u}}{\alpha} = \dfrac{1}{\sqrt{\pi}} \qquad\qquad \bar{\bar{r}}^2 = \dfrac{\bar{\bar{u}}^2}{\alpha^2} = \dfrac{1}{2} \qquad (381)$$

Group VI. Concentrations of Particles.

N = concentration, that is, total number of gas particles per cubic centimeter.

N_c = the number of particles per cubic centimeter with total velocities greater than a value c.

$p_c = \dfrac{N_c}{N} = (a)$ when multiplied by 100, the *per cent* of the total number of particles that have, at any given moment, a total velocity greater than c. Thus p_c is just the *fraction* of the total number whose velocities exceed c; also

(b) the *probability* that any particle has a velocity greater than c.

dN_c = the number per cubic centimeter with total velocities between c and $c + dc$, where

dc = an increment of c that is small relative to the range of velocities considered, yet large enough so that dN_c is a very large number. The other incremental quantities listed in Group II and Group IV are similarly proportioned.

$dp_c = dN_c/N = (a)$ the small *fraction* of the total number of particles that have velocities between c and $c + dc$, also,

(b) the *probability* that any particle's total velocity is between c and $c + dc$.

N_u = the number per cubic centimeter with x-directed velocities greater than a value u.

dN_u = the number per cubic centimeter with x-directed velocities between u and $u + du$; as with dc, the increment du must be small, yet dN_u must be a large number.

$dp_u = dN_u/N = (a)$ a small *fraction*, also (b) a *probability*, both as for dp_c except that x-directed instead of total velocities are involved.

111. Maxwellian Distribution Curves for Total Velocities in an Ordinary Gas. The curves and equations that appear in this and the following sections describe the most probable distributions. Velocity distribution rather than energy distribution curves are presented because of their greater general utility. It is usually not difficult to convert from a velocity to an energy distribution expression by using the reverse of the process indicated by Equation (335). The mathematical expressions given for the distributions can be derived by analysis of the familiar box full of shuttling particles, using suitable mathematical principles to discover the distribution that can occur in the greatest number of different ways.

When the results of the derivation are applied to the nondegenerate gases of every-day experience, a relatively simple distribution relation, originally derived by Maxwell and by Boltzmann, results; hence such gases are said to have the Maxwellian or Maxwell-Boltzmann velocity distribution. The distribution is Maxwellian at any point along the straight part of the solid line in Fig. 87.
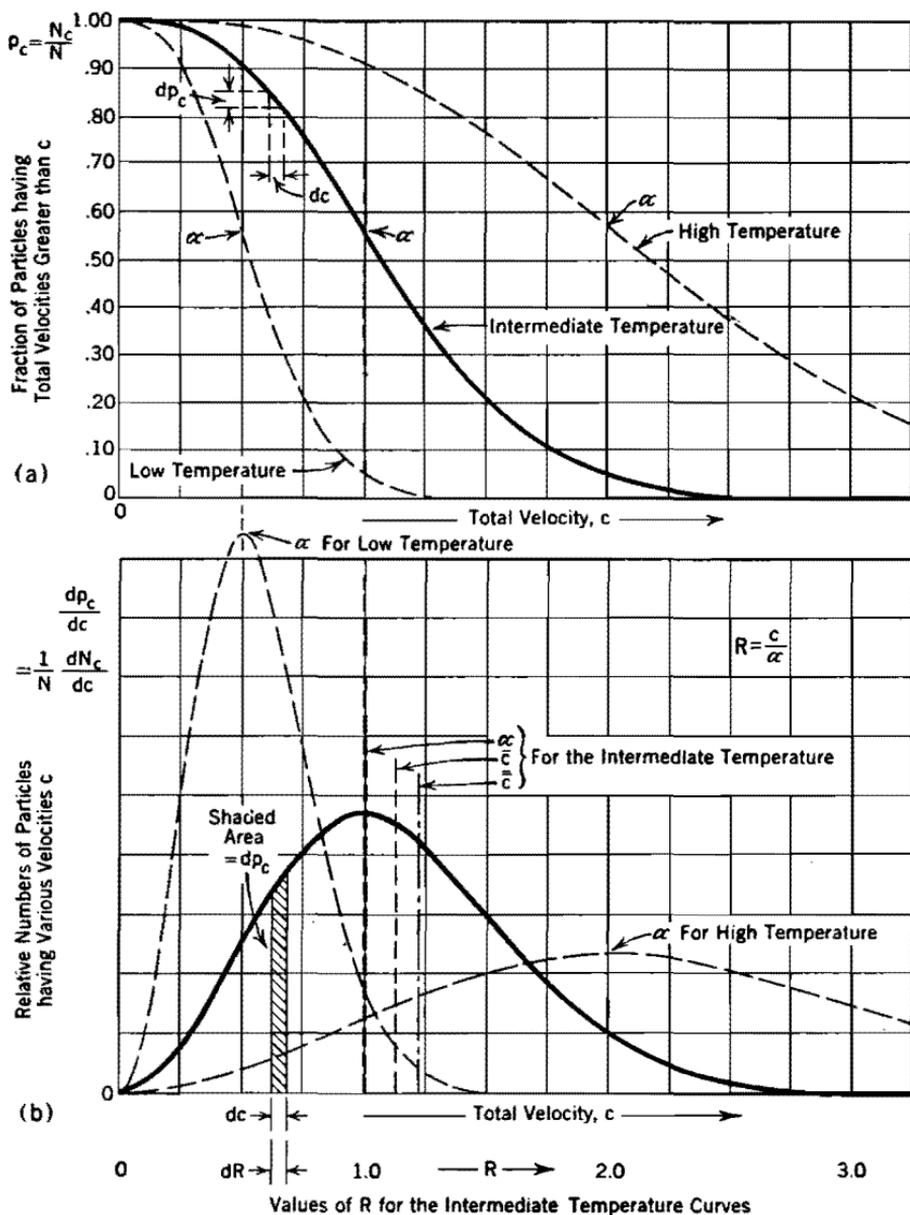
FIG. 88.   Maxwellian total velocity distributions.

(a) Integrated distributions of total gas-particle velocities, ordinary gases, Equation (391).

(b) Distributions of total gas-particle velocities, ordinary gases, Equations (382) and (383).

The three curves in Fig. 88a are of a type sometimes called *integrated distribution curves*. They describe, in a manner convenient for quantitative use, the Maxwellian distribution of total velocities in an ordinary gas for three different temperatures, low, intermediate, and high.

Ordinates in Fig. 88a describe simply the fraction, p_c, of the total number of particles whose total velocity exceeds a value c, which is measured along the horizontal axis. It will be seen that, for all three temperatures, about 57 per cent of the particles have velocities greater than the characteristic velocity α for that temperature. The meanings of the differential quantities dc and dp_c are indicated near the upper end of the intermediate-temperature curve.

To illustrate the manner of use of Fig. 88a, suppose that it is desired to discover how large a fraction of the total number of particles in a gas at 2000° K has total velocities corresponding to an energy greater than 0.3 electron volt. Then $E_c = 0.3$ and $E_T = 2000/11600 = 0.1723$. The ratio c^2/α^2, symbolized as R^2 in Group IV above, is simply $E_c/E_T = 0.3/0.1723 = 1.741$. Hence, $c/\alpha = \sqrt{1.741} = 1.32$. The velocity to be exceeded is thus 32 per cent greater than α. The point corresponding to $c = 1.32\alpha$ ($R = 1.32$, on the scale at the bottom of Fig. 88b) along the intermediate-temperature curve of Fig. 88a is now located. The ordinate at this point is observed to be about 0.33, which is the required fraction. Therefore about 33 per cent of the particles in a gas at 2000° K have total translational energies exceeding 0.3 volt.

Fig. 88b contains three *distribution curves* [W 73, V 82] that correspond exactly to the *integrated distribution curves* of Fig. 88a. The area, from some velocity $c = c$ out to $c = \infty$, under one of the curves in the lower figure, is just the ordinate of the corresponding curve in the upper figure, hence the use of the word "integrated" in the name for the upper curves. The *area* under an *increment*, dc, of a distribution curve is just dp_c, since it describes the increment of height in the corresponding integrated distribution curve. The total area under each of the lower curves must be just 1.00, as that is the maximum height of each upper curve. It follows from the general nature of the relationship between the two sets of curves that the ordinates of the lower ones are the slopes of the corresponding upper ones, except for a difference in sign.

112. Equations for Total-Velocity Maxwellian Distribution Curves; Average Total Velocity and Energy. The equation for the Maxwellian total-velocity distribution curves of Fig. 88b is [V 80]

$$\frac{dp_c}{dc} = \frac{4}{\alpha\sqrt{\pi}} \frac{c^2}{\alpha^2} \epsilon^{-\frac{c^2}{\alpha^2}} \tag{382}$$

This relation is somewhat more conveniently expressed in terms of the ratio R between c and α, as follows:

$$\frac{dp_c}{dR} = \frac{4}{\sqrt{\pi}} R^2 \epsilon^{-R^2} \tag{383}$$

also written

$$\frac{dN_c}{dR} = \frac{4N}{\sqrt{\pi}} R^2 \epsilon^{-R^2} \tag{384}$$

These are equations for the distribution curves of Fig. 88b when referred to a horizontal scale of R-values. A scale of that kind suitable for use with the intermediate-temperature curves of both Figs. 88a and 88b appears at the bottom of the diagram.

Distribution curves are not as satisfactory as integrated distribution curves for direct quantitative work, but they have compensating values. For example, they are graphically suggestive of the values of average and root-mean-square velocities, and of the corresponding Group V ratios; and it is the distribution-curve equation that is used in determining the various averages mathematically.

The root-mean-square velocity-ratio $\overline{\overline{R}}$ is just the radius of gyration, referred to the zero-velocity axis, of the geometric figure described by the distribution curve. The steps in its determination are as follows:

(a) The incremental area dp_c under the distribution curve at any increment dR is expressed, from Equation (383), as $4R^2\epsilon^{-R^2}dR/\sqrt{\pi}$.

(b) This incremental area is multiplied by the second power of R, because it is the "second moment," or moment of inertia, of the figure that is used in obtaining the radius of gyration. This gives $R^2 \cdot 4R^2\epsilon^{-R^2}dR/\sqrt{\pi}$.

(c) All such products are summed by integrating from $R = 0$ to $R = \infty$. If the figure were a sheet of unit mass per unit area, this would give its moment of inertia, also called its second moment.

(d) The result of the integration is divided by the area under the curve (mass of the sheet), which is in this case 1.00, thus giving $\overline{R}^2$, where $\overline{R}$ can be variously described as the root-mean-square value of R, the radius of gyration of the figure, and — because of the fact that the area is 1.00 — the second moment of the figure.

The mathematical statement of these operations is:

$$\overline{\overline{R}}^2 = \frac{\int_{R=0}^{R=\infty} R^2 \frac{4}{\sqrt{\pi}} R^2 \epsilon^{-R^2} dR}{1.00} = \frac{4}{\sqrt{\pi}} \frac{3}{8} \sqrt{\pi} = \frac{3}{2} \tag{385}$$

The integration formula used here, and other similar ones used later, will be found in Table VIII. On using Equation (385) in Equation (380), it appears that

$$\overline{E}_c = \tfrac{3}{2} E_T \tag{386}$$

Equation (386) confirms the statement, made in Sections 71 and 108,

that the average energy due to total translational velocities of gas particles is $\frac{3}{2}$ of the energy characteristic of the temperature.   Another form of this average energy relation is

$$\bar{\bar{c}} = \sqrt{\tfrac{3}{2}}\alpha = 1.224\alpha \tag{387}$$

Equation (387) confirms the upper right-hand formula in Equation (381).

The *average* values of c and R are found by a similar method.   The important change is that in the operation corresponding to (b) above, the incremental area is multiplied by the first instead of the second power of R, for $\bar{R}$ is the distance to the center of gravity of the distribution-curve figure, and is obtained by the use of the first moment.   The results are

$$\bar{R} = \frac{2}{\sqrt{\pi}} \tag{388}$$

$$\bar{c} = \frac{2\alpha}{\sqrt{\pi}} = 1.128\alpha \tag{389}$$

Equation (389) confirms the upper left-hand formula in Equation (381).

The occurrence of the crests of the distribution curves of Fig. 88b at points for which $c = \alpha$ illustrates graphically why the characteristic velocity α is sometimes called the *most probable velocity*.   These curves are higher at the points where $c = \alpha$ than elsewhere; therefore more particles must have velocities equal or near to α than to any other velocities. The points of inflection of the curves in Fig. 88a of course correspond to the crests in the lower figure.

113. Equations for Total-Velocity Maxwellian Integrated Distribution Curves.   The equation for the integrated distribution curves of Fig. 88a is obtained by integration of Equation (383), indicated as follows:

$$p_c = \frac{N_c}{N} = \frac{4}{\sqrt{\pi}} \int_{R=R}^{R=\infty} R^2 \epsilon^{-R^2}\, dR \tag{390}$$

This expression can be integrated by parts, as suggested in Table VIII, giving

$$p_c = \frac{N_c}{N} = \frac{2}{\sqrt{\pi}}\, R\epsilon^{-R^2} + (1 - \operatorname{erf} R) \tag{391}$$

The *error function*, also called the *probability integral*, used here, is defined as

$$\operatorname{erf} R = \frac{2}{\sqrt{\pi}} \int_{R=0}^{R=R} \epsilon^{-R^2}\, dR \tag{392}$$

Tables of numerical values of erf R, corresponding to upper limits between 0 and 3.0, will be found in Pierce's Tables of Integrals.[K][116]

Erf R rapidly approaches unity as R rises above 1.5, so that for large values of R the first term alone of Equation (391) gives a good approximation to p_c; see also Equation (394) below.

The following *asymptotic series* gives numerical values for erf R that are satisfactory, for most purposes, when R is greater than about 2.0:

$$\text{erf } R = 1 - \frac{\epsilon^{-R^2}}{R\sqrt{\pi}}\left[1 - \frac{1}{2R^2} + \frac{1\cdot 3}{(2R^2)^2} - \frac{1\cdot 3\cdot 5}{(2R^2)^3} + \cdots\right] \quad (393)$$

As usual with series of this type, the successive terms first decrease, then increase. Since the error in a numerical value for the series is not greater than the last term used, maximum accuracy is obtained by stopping with the smallest term, though often sufficient accuracy results from the use of only a very few terms. This is particularly true in the asymptotic series expression that results from combining Equations (391) and (393), which is

$$p_c = \frac{N_c}{N} = \frac{\epsilon^{-R^2}}{R\sqrt{\pi}}\left[2R^2 + 1 - \frac{1}{2R^2} + \frac{1\cdot 3}{(2R^2)^2} - \frac{1\cdot 3\cdot 5}{(2R^2)^3} + \cdots\right] \quad (394)$$

114. Curves and Equations for x-Directed Maxwellian Velocity Distributions. Figs. 89a and 89b are respectively integrated distribution curves and distribution curves for x-directed components of the Maxwellian velocities in an ordinary gas. The unit of measure along the horizontal scale is r, the ratio of a particular x-directed velocity to the characteristic velocity. Because of this choice of scale units, both curves can serve for any temperature at which a gas is " ordinary," that is, nondegenerate. Of course the curves and equations for y-directed and z-directed velocity distributions are identical in form with those given here for x-directed velocities.

The equation for the distribution curve, Fig. 89b, is [V 76]

$$\frac{dp_u}{dr} = \frac{1}{\sqrt{\pi}}\,\epsilon^{-r^2}, \quad \text{or,} \quad \frac{dN_u}{dr} = \frac{N}{\sqrt{\pi}}\,\epsilon^{-r^2} \quad (395)$$

The equation for the integrated distribution curve, Fig. 89a, is

$$p_u = \frac{N_u}{N} = \frac{1}{\sqrt{\pi}}\int_{r=r}^{r=\infty}\epsilon^{-r^2}\,dr = \frac{1}{2}\,(1 - \text{erf } r) \quad (396)$$

Numerical values of erf r are obtainable in the manner described in the preceding section. If the asymptotic series for erf r is used, Equation (396) becomes

$$p_u = \frac{\epsilon^{-r^2}}{2r\sqrt{\pi}}\left[1 - \frac{1}{2r^2} + \frac{1\cdot 3}{(2r^2)^2} - \frac{1\cdot 3\cdot 5}{(2r^2)^3} + \cdots\right] \quad (397)$$

The distribution curve in Fig. 89b is symmetrical about zero, because just as many particles must be moving east as are moving west, if the gas as a whole is to remain stationary. Hence the true average x-
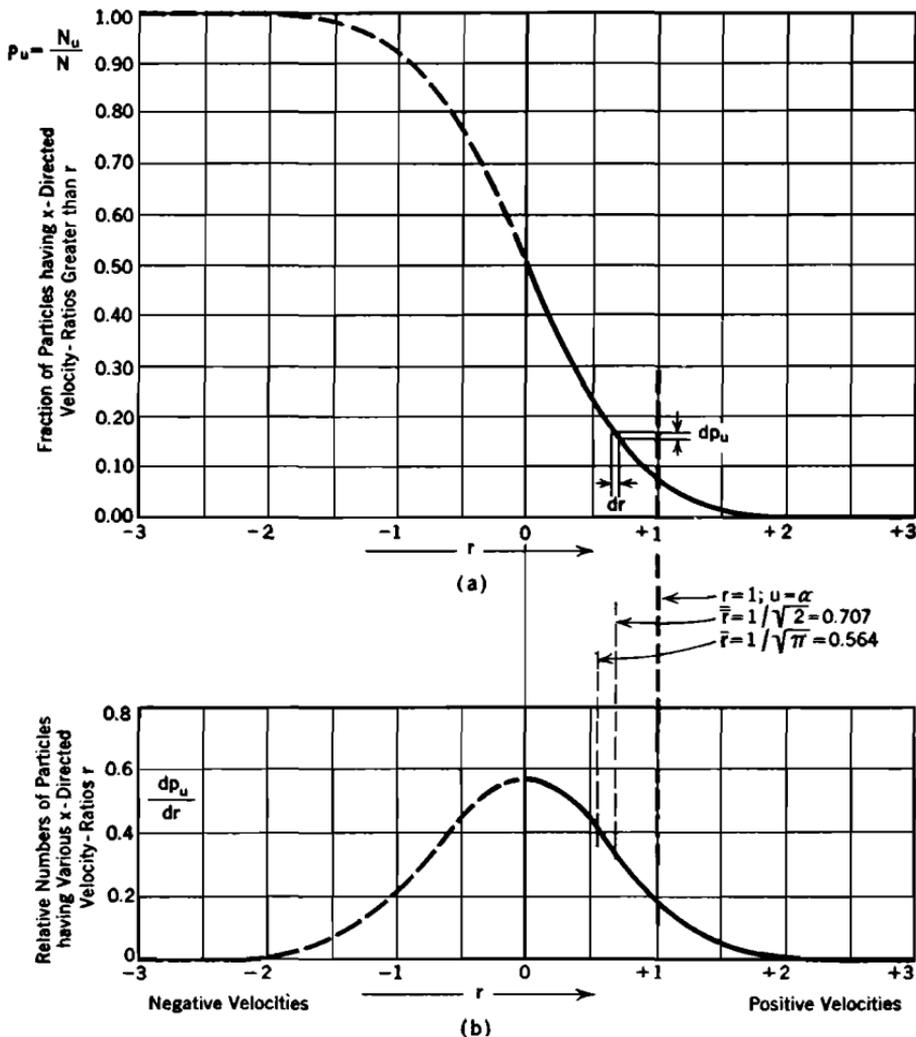
FIG. 89.   Maxwellian x-directed velocity distribution.

(a) Integrated distribution curve, Equation (396).

(b) Distribution curve, Equation (395).

directed velocity is zero, just as the true cyclic average value of a sine-wave alternating-current is zero.   Yet just as the average value of a *half-wave* of alternating current is of interest, so here the average velocity among the half of the particles with positive velocities is of interest and

is described by the symbol $\bar{u}$.   Hence $\bar{r}$ is the distance from the zero axis to the center of gravity *of one " wing "* of Fig. 89*b*. $\bar{\bar{r}}$ is the radius of gyration of either or both wings.

Integrations similar to that in Equation (385), applied however to only one wing, lead to the following results:

$$\bar{r} = \frac{1}{\sqrt{\pi}}, \quad \text{so that} \quad \bar{u} = \frac{\alpha}{\sqrt{\pi}} = 0.564\alpha \tag{398}$$

$$\bar{\bar{r}} = \frac{1}{\sqrt{2}}, \quad \text{so that} \quad \bar{\bar{u}} = \frac{\alpha}{\sqrt{2}} = 0.707\alpha \tag{399}$$

Equations (398) and (399) confirm the two lower formulas in Equations (381).   Equation (399) can be used to relate energies, as follows:

$$\bar{r}^2 = \frac{\bar{\bar{u}}^2}{\alpha^2} = \frac{\bar{E}_u}{E_T} = \frac{1}{2} \tag{400}$$

The following relations of particular interest grow out of the averages just stated and the corresponding total-velocity averages:

(1) On comparing Equations (389) and (398), it appears that *the average x-directed velocity is half the average total velocity*, that is,

$$\bar{r} = \frac{\bar{R}}{2} \quad \text{and} \quad \bar{u} = \frac{\bar{c}}{2} \tag{401}$$

(2) On comparing Equations (386) and (400), it appears that the average x-directed energy is, in agreement with what is called the *principle of equipartition of energy* (equipartition among the degrees of freedom), one-third of the average total translational energy, that is,

$$\bar{E}_u = \frac{1}{3}\bar{E}_c \tag{402}$$

115. Relations between Total and x-Directed Maxwellian Distribution Equations.   Agreement between total-velocity distribution Equation (383) and x-directed velocity distribution Equation (395) will be demonstrated by means of a velocity coordinate system similar to the lattice-point array in Fig. 70.   The relations between the velocity coordinates are shown in Fig. 90.

Just as many velocity points exist in this three-dimensional velocity space as there are gas particles in a single cubic centimeter of real space, each velocity point being placed in accordance with the proper u, v, w components.   It is convenient to introduce a symbol η (" eta ") to describe the concentration of points in this velocity space.   Just as N stands for the number of particles per cubic centimeter anywhere in real

space, so η stands for the number of particles per cubic centimeter per second at any location in the velocity space.   There are just N points altogether in the velocity space.

The analysis used with these velocity points will resemble that used in connection with the lattice points of Fig. 70; the following comparisons may be helpful:

(a) Coordinates in Fig. 70 described n-values which were directly proportional to velocities; here the velocity components themselves provide the coordinate scale.

(b) The present figure occupies all eight octants, not just one, for u, v, w are all as often negative as positive, while the n's could be only positive.

(c) In the present figure, points are located so as to describe only velocities *that actually do exist in the gas*. With Fig. 70 all possible lattice points were spotted, then a distinction made between "occupied" and "unoccupied" points.   In the present discussion, there is one point for each particle, no more, no less.
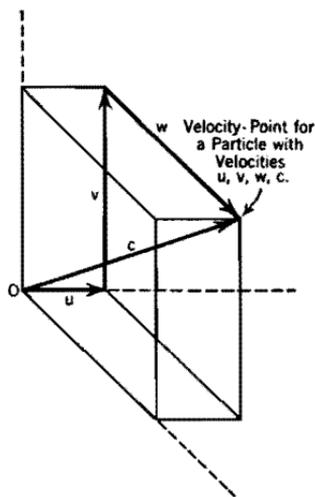
FIG. 90.   Coordinate system for velocity points.

dN_c has been defined as the number of particles per cubic centimeter having velocities between c and $c + dc$, regardless of the direction of motion.   It is, therefore, the number of points within a *spherical shell of velocity space* of radius c, thickness dc.   The quantity dN_u is similarly the number of points within a perfectly plane "slice," perpendicular to the u axis, located u from the origin and with thickness du.

No one direction of motion can have preference over any other, if the gas as a whole is to remain stationary; therefore the number of particles in any spherical shell must be equally distributed throughout it.   Hence η can be evaluated by dividing dN_c by $4\pi c^2\, dc$, the volume of the shell, and using Equation (382), as follows:

$$\eta = \frac{dN_c}{4\pi c^2\, dc} = \frac{N\, dp_c}{4\pi c^2\, dc} = \frac{N\epsilon^{-\frac{c^2}{\alpha^2}}}{\alpha^3 \pi^{\frac{3}{2}}} \tag{403}$$

The concentration η at any radius is now available, so that it is not difficult to determine the number of particles in the above-mentioned slice of thickness du at u.   Divide the slice into small rings about the axis, each of radius s and extent ds, where $s^2 = v^2 + w^2$ (s is obviously perpendicular to u).   All points of each ring are at a common distance $c = \sqrt{u^2 + s^2}$ from the origin; therefore η is uniform within each ring and the number of points in each is $2\pi\eta s\, ds\, du$.   Integration over all rings gives the number of points in the slice.   In carrying out the inte-

gration it is necessary to split the exponential in Equation (403) into the product $\epsilon^{-\frac{u^2}{\alpha^2}} \epsilon^{-\frac{s^2}{\alpha^2}}$. The following expressions result:

$$dN_u = \int_{s=0}^{s=\infty} 2\pi\eta s \, ds \, du = \frac{N}{\sqrt{\pi}} \epsilon^{-\frac{u^2}{\alpha^2}} d\left(\frac{u}{\alpha}\right) \int_{\frac{s}{\alpha}=0}^{\frac{s}{\alpha}=\infty} 2\frac{s}{\alpha} \epsilon^{-\frac{s^2}{\alpha^2}} d\left(\frac{s}{\alpha}\right) \quad (404)$$

Integration by the aid of Table VIII gives

$$dN_u = \frac{N}{\sqrt{\pi}} \epsilon^{-r^2} \, dr \quad (405)$$

Equation (405) is in entire agreement with Equation (395). The corresponding degenerate gas interrelation is demonstrated similarly in a later section.

The agreement between the two kinds of distributions in a Maxwellian gas can also be demonstrated by making use of the fact that the incremental directional *probabilities* dp_u, dp_v, dp_w are all independent of one another. That is, dp_u, expressed from Equation (395), is the probability that the velocity of a particle is between u and $u + du$, regardless of the values of v and w. As these three probabilities are wholly independent of one another, their product gives the probability of a particle's having velocities near u, v, w, simultaneously. This reasoning can be pursued to give Equation (383).

Since the directional probabilities are not independent in a degenerate gas, the velocity-slice method must be employed for such a gas. For example, in a completely degenerate gas a particle whose x-directed velocity is that corresponding to E_N can have no velocity at all in the other two directions. This restrictive influence has considerable effect at temperatures approaching degeneracy.

116. Total-Velocity Distribution Curves and Equations for a Degenerate Gas. It is not possible to draw a single distribution curve that will serve for all degenerate gases at all temperatures, as was done for ordinary gases. The most satisfactory procedure is to choose, for illustrative purposes, a particular electron gas that is frequently used, has properties typical of those in thermionic emitters generally, and makes a readable diagram. Accordingly, the solid lines in Fig. 91 illustrate the two familiar forms of total-velocity distribution curves for the electron gas within tungsten at the customary operating temperature of 2500° K, assuming two free electrons per tungsten atom.

Curves for the electron gases in other thermionic emitters are similar to these, in that they represent distributions that correspond to points near the lower end of the curve in Fig. 87, where the variation of average

energy with temperature has nearly ceased.   A gas in any such state is nearly but not completely degenerate.
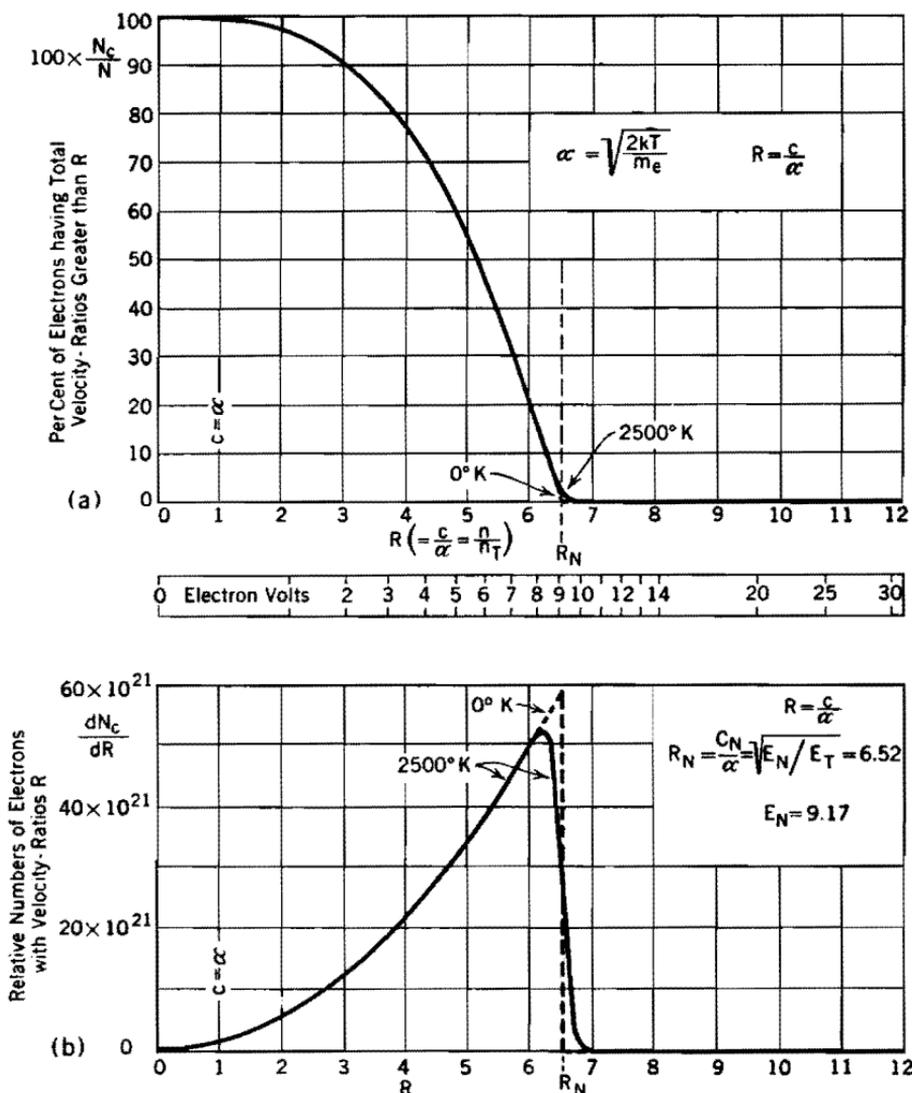
FIG. 91.   Total-velocity distribution among electrons in tungsten at 2500° K, two electrons per atom.

(a) Integrated distribution curve, Equation (408).
(b) Distribution curve, Equation (407).

The velocity distribution for the electron gas within tungsten in the completely degenerate condition (0° K, "normal state") is indicated by the dotted-line extension to a peak in Fig. 91b.   It is just as simple

as the corresponding energy distribution curve, Fig. 71a, page 175 being a parabola with a vertical axis.   The contrast between the dotted and solid lines in Fig. 91b shows clearly how the distribution curve changes when the temperature rises.   As a particular electron gas has been used, it has been possible to put an energy scale as well as a velocity scale along the horizontal axis.

It is convenient to write the equation for distribution curves like that in Fig. 91b in terms of dN_c rather than of dp_c.   An interesting form of the equation corresponding to Fig. 91b is the following: [69, U 543]

$$\frac{dN_c}{dc} = \frac{8\pi m^3 c^2}{h^3} \frac{1}{\epsilon^{\frac{c^2 - c_N^2}{\alpha^2}} + 1} \tag{406}$$

This expression shows clearly what happens at absolute zero, when α vanishes: If then $c < c_N$, the exponential in the denominator becomes $\epsilon^{-\infty}$ which is zero; dN_c/dc is then directly proportional to c^2.   At absolute zero, therefore, this equation describes a distribution curve that is a parabola for all values of c less than c_N, hence of R less than R_N, as shown by the dotted line.   If, however, $c > c_N$, the exponential becomes $\epsilon^{+\infty}$, which is infinite; therefore $dN_c/dc = 0$ for all values of R greater than R_N.

The following form is somewhat easier to read, and simpler to work from numerically, when the temperature is not zero: [69]

$$\frac{dN_c}{dR} = \frac{8\pi m^3 \alpha^3 R^2}{h^3} \frac{1}{\epsilon^{R^2 - R_N^2} + 1} \tag{407}$$

The following numerical relations apply for the solid-line curve in Fig. 91b:

$N = 122.3 \times 10^{21}$ electrons per cc, determined by using 2 electrons per atom, specific gravity 18.7, atomic weight 184.0, for tungsten.[Z]

$T = 2500°$ K, so that $E_T = 2500/11600 =$ ‹0.2155

$E_N = 9.17$ electron volts (from Equation 332)

$R_N^2 = E_N/E_T = 42.5$

$R_N = 6.52$

α is determined from Equation (372)

$m = m_e$

It is illuminating to study the numerical value of the fraction containing the exponential in Equation (407) at three different positions not far removed from $R = R_N$, as follows:

(1) $R = 6.0$.   Here $R - R_N^2 = 36 - 42.5 = -6.5$; the exponential term is $\epsilon^{-6.5} = 1/660$, which is quite insignificant relative to unity.   The fraction itself is,

therefore, practically unity, so that for all values of R less than about 6, dN_c/dR is proportional to R^2; that is, the curve follows the zero-temperature parabola.

(2) $R = R_N$. The exponential is $\epsilon^0 = 1$, the fraction is $\frac{1}{2}$; dN_c/dR is at exactly half the zero-temperature crest.

(3) $R = 7.0$. Here $R^2 - R_N = 49.0 - 42.5 = +6.5$, so that the exponential is $\epsilon^{+6.5} = 660$; the 1 is now insignificant, and the fraction a very small one; the curve is merging with the zero axis. Since the factor by which the fraction is multiplied is very large, dN_c/dR may still have a large numerical value, corresponding to the ability of the metal to release at this temperature a measurable current over the gross work function barrier of 13.71 electron volts.

The equation for the integrated distribution curve shown in Fig. 91*a* is the integral of Equation (407), evaluated from $R = R$ to $R = \infty$. The integration is indicated as follows:

$$p_c = \frac{N_c}{N} = \frac{8\pi m^3 \alpha^3}{N h^3} \int_{R=R}^{R=\infty} \frac{R^2\, dR}{\epsilon^{R^2 - R_N^2} + 1} \tag{408}$$

This integral can be expressed as the sum of two infinite series, one for the range $R < R_N$, another for the range $R > R_N$.

However, in this particular case, the magnitudes are such that for most values of R the distribution curve may be treated as a parabola, and very simply integrated. This takes care of all except the lower end of the integrated distribution curve; in preparing Fig. 91*a* the ordinates of the curve at its lower end were determined by graphical integration of the distribution curve.

Equations (382) and (406) are both special forms to which a more general equation reduces at extreme or near-extreme conditions. The general equation is expressible only by the use of an infinite series.[69] Many terms of the series are necessary to describe accurately the velocity distributions that correspond to points that lie along the bending portion, yet not near the lower end, of the average-energy curve, Fig. 87.

117. *x*-Directed Velocity Distribution Curves and Equations for a Degenerate Gas. Fig. 92 contains the *x*-directed integrated distribution and distribution curves that correspond to the total-velocity curves in Fig. 91. The *x*-directed distribution curve is, as in Fig. 89*b* and for the same reason, symmetrical about zero velocity. No attempt has been made to show the completely degenerate curves here; they would be very similar to the 2500° curves except that both would become zero at $r = r_N$.

The equation for curves like that in Fig. 92*b* is [82]

$$\frac{dN_u}{dr} = 2\pi \left(\frac{\alpha m}{h}\right)^3 \log\left(\epsilon^{r_N^2 - r^2} + 1\right) \tag{409}$$

When r is appreciably smaller than r_N, the quantity $\epsilon^{r_N^2 - r^2}$ is so large
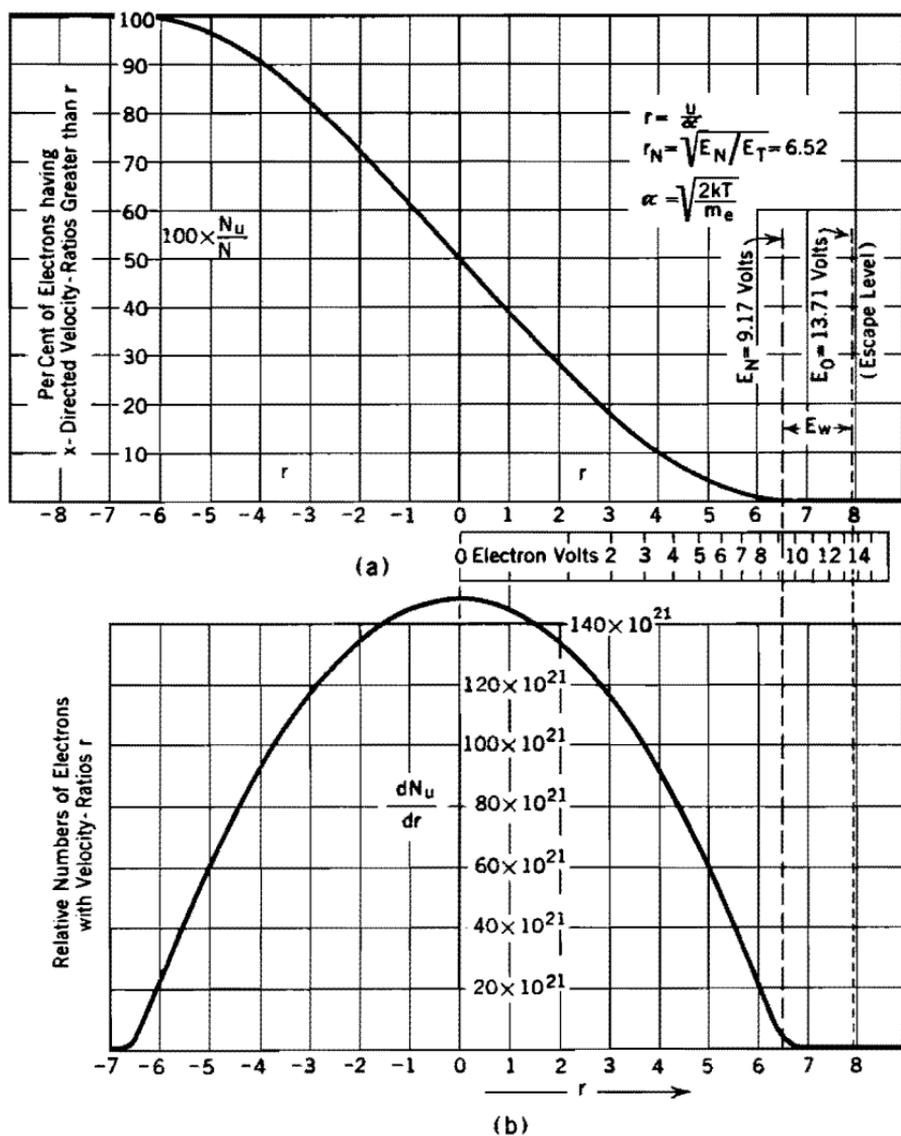
FIG. 92. x-directed velocity distributions among electrons in tungsten at $2500°$ K, two electrons per atom.

(a) Integrated distribution curve.

(b) Distribution curve, Equation (409).

that the 1 added to it can be neglected. Then the logarithmic factor reduces as follows:

$$\log \ (\epsilon^{r_N^2 - r^2} + 1) \rightarrow \log \epsilon^{r_N^2 - r^2} = r_N^2 - r^2 \tag{410}$$

Equation (410) shows why the x-directed distribution curve is, for the most part, parabolic in form. By writing the exponential in a form similar to that used in Equation (406) and letting α vanish, it can be shown that the distribution curve must plunge into the zero axis at $r = r_N$ when the temperature is 0° K.

Integration to obtain an exact mathematical expression for the integrated distribution curve is, as in the previous section, very difficult. The greater part of the curve can, however, be satisfactorily located by integration of the parabolic portion of the distribution curve, and the rest by a step-by-step method, graphical or otherwise.

x-directed velocity distribution Equation (409) can be derived from total-velocity distribution Equation (407) in exactly the manner used in Section 115 for demonstrating agreement between the corresponding ordinary gas equations. The steps are as follows, referring to the velocity coordinate system of Fig. 90 and using the symbolism of Section 115:

$$\eta = \frac{dN_c}{4\pi c^2 \, dc} = \frac{8\pi m^3 c^2}{h^3} \frac{1}{\left[\epsilon^{\frac{c^2 - c_N^2}{\alpha^2}} + 1 \right]} \frac{1}{4\pi c^2} \tag{411}$$

$$= 2\frac{m^3}{h^3} \frac{1}{\left[\epsilon^{\frac{c^2}{\alpha^2} - \frac{c_N^2}{\alpha^2}} + 1 \right]} \tag{412}$$

Now

$$\frac{c^2}{\alpha^2} = \frac{u^2}{\alpha^2} + \frac{s^2}{\alpha^2} \tag{413}$$

So that

$$\frac{c^2}{\alpha^2} - \frac{c_N^2}{\alpha^2} = \frac{u^2}{\alpha^2} - \frac{c_N^2}{\alpha^2} + \frac{s^2}{\alpha^2} \tag{414}$$

$$= r^2 - r_N^2 + \frac{s^2}{\alpha^2} \text{ (for } R_N = r_N\text{)} \tag{415}$$

This used in Equation (411) gives

$$\eta = 2 \frac{m^3}{h^3} \frac{1}{\left[\epsilon^{r^2 - r_N^2} \, \epsilon^{\frac{s^2}{\alpha^2}} + 1 \right]} \tag{416}$$

From Equation (416) dN_u is determinable, for

$$dN_u = \int_{s=0}^{s=\infty} 2\pi \eta s \, ds \, du \tag{417}$$

$$= 4\pi \, \frac{\alpha^3 m^3}{h^3} \, d\left(\frac{u}{\alpha}\right) \int_0^\infty \frac{\frac{s}{\alpha} d\left(\frac{s}{\alpha}\right)}{\left[\epsilon^{r^2 - r_N^2} \, \epsilon^{\frac{s^2}{\alpha^2}} + 1 \right]} \tag{418}$$

On using Part E of Table VIII to evaluate the integral, Equation (418) becomes just Equation (409).

118. High-Velocity x-Directed Distribution for the Electrons within a Metal. One of the major reasons for the presentation, in this text, of the velocity distribution curves and equations for degenerate gases is the fact that such distributions control the rate of escape of electrons over the work-function barrier at a thermionic surface. For example, quantitative study of the thermionic emission problem, which leads to the derivation of Dushman's Equation (299) for thermionic current density, requires the use of the x-directed velocity distribution described by Equation (409).

The thermionic emission analysis involves only the electrons whose x-directed velocities are more than enough to permit escape over the barrier. Therefore the distribution of velocities in the higher ranges, where r is substantially greater than r_N, is of particular interest.

For this high-velocity range, the quantity $\epsilon^{r_N^2 - r^2}$ in Equation (409) is small relative to unity, for the exponent $r_N^2 - r^2$ has a considerable negative value. It is, therefore, satisfactory to use the relation described by Equation (122); that is, all except the first powers of $\epsilon^{r_N^2 - r^2}$ can be dropped out in the type of expansion indicated by Equation (121). Hence in the interesting range the x-directed velocity distribution equation for electrons within a metal has the form

$$\frac{dN_u}{dr} = 2\pi \left(\frac{\alpha m}{h}\right)^3 \epsilon^{-(r^2 - r_N^2)} \tag{419}$$

Now

$$r^2 = \frac{E_u}{E_T} \quad \text{and} \quad r_N^2 = \frac{E_N}{E_T}$$

so that

$$r^2 - r_N^2 = \frac{E_u}{E_T} - \frac{E_N}{E_T} \tag{420}$$

Equation (409) can therefore be written

$$\frac{dN_u}{dr} = 2\pi \left(\frac{\alpha m}{h}\right)^3 \epsilon^{-\frac{E_u - E_N}{E_T}} \tag{421}$$

when E_u is a volt or so larger than E_N. Equation (421) is of particular interest in that it shows that the number of electrons having a particular x-directed velocity *depends primarily on how much their x-directed energy exceeds the normal maximum energy* E_N. Equation (419) will be employed in a later section for the derivation of Dushman's equation.

PROBLEMS

CHAPTER X

1. To what gas concentration do the numerical values used in Fig. 87 correspond?

2. If the *shape* of the solid-line curve in Fig. 87 were quantitatively correct, to what temperature would helium gas, at the concentration represented by standard atmospheric conditions, have to fall before its actual average energy would become 25 per cent more than that predicted by the ideal gas law?

3. Demonstrate that, as stated in Section 109, the probability that a throw of dice will turn up some one of the numbers 6, 7, or 8 is $\frac{16}{36}$.

4. The electron gas of a certain gaseous discharge has a temperature of 15,000° K and a Maxwellian velocity distribution.

(a) What is the voltage equivalent of this temperature?

(b) How large a fraction of the total number of electrons have total velocities exceeding 6×10^8 cm per sec?

(c) How large a fraction have x-directed velocities exceeding 6×10^8 cm per sec?

(d) What kinetic energy, in electron volts, corresponds to this velocity?

(e) Find the average total velocity. If each electron travels on the average 8 cm between successive collisions with gas particles (not with other electrons), how many such collisions will each electron experience per second?

5. (a) What per cent of the particles in a Maxwellian distribution have total velocities greater than the average total velocity?

(b) What per cent have energies greater than the average total energy?

(c) What per cent have positive x-directed velocities greater than the average positive x-directed velocity?

(d) What per cent have energies due to positive x-directed motion that are greater than the average x-directed energy?

6. What per cent of the particles in a Maxwellian distribution have total velocities greater than 10 times the characteristic velocity? What per cent have total energies greater than 10 times the characteristic energy?

7. Derive an expression for the number of Maxwellian gas particles whose velocities in all directions *perpendicular to some definite axis* have values between s and $s + ds$ (see Section 115).

8. What is the most probable total energy, in terms of E_T, in a Maxwellian distribution? The most probable x-directed energy?

9. Demonstrate mathematically the agreement between Equation (333) and Equation (406), if the temperature is 0° K.

10. Determine what fraction the average x-directed energy in a completely degenerate gas is of the normal maximum energy.

11. Find how large a fraction of the electrons in tungsten occupy levels more than 20 per cent above the normal maximum level,

(a) at 1500° K.

(b) at 3000° K. (Expand Equation (408) as an infinite series, for the case $R > R_N$, and integrate. See also Table VIII.)

12. Derive an expression for the number of degenerate gas particles whose x-directed energies exceed a value E_u electron volts.

13. Suppose that the distribution of velocities v among falling raindrops is expressible mathematically as follows:

$$dN_v = \frac{2N}{b\sqrt{\pi}} \, \epsilon^{-\left(\frac{v-a}{b}\right)^2} dv$$

Here N is the number of raindrops per cubic foot, a a characteristic velocity in feet per second, and b a velocity roughly descriptive of the deviation of individual velocities from the characteristic velocity.

(a) Plot the corresponding distribution curve of dN_v/dv against v, letting $b = \frac{1}{10}a$. Locate points at $v = a - b, v = a, v = a + b$.

(b) Write the equation for and plot the corresponding integrated distribution curve.

(c) Derive mathematical expressions for the average, most probable, and root-mean-square velocities, in terms of a and b.

CHAPTER XI

ELECTRICAL EFFECTS OF RANDOM MOTIONS

119. Rate at which Gas Particles Arrive at a Boundary Wall. The derivation of Dushman's equation for thermionic emission, and the handling of a number of problems having to do with conducting gases, require a knowledge of the number of gas particles or electrons that hit per second within an area of one square centimeter of an enclosing boundary wall; also of the number that so hit with x-directed energies exceeding any given value.

Fig. 93 illustrates a portion of an enclosing wall. Suppose that it is desired to determine the number of particles whose random motions cause them to strike one square centimeter of the boundary in a short time interval, dt seconds. There are shown in the figure imaginary boundaries of a region shaped like a rectangular box adjacent to the wall, the two ends each being one square centimeter in area. The length of the region perpendicular to the wall is $u\,dt$ centimeters, where u is any particular x-directed velocity. Any particle *within this region* that has a velocity between u and $u + du$ centimeters per second at the beginning of the time interval dt must of course arrive at the wall by the end of the interval. Those initially near the wall arrive in very much less time than dt; those far out just barely arrive in the time dt.
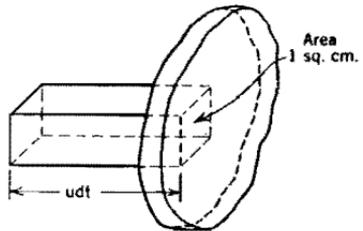
FIG. 93.   Rate of arrival at a wall of particles with velocity u.

The y- and z-directed velocities that the various particles possess are of no consequence, for as many particles of any selected x-directed velocity must enter from one side of the imaginary boxlike region as leave from the opposite side, during the chosen interval dt.

Thus the number of particles with velocities between u and $u + du$ which, in dt seconds, strike the square centimeter that terminates this region on the wall is just the number that was in it at the beginning of the chosen interval. This number can be determined by multiplying the volume $u\,dt$ of the region by the number dN_u of such particles per

233

cubic centimeter.   Mathematically:

$$\left.\begin{array}{l}\text{Number of particles with velocities}\\\text{between } u \text{ and } u + du \text{ that strike}\\\text{one square centimeter in } dt \text{ seconds}\end{array}\right\} = u \, dt \cdot dN_u \qquad (422)$$

Division by the duration dt of the time interval gives a *rate*, that is, the number of particles of the selected velocity that arrive per second, symbolized by dL_u, as follows:

$$dL_u = u \, dN_u \qquad (423)$$

Then L, *the number of particles of all velocities* that arrive per square centimeter per second is obtained by integration over all values of u; the integration is indicated as follows:

$$L = \int_{u=0}^{u=\infty} u \, dN_u \qquad (424)$$

Later on, in Section 127, it will be shown how to use this equation for determining the magnitudes of the random ion and electron current densities in a conducting gas.

L_u, the number of particles arriving per square centimeter second with x-directed velocities *exceeding* some definite value u, is obtained by integration between $u = u$ and $u = \infty$, indicated as follows:

$$L_u = \int_{u=u}^{u=\infty} u \, dN_u \qquad (425)$$

All these equations can be expressed in terms of the velocity ratio $r(= u/\alpha)$; for example:

$$dL_u = u \, dN_u = \alpha \frac{u}{\alpha} dN_u = \alpha r \, dN_u \qquad (426)$$

When Equation (426) is used in Equation (425), it appears that the rate of arrival of particles with x-directed velocities exceeding a definite value u, corresponding to a definite velocity-ratio r, is:

$$L_u = \alpha \int_{r=r}^{r=\infty} r \, dN_u \qquad (427)$$

Equations (424), (425), and (427) can be made to suit either ordinary or degenerate gases by the respective use of Equation (405) or Equation (409) to express dN_u.   For the thermionic emission problem, Equation (419) is satisfactory.

120. Derivation of Dushman's Equation for Thermionic Current Density.   The following operations are necessary to convert Equation

(427) into an expression for the thermionic current density due to the escape of electrons over the gross work function barrier:*

(a) The lower limit of integration in Equation (427) must be r_G, corresponding to the escape-velocity u_G. Only electrons with x-directed energies greater than the gross work function E_G can escape; see Groups III and IV, Section 110.

(b) L_u must be multiplied by e, the electronic charge, to give the current density carried by the electrons. With the lower limit established as in (a), this operation gives J_{th}, the thermionic current density.

(c) Equation (419) can be used to express dN_u, because the gross work function is invariably high enough above the normal maximum level to make valid the approximation used in arriving at Equation (419).

(d) α must be expressed in terms of temperature.

(e) The exponent in the resulting expression must be expressed in terms of energy values, that is, using Equation (377)

$$r_G^2 - r_N^2 = \frac{E_G}{E_T} - \frac{E_N}{E_T} = \frac{E_G - E_N}{E_T} = \frac{E_W}{E_T} \tag{428}$$

The mathematical statement of the operations just described is as follows:

$$J_{th} = e\alpha \int_{r=r_G}^{r=\infty} r\, dN_u \tag{429}$$

$$= \frac{2\pi e m_e^3}{h^3} \alpha^4 \epsilon^{r_N^2} \int_{r_G}^{\infty} r\epsilon^{-r^2}\, dr \tag{430}$$

$$= \frac{\pi e m_e^3}{h^3} \alpha^4 \epsilon^{-(r_G^2 - r_N^2)} \tag{431}$$

Using Equation (428), and substituting for α its value in terms of T, this becomes

$$J_{th} = \frac{4\pi m_e e k^2}{h^3} T^2 \epsilon^{-\frac{E_W}{E_T}} \tag{432}$$

Equation (432) is in fact Dushman's equation; a more familiar form, first used in Section 70, is

$$J_{th} = A_0 T^2 \epsilon^{-\frac{E_W}{E_T}} \tag{433}$$

According to this derivation, the constant A_0 in Dushman's equation should be

$$A_0 = \frac{4\pi m_e e k^2}{h^3} = 120.4 \frac{\text{amperes}}{\text{cm}^2 \text{ deg}^2} \tag{434p}$$

An important feature of the theoretical derivation of Dushman's equation is that it predicts the same value of A_0 for all clean metals; thus

---

* The derivation of Dushman's equation from Fermi statistics, given by Reimann,[T IV] uses the energy distribution rather than the velocity distribution approach.

Equation (434) contains only universal constants of nature. The convergence of the various lines of Fig. 64 toward a common point confirms this prediction rather strikingly.

The value most commonly obtained for A_0 experimentally is 60, just half the theoretical prediction. A great deal of thought and experimentation has been expended, with only moderate success, in the search for the reason for this two-to-one disagreement. Yet the fact that this theory predicts, entirely from universal constants of nature, as nearly correct a value of A_0 as it does is a rather remarkable confirmation of the underlying basis of the theory.

If it is assumed that half the electrons that might normally be expected to escape are, for one reason or another, *reflected* by the work function barrier, the discrepancy between predicted and experimental values of A_0 is at once explainable. It is probably quite accidental that the *reflection coefficient* is just $\frac{1}{2}$. Most of the thought that has gone into this problem has been directed toward analyses of various possible reflecting mechanisms; a discussion of the possibilities will be found in Reimann's book.[T V] It is very difficult to explain the existence of values of A_0 substantially larger than 120, as exhibited very markedly by platinum; see Table III.

121. Distribution of Initial Velocities among Electrons Emitted from a Thermionic Surface. It is sometimes desirable to know how large a fraction of thermionically emitted electrons have initial x-directed velocities greater than any selected value. Fig. 94a is an integrated distribution diagram that conveys this information, and Fig. 94b is the corresponding distribution diagram. This distribution is predicted by theory and has been verified experimentally. The term *initial velocity* here signifies the velocity possessed by an electron after surmounting the gross work function barrier, but before it has been affected by whatever external field may exist due to space charge, interelectrode potentials, or both.

As applied to Fig. 78, initial velocity signifies an escaping electron's velocity at the point where the zero level of potential energy is reached or crossed, whether the potential-energy line is that corresponding to space-charge-limited (upper line) or temperature-limited (lower line) conditions.

Fig. 94 is a rather strikingly different *kind* of distribution diagram from those previously presented, for reasons which are not at once obvious. It is important to interpret each kind correctly. A satisfactory discussion of the contrasts between the two kinds can be built around the fact that Fig. 94, in addition to describing the x-directed velocity distribution of escaping electrons, *also gives the distribution of x-directed*

velocities among the particles that hit the inside wall of an enclosure containing a Maxwellian gas.   This statement is at first a little bothersome,
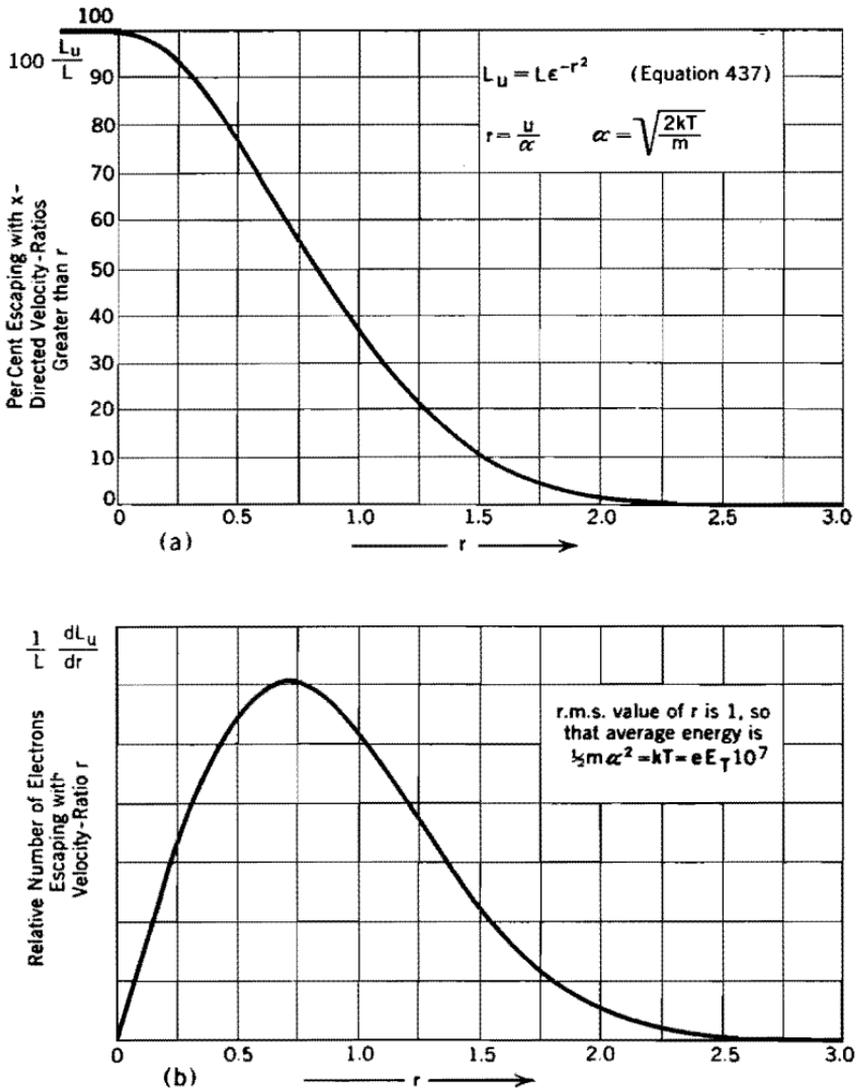
FIG. 94.   Time-exposure-over-a-surface velocity distribution of thermionically-emitted electrons, also of Maxwellian gas particles striking a boundary.

(a) Integrated distribution curve, Equation (437).
(b) Distribution curve, Equation (442).

because Fig. 94 is obviously not the same as Fig. 89; yet both purport to describe Maxwellian x-directed velocity distributions.   The two diagrams have different meanings, and must be differently interpreted.

All of the velocity distribution curves and equations presented *prior* to Fig. 94 have one characteristic in common: they describe in one way or another the range of individual velocities or energies as measured by a " *snap-shot observation throughout a volume.*" For example, all the integrated distribution curves *except* that in Fig. 94*a* specify percentages among particles that are observed to be present

(*a*) within a given volume
(*b*) at a given instant.

In sharp contrast to this, the curves in Fig. 94 describe the distribution as recorded by a " *time-exposure observation over a surface.*" That is, Fig. 94*a* specifies percentages among all the particles that are observed to reach or pass through

(*a*) a given area
(*b*) during any given time interval, say one second.

With this contrast in mind, the differences between Figs. 89 and 94 are easily explained, as follows:

(1) A particle cannot arrive at a surface while going away from it; therefore only positive velocities appear in Fig. 94.   In contrast with this, the curve in Fig. 89*b* is symmetrical about zero velocity, because a snapshot throughout a volume must record as many easterly as westerly motions.

(2) A particle cannot reach or cross any surface while standing still; therefore the curve in Fig. 94*b* has a zero ordinate at zero velocity.   Comparison shows that in Fig. 89*b* the maximum ordinate occurs at zero velocity; that is, a snapshot throughout a volume discovers more particles with zero than with any other *x*-directed velocity.

(3) During an appreciable time interval an individual particle may shuttle back and forth in an enclosure many times, hitting the surfaces on each round trip. Rapidly moving particles make more round trips per second, so hit more times per second, than slowly moving ones.   For that reason the average *x*-directed energy of the particles on arrival at a surface is greater than the average observed throughout a volume; it is, in fact, just twice as much in a Maxwellian distribution, as first stated in Section 72, and demonstrated in Section 122.

122. Equations and Averages for the " Time-Exposure-over-a-Surface " Velocity Distribution of Escaping Electrons.   If, as has been stated, the curves in Fig. 94 describe equally well the variations in velocities among:

(*A*) particles of ordinary gases impinging on an enclosing wall, and
(*B*) electrons escaping from a thermionic surface,

the equations for the curves must be derivable from expressions for both Maxwellian and degenerate velocity distributions.   Both derivations appear in the following paragraphs.

Fig. 94*a* describes how large a fraction of the recorded particles arrive or pass with *x*-directed velocity ratios exceeding any selected value *r*.

This fraction is just the ratio of L_u, Equation (425), to L, Equation (424); that is, it is

$$\frac{L_u}{L} = \frac{\text{Rate of arrival of particles with } x\text{-directed velocities greater than } u}{\text{Rate of arrival of particles with all velocities}}$$

(435)

The u's in both the equations mentioned can be converted into r's by the process used in Equation (426). The α's cancel in the resulting fraction, which has the form

$$\frac{L_u}{L} = \frac{\int_r^\infty r\, dN_u}{\int^\infty r\, dN_u}$$

(436)

This expression is applicable equally well to the (A) and (B) conditions to be treated; the difference between them is introduced in the choice of an expression for dN_u.

(A) In order to obtain the expression for the velocity distribution registered by Maxwellian particles impinging on an enclosing wall, Maxwellian x-directed distribution Equation (405) is used for dN_u. The integration is simple and leads to the following result:

$$\frac{L_u}{L} = \epsilon^{-r^2}$$

(437)

(B) In order to obtain this identical form as a description of the x-directed velocity distribution among the electrons escaping from a thermionic surface, it is necessary (1) to express, by means of a relation similar to Equation (436), the time-exposure-over-a-surface integrated distribution, *while still within the metal*, of the electrons *about to escape*, then (2) to express the result in terms of the much smaller velocities possessed after escape. The first operation employs high-velocity degenerate-gas distribution Equation (419) for dN_u; the constant factors cancel out, leaving

$$\frac{L_u}{L_G} = \frac{\int_r^\infty r\epsilon^{-r^2}\, dr}{\int_{r_G}^\infty r\epsilon^{-r^2}\, dr} = \epsilon^{-r^2 + r_G^2}$$

(438)

Here L_G signifies the rate of arrival, at the inside of the boundary, of particles for which $r > r_G$, enabling them to escape. This explains the use of r_G for the lower limit in the denominator.

In order to carry out the second operation let u', r', and E_u' describe the x-directed velocities, velocity ratios, and energies *after escape* of

particles for which u, r, and E_u describe the corresponding properties before escape. Since each electron during escape loses energy corresponding to the gross work function E_G,

$$E_u' = E_u - E_G \qquad (439)$$

so that

$$r'^2 = r^2 - r_G^2$$

or,

$$- r^2 + r_G^2 = - r'^2 \qquad (440)$$

In order to make the symbolism homogeneous, let L' stand for the rate of escape of particles with all velocities, that is, $L' \equiv L_G$. Also let L_u' describe the number emerging with x-directed velocities *after escape* exceeding u; L_u' is the same as L_u, for u and u' are the inside and outside velocities of the same particle. Using these facts and Equation (440), Equation (438) becomes, in terms of velocity ratios *after* escape:

$$\frac{L_u'}{L'} = \epsilon^{-r'^2} \qquad (441)$$

Equation (441) is identical with Equation (437), but refers to escaping electrons. This completes the (B) demonstration.

The equation for the curve in Fig. 94b is the negative derivative of Equation (437), that is,

$$\frac{d(L_u/L)}{dr} = 2r\epsilon^{-r^2} \qquad (442)$$

This distribution curve has the following interesting properties, determined in the usual way (see Section 112):

Average x-directed velocity: $\dfrac{\sqrt{\pi}\alpha}{2}$

Root-mean-square x-directed velocity: α

Average energy due to x-directed motion: E_T

Most probable (crest, Fig. 94b) x-directed velocity: $\dfrac{\alpha}{\sqrt{2}}$

The distribution specified by Equation (442) does not depend on the work function of the emitting surface. Therefore the electrons that succeed in surmounting an external potential-energy crest due to space charge, as in Fig. 78, must have just the same velocity distribution as those that merely reach the zero level, for the additional external obstruction has just the same kind of effect on the electrons that an increase in the work function would have. It is true that only those with originally greater energy reach the higher point; but, in so doing, they lose just enough energy to make the distribution among them the same

as formerly existed among the larger number that reached a point farther back.

123. Average Energies of Arriving Maxwellian Particles and of Escaping Electrons. Referring to Section 121, it is possible of course to imagine a time-exposure apparatus that records y-directed and z-directed velocities as well as those normal to the surface. But only x-directed movement brings particles to a y, z, surface, so that there is no tendency for particles with large sideways velocities to hit more often than those with small sideways velocities. Time-exposure observations over a y, z, surface must, therefore, record just the y-directed and z-directed velocity distributions that result from snapshot observations throughout a volume, Fig. 89. Only the x-directed distribution is recorded differently by the two methods of observation.

It was shown in the preceding section that the average energy of x-directed motion of Maxwellian particles on arrival at a surface is E_T. It was shown in Section 114 that the average snap-shot-throughout-a-volume energy of y-directed and z-directed motion among such particles is $\frac{1}{2}E_T$ for each of the two directions, in fact for all three directions. Therefore the average total energy of Maxwellian particles on arrival at a surface is $E_T + \frac{1}{2}E_T + \frac{1}{2}E_T = 2E_T$, as stated in Section 72.

It has now been shown that the average x-directed energy of *electrons emitted thermionically* is E_T, just as for Maxwellian gas particles striking a wall. The escaping electrons lose no sideways energy during escape; therefore their y-directed and z-directed average energies after escape are the same as while within the metal. These energies are, *for the group of high-velocity electrons able to escape,* just the same as for Maxwellian particles, $\frac{1}{2}E_T$ in each sideways direction, so that the total energies of thermionically emitted electrons should also average $E_T + \frac{1}{2}E_T + \frac{1}{2}E_T = 2E_T$. This expectation has been verified experimentally.

To explain why electrons about to escape have Maxwellian sideways energies it can be observed that their large x-directed velocity components make their total velocities large, so that the 1 in the denominator of Equation (412) becomes insignificant relative to the exponential term. Therefore η is proportional to $\epsilon^{-\frac{c^2}{\alpha^2}}$ for the portion of velocity-space that lies far out along the u axis. This is exactly the proportionality that appears in Maxwellian-distribution Equation (403) for the entire velocity-space; it is exactly this dependence that leads to the value $\frac{1}{2}E_T$ for sideways average energy. The proportionality factor is different in the two equations; but cancels out in the averaging process, which must be restricted to velocity points for which u^2 is somewhat larger than u_N^2.

124. Effect of Initial Velocities on Space-Charge-Limited Current Density, Plane Cathode. The derivation, in Sections 51 and 54, of the fundamental space-charge-limited current equations was based on the assumptions that the potential gradient into which the electrons emerge is zero just at the cathode surface, and that the electrons emerge with negligible initial velocities. Actually the point of zero potential gradient is at the bottom of a negative dip, illustrated in Fig. 77, and the electrons that succeed in passing the bottom of the dip do so with velocities distributed in accordance with Fig. 94 and Equations (437) and (442). It will be recalled that the effect of the dip on the electrons is the same as that of an addition to the work function, and that the velocity distribution of departing electrons is the same for all work functions.

The bottom of the negative dip may satisfactorily be thought of as a *virtual cathode*, located at a distance x_m from the true cathode, and with a potential E_m relative to the true cathode. See Curve (4), Fig. 77. x_m is always positive, E_m negative. Langmuir [74, 62] has given a complete mathematical solution for the potential distribution that exists between parallel planes when initial velocities give rise to such a virtual cathode. He also gives the following approximate equation for space-charge-limited current density, which has the special merit of indicating clearly the nature of the effects of the initial velocities:

$$J_p = \frac{\sqrt{2}}{9\pi} \sqrt{\frac{e}{m}} \frac{(E_s - E_m)^{\frac{3}{2}}}{(s - x_m)^2} \left[1 + 2.66 \left(\frac{E_T}{E_s - E_m} \right)^{\frac{1}{2}} \right] \quad (443 \text{ esu})$$

Or, in practical units:

$$J_p = 2.331 \times 10^{-1} \frac{(E_s - E_m)^{\frac{3}{2}}}{(s - x_m)^2} \left[1 + 2.66 \left(\frac{E_T}{E_s - E_m} \right)^{\frac{1}{2}} \right] \quad (444 \text{ p})$$

In these and the following equations J_p, J_{th}, and E_T have the same meanings as in Section 98; x and E are distance and potential measured from the cathode. Also:

$-F_0$ = space-charge-free potential gradient just off the cathode surface.

s = (1) in a two-electrode tube, spacing distance between cathode and plate, or,

(2) in a triode, spacing distance between cathode surface and intersection, at point S, of space-charge-free and space-charge-flow lines; see Figs. 44 and 45 and associated discussion.

$E_s = -F_0 s$ = potential corresponding to spacing distance s; E_s is the plate voltage for a diode, but simply $-F_0 s$ for a triode.

A comparison of Equations (443) and (444) with the earlier corresponding equations, (236) and (237), suggests the three distinct effects that initial electron velocities have on the space-charge-limited current density, all of them tending to make the actual current larger than that predicted by the earlier equations. The three effects are as follows:

(1) The virtual cathode is closer to the plate than is the true cathode; correspondingly $(s - x_m)^2$ in the denominator of Equation (444) is numerically less than s^2 that occurs similarly in Equation (237).

(2) The potential difference between virtual cathode and plate is greater than that between true cathode and plate; correspondingly $(E_s - E_m)^{\frac{3}{2}}$ in the numerator of Equation (444) is numerically greater than $E_p^{\frac{3}{2}}$ that appears similarly in the earlier equation, for E_m is negative as used here.

(3) The electrons that pass the virtual cathode do so with velocities whose average is proportional to α, so to $\sqrt{E_T}$. The expression for current density should therefore include an additive term containing the factor $\sqrt{E_T}$. Equations (443) and (444) do have such a term.

The form of Equations (443) and (444) shows that all three of these effects become progressively less important as the plate potential and spacing distance are increased relative to E_m and x_m.

Equations (443) and (444) are not satisfactory for numerical computations, because values of E_m and x_m must be determined before they can be used. Langmuir states the complete solution in terms of two very general nondimensional ratios, η (in no way related to η of the velocity-space, Chapter X) and ξ, called *eta* and *xi*, which are direct measures, respectively, of distance and of potential, *relative to values at the virtual cathode*. These two new quantities are related to familiar variables as follows:

$$\eta = \frac{E - E_m}{E_T} \tag{445}$$

$$\xi = \sqrt{J_p}\left[\frac{32\pi^3 m_e}{E_T^3 \, e}\right]^{\frac{1}{4}} (x - x_m) \tag{446 esu}$$

or, in practical units

$$\xi = 410.6\left[\frac{J_p^2}{E_T^3}\right]^{\frac{1}{4}} (x - x_m) \tag{447 p}$$

Equations (445), (446), and (447) show that the scale factors relating η to potential differences, and ξ to distance differences, are respectively $1/E_T$ and $410.6\,[J_p^2/E_T^3]^{\frac{1}{4}}$. See Fig. 95, discussed below.

The derivations of space-charge-limited current equations in Sections 51 and 54 make use of three basic relationships: the energy equation, Poisson's equation, and the equation of flow. Langmuir [62] uses

these, and in addition takes into account the initial velocities, using velocity distribution Equation (442). By a process in general similar to that used in Section 51, he then arrives at the function given in Equation (448) below, relating η and ξ. This relation is valid for any combination of values of cathode temperature, space-charge-free gradient, and spacing that produces a virtual cathode. The relation arrived at is

$$\xi = \int_0^\eta \frac{d\eta}{\sqrt{\epsilon^\eta - 1 \pm \epsilon^\eta \operatorname{erf} \sqrt{\eta} \mp \frac{2}{\sqrt{\pi}} \sqrt{\eta}}} \qquad (448)$$

The upper or lower signs are to be used here according as $(x - x_m)$ is negative or positive.

Numerical values, of ξ in terms of η, which satisfy this equation have been calculated by Langmuir and others and appear in Table IX. Equation (443) is obtained by rearranging a series expansion of Equation (448) after dropping out all but the first two terms.
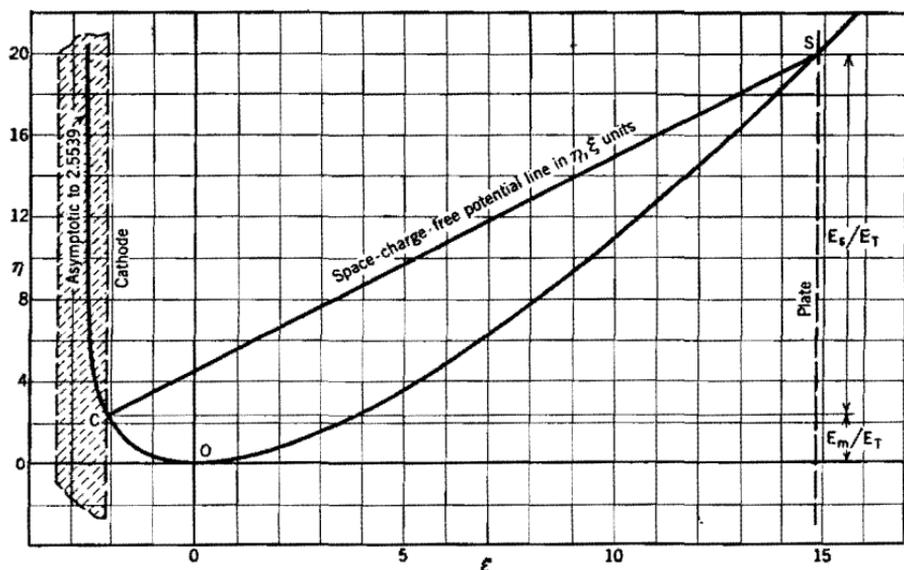
FIG. 95. Effect of initial electron velocities on the space-charge-limited potential distribution outside a plane cathode. See Equations (445) and (446). η is a measure of potential, ξ of distance.

Fig. 95 contains a curve plotted from Table IX. As η is proportional to potential and ξ to distance, this curve is essentially a potential distribution curve like Curve (4) in Fig. 77, referred to an origin at the virtual cathode. By proper location of a new origin to correspond

to the true cathode, and the use of horizontal and vertical scale factors, it can be used to describe the potential distribution that results from *any* combination of values of space-charge-free gradient, spacing distance, and cathode properties that results in the existence of a plane virtual cathode, providing the emitted electrons have velocities distributed according to Equation (442).

To illustrate this fact, there is shown in Fig. 95 a particular true cathode location and a straight line that corresponds to a particular space-charge-free potential line. This straight line intersects the ξ, η curve at C on the true (dotted) cathode; also at a point S which corresponds·to a particular plate location, or, more generally, to a particular intersection of the prolonged space-charge-free and space-charge-flow potential lines. The placement of C and S in Fig. 95 satisfies the following conditions:

$J_{th} = 0.16$, $J_p = 0.016$, amperes per square centimeter. (J_{th} is emission at C, J_p is plate current passing O, in Fig. 95.)

$s = 0.05$, $x_m = 0.0062$ centimeter.

$-F_0 = 73.0$ volts per centimeter.

$E_T = 0.2065$, corresponding to a temperature of 2400° K.

$E_s = -F_0 s = 3.65$ volts (plate voltage, if a diode).

$E_m = -0.48$ volt.

A solution for a specific problem in which J_{th}, J_p, E_T, and s are given, E_s to be found, is obtained as follows:

(1) Determine E_m from the following relation:

$$\frac{J_p}{J_{th}} = \epsilon^{\frac{E_m}{E_T}} \qquad (449)$$

This is in fact Equation (361), Section 98. Both here and there E_m is the potential difference between virtual and true cathodes; as used in Equation (361) its numerical value is positive, because its effect is like that of work function, which is customarily used as a positive numerical quantity. The present setup is such as to give E_m a negative numerical value, so that E_m here $= -E_m$ there.

(2) Use Equation (447) to find ξ at $x = 0$ and at $x = s$, called respectively ξ_c and ξ_s; these are the ξ coordinates of points C and S, Fig. 95.

(3) Read η_c and η_s from the curve of Fig. 95 at C and S.

(4) Find E_s from $\eta_s - \eta_c$, using Equation (445) and the facts that $E = 0$ when $\eta = \eta_c$ and $E = E_s$ when $\eta = \eta_s$.

If E_s is given, and J_p is not given, but rather to be found, trial values of J_p must be used as above until one is found which gives the correct value of E_s at S.

This method is valid for any set of values of the specified quantities for which the potential line has a minimum. $-F_0$ may even be negative, that is, the potential at S may be less than that at the cathode, and x_m

may be a large fraction of s, as in the case of a triode approaching cut-off. The true cathode and the virtual cathode coincide, and $x_m = E_m = 0$, when $J_p = J_{th}$.

Actual cathodes are only very rarely built with plane surfaces, so that the direct application of these curves and equations is somewhat limited. They are, however, helpful as an aid to understanding electronic behavior in geometries more often used but less easily analyzed than the parallel-plane type. They also illustrate how a fundamental relation between nondimensional ratios can be used for solving specific problems.

125. Effect of Initial Electron Velocities on Space–Charge–Limited Current Flow from a Cylindrical Cathode. No complete mathematical analysis, comparable with that used in the previous paragraph, has been made of the effect of initial velocities on the space–charge–limited current from a cylindrical cathode. Exact analysis is made difficult by the existence of nonradial motion due to sideways components of initial velocities, and the urge to obtain an exact solution is lessened by the fact that the initial velocities are known to be relatively less effective in cylindrical than in parallel-plane geometry.

As with a plane cathode, a virtual cathode exists a little outside a cylindrical cathode when the current is space-charge-limited. The three radii of consequence in the analysis of initial-velocity effects are:

r_c = true cathode radius

r_m = virtual cathode radius

r_s = radius to the plate, in a diode, or to the intersection point S (see Fig. 48) in a triode.

In the present problem the quantity β^2 that appears in Equation (257) and other similar ones depends on r_s/r_m rather than on r_s/r_c, but in the usual way. If r_s/r_m is 10 or larger, β^2 is very little affected by changes in the radius ratio and is not greatly different from 1; see Fig. 47. With very low space-charge-free off-cathode gradients, such as exist in a triode near cut-off, r_m may be quite large and β^2 quite small.

The electrons that pass the virtual cathode do so with average energies proportional to the temperature. The average energy due to initial *radial* motion is E_T; that due to initial *sideways* motion, perpendicular to a radius and to the axis, is $\frac{1}{2}E_T$. Both radial and sideways motion carry an electron nearer to a receiving electrode. Energy due to motion parallel to the axis has no effect on the current, because it does not affect the rate of travel toward an outer point. Hence, in his analysis of this problem, Langmuir considers only the contributions of the two kinds of motion first mentioned, totaling $\frac{3}{2}E_T$, as the average

contribution of initial velocities to the kinetic energy of outward motion.

The approximate cylindrical-cathode analysis given by Langmuir [74] starts with the assumption that an electron at radius r, potential E relative to the true cathode, has on the average a radial energy component of $E - E_m + \frac{2}{3}E_T$ electron volts. $\frac{2}{3}E_T$ of this is the initial energy contribution, arrived at as stated above. The remainder is the energy acquired from the field after passing the virtual cathode. This total average radial energy, rather than just E, is used in setting up the differential equation corresponding to Equation (251), Section 54. The result is

$$\frac{d^2E}{(d\log r)^2} = -2\pi J'\sqrt{\frac{2m_e}{e}}\frac{\epsilon^{\log r}}{\sqrt{E_s - E_m + \frac{2}{3}E_T}} \quad (450 \text{ esu})$$

An approximate solution for this equation, valid only when

$$\tfrac{2}{3}E_T \ll E_s - E_m$$

is then obtained in the general manner outlined in Section 54. The result is given as

$$J'_p = \frac{2\sqrt{2}}{9}\sqrt{\frac{e}{m_e}}\frac{1}{r_s\beta^2}\left[E_s - E_m + \tfrac{3}{8}E_T\left(\log\frac{2E_s}{3\lambda E_T}\right)^2\right]^{\frac{3}{2}} \quad (451 \text{ esu})$$

In practical units, the assembly of numbers and universal constants in this equation becomes $2\pi \times 2.331 \times 10^{-6}$, as in Equation (249).

The symbolism used here is the same as in Section 54, with E_T and E_m added. λ is a numerical constant whose value probably lies between 1 and 2 and cannot be definitely determined except by experiment. This treatment is not valid near cut-off in a triode, for then E_s is very small, and in some cases negative.

Comparison of the earlier Equation (257) with Equation (451) indicates that the latter, like Equation (443), takes the three modifying effects of initial velocities into account separately, as follows:

(1) The effect of the outward shift of the cathode from the true to the virtual position appears in the value of β^2. This is more a formal than a real accounting, because the condition that $\frac{2}{3}E_T$ must be small relative to $E_s - E_m$ is usually satisfied only when r_s is so large relative to both r_c and r_m that β^2 is nearly constant at a value near 1.

(2) The use of $E_s - E_m$ rather than E_p takes account of the dip to virtual cathode potential.

(3) The term containing E_T accounts for the effects of the random energies possessed by the electrons as they pass the virtual cathode.

Even without knowing definitely the proper value to use for λ, it is possible to estimate how much larger the current specified by Equation (451) is than that given by Equation (257), which neglects initial elec-

tron velocities. Langmuir's calculations show that the per cent increase is in general only about one-fourth or one-fifth of that in the corresponding plane cathode comparison.

In contrast with Equation (443), the comparable approximate plane cathode expression, Equation (451) *can* be used rather easily for numerical calculations, although a trial and error process is necessary. Suppose that J'_{th}, E_T, E_s, r_s, and r_c are all given, and that 1 is to be used for β^2 and for λ. Equation (449) takes the form

$$\frac{J'_p}{J'_{th}} = \epsilon^{\frac{E_m}{E_T}} \tag{452}$$

and permits determination of a value of E_m (negative) corresponding to an arbitrarily chosen preliminary value of J'_p. This value of E_m used in Equation (451) either will or will not give the assumed value of J'_p. If it does, the preliminary choice is the correct one; if not, another trial must be made. This simple handling is made possible by the fact that the radius of the virtual cathode does not appear in Equation (451).

126. Shot Effect and Voltages within Conductors Due to Random Motions; Noise Level. The term " noise level " refers basically to the intensity of miscellaneous noises that speech or music must rise above in order to be clearly audible. In a busy street, the noise level has to do with the amount and kind of traffic passing; in a telephone receiver or radio speaker, it depends on small current variations which result from various properties of electric circuit elements. The word " noise " suggests variations within the frequency range audible to human ears, but the phrase " noise level " is in fact used to describe the intensity of miscellaneous troublesome variations of currents at any or all frequencies.

Some causes of noise are fairly obvious. The sensitive element in many microphones and most telephone transmitters is a pair of diaphragms separated by closely packed carbon granules. Movement of the diaphragms in response to air vibrations of music or speech changes the current in the circuit by changing the resistance of contact points between the granules. There is always some agitation among the granules, with resulting noise, even when the diaphragm movement is vanishingly small. Other causes of noise are poor connections, run-down batteries, and vibration of electrodes within tubes. Radio " static " is fundamentally a high-frequency noise. It is possible, though sometimes difficult, to devise circuits in which these *obvious* noise effects are eliminated.

There are other sources of noise which are present in any circuit containing a vacuum-tube amplifier, no matter how much care is taken in the circuit assembly, because they result indirectly from the random

motions of the current-carrying particles. The two important causes of noise that come under this heading are the raindrop patter of electrons on the plate, called " shot effect " from its similarity to the effect of a shower of small shot, and the random deviations from uniformity of the distribution *in space* of the electrons within metal conductors, usually referred to by the use of the phrase *thermal agitation.*[83, I 208]

The velocity distributions of gas particles discussed in the previous chapter are merely *most probable* distributions, as explained in Section 109. The actual velocity distribution at any moment is invariably near to the most probable; but the smaller the volume selected, the greater is the chance of discovering within it appreciable momentary deviation from the most probable.

The most probable *space* distribution of electrons throughout the interior of a metal is of course such as to make the electron concentration everywhere the same, but again this is only a *most probable* space distribution. A momentary observation is likely to discover appreciable piling up of electrons here and there, with resulting variation, according to Poisson's law, of potential from point to point within the metal. This occurs in the conductors that make up the connection between cathode surface and grid of a triode (this connection includes the hot filament wire itself) and results in small, rapid, random changes in grid potential. These changes are amplified by the tube and appear as noise in the plate circuit. It is hopeless to attempt to amplify signal voltages no larger than these grid-circuit *thermal agitation voltages,* for the noise is amplified along with the signal voltage; the two are indistinguishable.

In a heated cathode, variations of the type just described in the space distribution of electrons, combine with similar random variations in the velocity distribution to cause random changes in the rate of arrival of high-energy electrons at the work-function barrier. As a result there are small random variations from point to point and from moment to moment in the thermionic current density. The higher the temperature, the more pronounced are the variations. If the current in a tube is temperature-limited, this results in a variation of the rate at which electrons rain on the plate, and noise in the plate circuit results. This kind of noise is said to be due to *shot effect.*

If the current is space-charge-limited, the random variations in thermionic current density cause compensating variations in the magnitude of the negative potential dip just outside the cathode, the changes in plate current being insignificant. Thus shot effect is important only in circuits containing tubes whose current is wholly or partially temperature-limited, so is of no consequence in most amplifier circuits.

The effects of these various random variations of voltage and current

in resonant circuits depend on the frequencies of the random variations. Their frequency distribution can be predicted from the velocity distribution laws, and checked experimentally; this fact has permitted experimental determination of the magnitude of the electronic charge from measurements of shot effect.

127. Random Current Density in an Ion or Electron Gas. There is a marked similarity between the mechanism of electric conduction in a metal and that in most conducting gases, for in both the current is carried primarily by the migration, or " drift " of an electron gas.

In both metals and conducting gases the space charge that might be expected to result from the presence of the electrons is entirely neutralized by an equal concentration of positive ions. The ions are of course atoms that have each contributed an electron to the makeup of the electron gas. The essential contrast between metallic and gaseous conduction is that, in a metal, the positive ions are fixed in position, but in a conducting gas they are free to move. Because of their large mass, they move so slowly that the ion drift motion makes a negligible contribution to the current flow. Their outward diffusion owing to *random* motions is important, however, because it ultimately removes them from the discharge. Continuous energy input is necessary to replace them.

This explains in a general way the interest in what is called the *random ion current density* in a conducting gas. Section 119 discusses the rate of arrival of gas particles at a square centimeter of a boundary wall. However, the entire story told there applies equally well to a square centimeter of imaginary area placed at any location within the body of the gas. Equation (424) can, therefore, be used to describe the rate at which gas particles pass through from one side to the other of any such *internal* square centimeter of area. If the gas particles are ions, they carry a current, called the *random ion current*, through the chosen area. Of course the random ion current in one direction across any internal area is the same as that in the reverse direction, so that the concept of random ion current flow does not imply a net transfer of charge from one part of the gas to another.

The magnitude of the *random ion current density* in an ordinary conducting gas is obtained by using, in Equation (424), the expression for dN_u in a Maxwellian gas, and multiplying it by e, the charge carried by each particle, as follows:

$$J_{ri} = e\alpha_i \int_{r=0}^{r=\infty} r \, dN_u = \frac{N_i e \alpha_i}{\sqrt{\pi}} \int_0^\infty r \epsilon^{-r^2} dr \qquad (453)$$

This integrates into

$$J_{ri} = \frac{N_i e \alpha_i}{2\sqrt{\pi}} \qquad (454)$$

Here J_{ri} = random ion current density
            N_i = ion concentration
            α_i = the velocity that is characteristic of the temperature of the ionic gas

An exactly similar expression for *random electron current density* J_{re}, the i subscripts being changed to e subscripts throughout, with obvious significance, is as follows:

$$J_{re} = \frac{N_e e \alpha_e}{2\sqrt{\pi}} \tag{455}$$

Even though $N_e = N_i$ in a conducting gas, the random electron current density is ordinarily hundreds of times larger than the random ion current density. This results primarily from the fact that the random currents are proportional to the α's, which vary inversely as the square roots of the masses, and ions have *very* much larger masses than electrons. It is the familiar story that large bodies move slowly. In addition, it is found by experiment that the electron temperature is often several times that of the ions and neutral gas, and this of course accentuates the contrast between the magnitudes of the random current densities.

128. Boundary Currents in a Conducting Gas; Sheath Penetration. Electrical discharges often have material boundaries, such as the walls of enclosing glass or metal envelopes. The random ion and electron current densities describe not only the rates at which ions and electrons pass back and forth across internal areas; they also describe the rates at which random motions make ions and electrons approach the boundaries.

The relation of boundary currents to discharge properties is an important part of the discussion of electrical discharges in general, and will be treated in detail in Chapter XIX.

For reasons discussed there, a "potential barrier" often exists between the main body of a conducting gas and a boundary wall. This barrier obstructs the flow of electrons to the wall. The term "barrier" is used here merely to remind the reader of the work-function barrier that similarly obstructs the escape of electrons from a metal. The barrier now under discussion is simply a drop in potential that occurs within a thin region, often called a "sheath," adjacent to the wall; see Fig. 96, page 252. The potential drop through the sheath is intimately associated, by way of Poisson's law, with the positive space charge that exists within it because of the absence of electrons.

A few high-random-energy electrons penetrate through the sheath from plasma to wall in spite of the barrier it offers, in very much the

way that a few electrons escape from a hot metal in spite of the work-function barrier. The determination of an expression for the electron current density that passes through the sheath in spite of the obstruction offered is arrived at by means of Equation (427), just as in the thermionic emission problem. Of course in the present problem dN_u has the Maxwellian form. The derivation is as follows:

$$J_{se} = e\alpha_e \int_{r=\frac{u_s}{\alpha_e}}^{r=\infty} r\, dN_u = \frac{N_e e\alpha_e}{\sqrt{\pi}} \int_{\frac{u_s}{\alpha_e}}^{\infty} r\epsilon^{-r^2}\, dr \qquad (456)$$

Equation (456) integrates into

$$J_{se} = \frac{N_e e\alpha_e}{2\sqrt{\pi}}\, \epsilon^{-\frac{u_s^2}{\alpha^2}} \qquad (457)$$

also written

$$J_{se} = J_{re}\, \epsilon^{-\frac{E_s}{E_{eT}}} \qquad (458)$$

The symbolism here is the same as in Equation (455), with the following additions:

J_{se} = current density due to passage of electrons *through* the sheath.

E_s = potential difference that these electrons must overcome in penetrating the sheath.

E_{eT} = voltage equivalent of electron temperature.

u_s = x-directed velocity necessary for penetration.

129. " Time Exposure-over-a-Surface " Distribution of Penetrating Electrons. The " time-exposure-over-a-surface " velocity distribution (see Section 121) among Maxwellian electrons that succeed in penetrating a sheath is the same as that among the much larger number that first enter the sheath from the conducting gas, and is described by Fig. 94 and Equations (437) and (442). Consequently the penetrating electrons reach the wall with average energies $\frac{1}{2}E_{eT} + \frac{1}{2}E_{eT} + E_{eT} = 2E_{eT}$ total.

FIG. 96. Sheath adjacent to the boundary wall of a conducting gas.

That this is true can be demonstrated in a way almost identical with that used in Section 122 to derive the velocity distribution of electrons escaping from a thermionic emitter. The two steps, as under heading (B) in that Section, are (1) to express the time-exposure-over-a-surface integrated distribution, *before entering the sheath*, of the electrons *about to pass through it*, then (2) to express the

result in terms of the much smaller velocities possessed on arrival at the boundary wall. The first operation employs Maxwellian distribution Equation (405) for dN_u; the constant factors cancel out, leaving

$$\frac{L_u}{L_s} = \frac{\int_r^\infty r\epsilon^{-r^2}\,dr}{\int_{r=r_s}^{r=\infty} r\epsilon^{-r^2}\,dr} = \epsilon^{-r^2+r_s^2} \tag{459}$$

Here L_s signifies, at the conducting-gas side of the sheath, the rate of arrival of electrons for which $r > r_s$, where

$$r_s = \frac{u_s}{\alpha_e} = \sqrt{\frac{E_s}{E_e T}} \tag{460}$$

From this point on, the derivation parallels that in Section 122 exactly, and leads to an expression for rates of arrival at the wall that is identical with Equation (441). Thus it is demonstrated that Fig. 94 describes the x-directed velocity distribution among penetrating as well as among entering electrons. The distributions in the other two directions are of course unchanged by passage through the sheath.

130. Richardson's Equation for Thermionic Emission. Early attempts were made to explain thermionic emission from metals by assuming an internal electron gas with a Maxwellian distribution. Such treatment leads to an expression for thermionic current density exactly like Equation (458). Work function E_W is used in place of E_s, and the distinction between net and gross work function vanishes; the E_T of the emitter is used in place of $E_e T$. Since α is proportional to the square root of the temperature, the following general form, widely used and called " Richardson's equation " results.[T 22]

$$J_{th} = a T^{\frac{1}{2}} \epsilon^{-\frac{E_W}{E_T}} \tag{461}$$

Equation (309) is a more usual form of Equation (461). A brief comparative discussion of Richardson's and Dushman's equations is given in connection with Equation (309).

131. Equilibrium between Different Potentials in an Enclosure; the Boltzmann Relation. Fig. 97 represents an enclosure containing an electron gas, its space charge being everywhere just neutralized by equal numbers of relatively stationary positive ions. The dotted vertical lines in the middle indicate that the enclosure is screened into two compartments differing in potential by E_s volts, as suggested by the potential lines at the bottom of the figure. Electrons are able to pass freely through the screens. The potential within each compartment is uniform,

the potential change taking place entirely within the intermediate boundary zone. Situations similar to this occur in some gaseous-conducting electronic devices.

Random electron velocities must in general result in a flow of electrons both ways through the boundary zone, though the flow from left to right is affected by the potential difference between the compartments. Since any electron as it leaves or enters a compartment has some kinetic energy of motion, the electron flow results in a two-way energy transfer. The rates of flow of particles, also the rates of transfer of energy, are equal in the two directions when the following equilibrium conditions exist:

Fig. 97. Equilibrium between electron gases at different potentials.

(1) $E_{T_1} = E_{T_2}$ (that is, the electron temperatures are equal) (462)

(2) $\dfrac{N_1}{N_2} = \epsilon^{-\frac{E_s}{E_T}}$ (463)

where N_1 and N_2 are the respective electron concentrations.

This illustrates what is rather widely known as the "Boltzmann relation." The Boltzmann relation in general specifies just such an exponential relation, involving the temperature, for the *equilibrium* distribution of particles among various available energy states, which may be of a great variety of types.

Equation (405) can be used to illustrate the application of the Boltzmann relation to the distribution of x-directed gas particle energies. Consider two increments of x-directed velocities with equal dr's, but widely differing x-directed energies E_u' and E_u'', as follows:

$$dN', dr \quad \text{at} \quad r' = \sqrt{\frac{E_u'}{E_T}} \tag{464}$$

$$dN'', dr \quad \text{at} \quad r'' = \sqrt{\frac{E_u''}{E_T}} \tag{465}$$

Divide the equation using the first set by that containing the second; the result is

$$\frac{dN'}{dN''} = \epsilon^{-\frac{E'-E''}{E_T}} = \epsilon^{-\frac{\Delta E}{E_T}} \tag{466}$$

which is identical in nature with Equation (463).

No general demonstration of the Boltzmann equilibrium relation will be given here, but the derivation of Equation (463) will be presented, for it is simple and direct.* Let it be assumed that equilibrium of electron flow and energy transfer exists between the two compartments of Fig. 97. Equality of the electron temperatures in the two regions has not yet been demonstrated, so distinct symbols E_{T_1} and E_{T_2}, α_1 and α_2, will be used for characteristic electron temperature energies and velocities in the two compartments. The current due to electron flow across screen AA from left to right is simply the random current density for the compartment 1. That in the reverse direction, corresponding to the rate at which electrons penetrate a potential barrier, is given by an expression like Equation (458), using the concentrations and temperatures for compartment 2. The requirement that these two flow rates must be equal is stated as follows:

$$\frac{N_1\alpha_1}{2\sqrt{\pi}} = \frac{N_2\alpha_2}{2\sqrt{\pi}}\epsilon^{-\frac{E_s}{E_T}} \qquad (467)$$

Electrons passing from left to right through screen AA have time-exposure-over-a-surface total average energies $2E_{T_1}$; those passing through it in the reverse direction have similar total average energies $2E_{T_2}$, for they originate in a region of characteristic temperature energy E_{T_2}. It was demonstrated in Section 129 that the velocity and energy distributions and average energy among those that penetrate are the same as exist among the larger number that arrived at BB. Hence the equality of energy transfer is expressed by multiplying the rate of electron movement each way by the average energy carried, as follows:

$$\frac{N_1\alpha_1}{2\sqrt{\pi}}2E_{T_1} = \frac{N_2\alpha_2}{2\sqrt{\pi}}\epsilon^{-\frac{E_s}{E_T}}2E_{T_2} \qquad (468)$$

If Equation (468) is divided by Equation (467) it appears that

$$E_{T_1} = E_{T_2} \qquad (469)$$

That is, in the equilibrium condition the temperatures must be equal. Using this fact in Equation (467) the α's cancel, leaving just Equation (463).

The differential form of Equation (463) is of interest. Suppose that N_1 and N_2 differ by only a small amount dN, and that E_s becomes a differential increment dE. Equation (463) can then be

---

The derivation given here parallels one given by Loeb [V 109] relative to the mechanism of evaporation and condensation at a liquid surface.

written

$$\frac{N + dN}{N} = \frac{\epsilon^{-\frac{E+dE}{E_T}}}{\epsilon^{-\frac{E}{E_T}}} = \epsilon^{-\frac{dE}{E_T}} \tag{470}$$

Since dE/E_T is very small, a series expansion of the exponential reduces this to

$$1 + \frac{dN}{N} = 1 - \frac{dE}{E_T} \tag{471}$$

or

$$\frac{dN}{N} = -\frac{dE}{E_T} \tag{472}$$

132. Free Paths of Gas Particles. The velocities of drift of electrons and ions in response to electric fields, and the rates of production of ions by collision between electrically accelerated and stationary particles, depend largely on the distances traveled by particles between collisions. These are called "free paths." [W II, V III] The average free path is usually called the *mean free path*, symbol *l*. Fig. 99, page 259, illustrates the distribution of free paths above and below the mean.

The most common direct use of the mean free path in the study of gaseous discharges is to determine the number of collisions experienced per second by a particle. If the average velocity is known, the number of centimeters traveled in *one* second is also known. This divided by the mean free path between collisions gives the number of collisions per second.

Approximate mathematical expressions for the free path distribution and for the mean free path are derived in the following paragraphs. The method used is chosen partly because it is simple, partly because its order is logical, but chiefly because the reasoning process employed appears again in discussions of backfire frequency in mercury arc rectifiers and in other similar problems of arc and spark initiation. In those later problems, time instead of distance becomes the important variable, but the mathematical and logical processes are unchanged.

The method involves probabilities or " chances," and can be illustrated by the following example. Suppose it is required to estimate the chance that a football player will gain between fifteen and sixteen yards on a play. In order to make this gain he must first succeed in advancing fifteen yards, then he must be stopped within the next one yard. Two distinct chances are involved: first, that of advancing any amount greater than fifteen yards; second, that of being stopped within one yard.

Suppose that in the game in question, it is reasonable to expect him to

gain fifteen yards or more in 5 per cent of the plays; that is, the probability of his advancing fifteen or more yards is 0.05. Suppose also that he has a 20 per cent chance of being downed in any one yard that he reaches; that is, the probability of his being stopped in any one yard is 0.20. Then the probability of going more than fifteen yards, then being stopped within the sixteenth yard, is the first probability multiplied by the second, that is

$$0.05 \times 0.2] = 0.01 \qquad (473)$$

This set of events should therefore be expected to occur in about one per cent of the plays. The chance of his going *more* than sixteen yards is that of going more than fifteen, minus that of stopping within the sixteenth yard, or $0.05 - 0.01 = 0.04$.

Now to return to the gas particles: let the solid-line, crosshatched circle at A, Fig. 98, represent a gas particle at a moment when it has already advanced x centimeters from the point of the last collision, and the dotted-line circle at B the same particle after it has gone dx centimeters farther. The increment dx corresponds to the one yard between fifteen and sixteen in the football example. Presume that all gas particles are spheres with radius b centimeters. If the *center* of any other approaches closer than $2b$ to the center of the one illustrated, collision takes place. Spheres of radius $2b$ are drawn at both positions in the figure to indicate the limit of approach of other noncolliding particles.
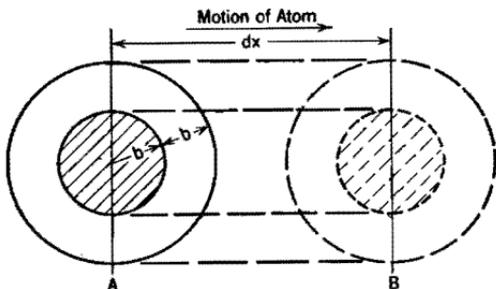
Fig. 98. Increment dx of an atom's free path. Volume swept out in distance dx by imaginary sphere of radius $2b$ is $\pi(2b)^2 dx$. Chance of collision therein is $N\pi(2b)^2 dx$.

Let P_x symbolize the probability that a gas particle will go a distance greater than x without collision; that is, it is the ordinate of the integrated distribution curve, Fig. 99a. When the football player's x was fifteen yards, his P_x was 5 per cent.

Let $p\,dx$ symbolize the probability of a collision terminating the free path within the distance dx; thus $p\,dx$ corresponds to the football player's 20 per cent chance of stopping in any one yard.

The *change* dP_x in P_x while dx is being traversed is of course the probability of reaching and passing x without collision, but of experiencing a collision before $x + dx$ is reached. Therefore dP_x is the product of (1), the probability of advancing a distance greater than x, by (2), that of stopping within dx. This product is negative, because P_x

decreases as x increases.   Mathematically,

$$dP_x = -P_x \cdot p\,dx \qquad (474)$$

Equation (474) can be written

$$\frac{dP_x}{P_x} = -p\,dx \qquad (475)$$

and integrated into

$$\log P_x = -px + K \qquad (476)$$

or

$$P_x = \epsilon^{-px+K} \qquad (477)$$

The probability of a free path terminating at a distance greater than zero is 100 per cent, so that $P_x = 1$ when $x = 0$.   Hence the integration constant K vanishes, giving

$$P_x = \epsilon^{-px} \qquad (478)$$

The quantity p is determinable in the gas particle case from the fact that $p\,dx$ must be the probability of collision with another gas particle within the distance dx.   While going from A to B, Fig. 98, the sphere of radius $2b$ sweeps out a volume $\pi(2b)^2\,dx$.   The chance of collision meanwhile is that of finding the center of another particle within this volume, which is the volume itself multiplied by the concentration N of particles.   Therefore

$$p\,dx = N \cdot \pi(2b)^2\,dx \qquad (479)$$

so that

$$p = 4\pi Nb^2 \qquad (480)$$

Using this in Equation (478)

$$P_x = \epsilon^{-4\pi Nb^2 x} \qquad (481)$$

Let a distance l be defined as follows:

$$\frac{1}{4\pi Nb^2} = l \qquad (482)$$

Equation (481) can then be written

$$P_x = \epsilon^{-\frac{l}{x}} \qquad (483)$$

This is in fact the equation for the integrated distribution curve of Fig. 99a.[W 68, V 45]   Its negative derivative is the equation for the distribution curve, Fig. 99b, that is

$$\frac{dP_x}{dx} = \frac{1}{l}\epsilon^{-\frac{x}{l}} \qquad (484)$$

For this particular distribution the two curves have identical shapes.

The average or mean free path is obtained by averaging the distribution curve, Equation (484), in the usual way, as described in Section 112; the result is just l, so that

$$\text{Mean Free Path of Gas Particles} = l = \frac{1}{4\pi Nb^2} \tag{485}$$

Equation (485) for the mean free path of gas particles is only approximately correct, because in the derivation it has been assumed that all particles other than the one illustrated are stationary; whereas in fact they have motions distributed according to the Maxwellian law. These motions may make some particles initially in the volume $4\pi b^2 dx$ move
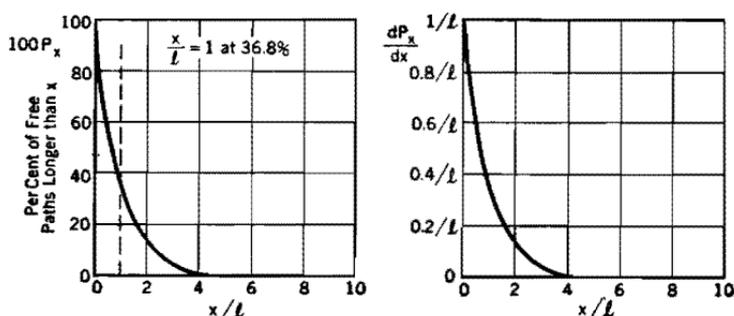
(a) Integrated distribution

$$P_x = \epsilon^{-\frac{x}{l}}$$

$$\frac{x}{l} = \frac{\text{Individual free path}}{\text{Mean free path}}$$

(b) Distribution

$$\frac{dP_x}{dx} = \frac{1}{l}\epsilon^{-\frac{x}{l}}$$

See Equations (485), (486), (487) for l.

FIG. 99.   Distribution of free paths of gas particles.

out in time to escape being hit and make others move in from the outside in time to be hit.   Careful analysis shows that the chance of moving in is greater than that of moving out, so that the actual probability of collision in the distance dx is greater than that given in Equation (479), by the factor $\sqrt{2}$.   Thus p becomes $4\sqrt{2}\pi Nb^2$; the mean free path, called *Maxwell's* mean free path, becomes[V 97]

$$\begin{array}{l}\text{Mean free path of gas particles}\\ \text{taking into account random velocities}\end{array} = \frac{1}{4\sqrt{2}\,\pi Nb^2} \tag{486}$$

An expression for the mean free path of an electron moving among gas molecules and ions is obtainable by recognizing that the space occupied by an electron is negligible relative to that occupied by an atom, so that the crosshatched circles in Fig. 98 shrink to points representing the electron. b is still used for the gas particle radius.   Since collision occurs if an electron approaches within a distance b of the center of a

gas particle, the larger sphere that describes the limit of approach without collision must have a radius b rather than $2b$. The volume swept out during the electron's advance from A to B, Fig. 98, becomes $\pi b^2 \, dx$. This is related as before to the mean free path, so that[V 51]

$$\text{Mean free path of an electron moving among gas particles} = \frac{1}{\pi b^2 N} \qquad (487)$$

No account need be taken of the random velocities in connection with the electronic mean free path. All the electron's collisions are with gas particles or ions, whose random velocities are small relative to those of the electrons because of the large mass ratio. For similar reasons Equation (485) should be expected to apply in the case of an ion's being driven by an electric field at a velocity much higher than the average random velocities of the gas particles among which it moves.

Experiments have shown that electron mean free paths are in fact considerably dependent on electron energy. In particular, high-velocity electrons often appear to pass completely through atoms without appreciable deflection. Table XV contains information from which the mean free paths of electrons of various velocities in various gases can be calculated.

PROBLEMS

Chapter XI

1. The supposed raindrop velocity distribution equation given in Problem 13, Chapter X, describes of course a " snap-shot-throughout-a volume " distribution. Using that equation,

(a) Derive an expression for the number of raindrops striking one square foot of ground per second.

(b) Derive expressions for the time-exposure-over-a-surface velocity distribution and the integrated velocity distribution among the raindrops as they strike the ground.

(c) Determine the average velocity among the arriving drops.

(d) Determine the average energy among the arriving drops, if each one's mass is m grams.

(e) How large a fraction of the drops strike the ground with velocities greater than $a + b$? With velocities less than $a - b$?

2. In Section 124 there will be found a set of numerical values of J_{th}, J_p, s, F_0, etc., that are consistent with the locations of points C and S in Fig. 95. For the set of conditions so described, draw carefully a *potential energy* diagram that shows the relations of the various energy levels within the cathode and plate to each other. (See Fig. 78 and associated discussion.) The following values apply, in electron volts:

	Cathode	Plate
Work function, E_W	4.5	3.5
Normal maximum energy, E_N	8.0	6.0

3. Suppose that for a particular set of conditions the space-charge-free potential line on a ξ, η curve like that of Fig. 95 extends from a point C at $\eta = 4$ to a point S at $\xi = 10$.

(a) Select a set of values of J_{th}, T, s, and E_s that will fit such a line.

(b) Select another set having twice as large a spacing between electrodes as the value selected in (a).

4. A plane tungsten cathode, temperature 2500° K, lies 0.1 cm from a plane plate surface. Using Fig. 95 or Table IX, or both, find the potential the plate must have in order to make the actual current (a) 1 per cent, (b) 10 per cent, (c) 100 per cent, of the temperature-limited value. For J_{th}, use 0.27 amperes per sq cm. Determine E_m and x_m in each case.

5. Derive Equation (448) from fundamental relations.

6. Solve part b of Problem 2, Chapter IX, by the method outlined in connection with Equations (451) and (452), using $\lambda = 1$, then using $\lambda = 2$. Comparison of these results with the solution obtained by the methods of earlier chapters illustrates quantitatively the effect of initial electron velocities in a cylindrical arrangement.

7. In an electron gas containing 10^{12} electrons per cc at a temperature of 20,000° K,

(a) What is the voltage equivalent of this electron temperature?

(b) What is the most probable electron velocity?

(c) How many electrons having x-directed energies exceeding 10 electron volts will strike each square centimeter of the enclosing boundary per second?

(d) What will be the average x-directed energy of these electrons?

(e) What will be their average total energy?

(f) Answer questions (d) and (e) for the same group of electrons *after penetration through a ten-volt barrier encountered at the surface.*

8. It is conceivable that the work function of a surface might depend on the temperature. Find what function E_W would have to be of E_T and a constant E'_W in order to convert Dushman's equation for thermionic current density into the following form:

$$J_{th} = A'_0 \, T_\epsilon^{-\frac{E'_W}{E_T}}$$

9. Two parallel plane hot surfaces are both emitting at the rate of 0.160 amperes per sq cm into the region between them. Both are at a temperature of 2400° K, and are at the same potential. There is of course no net current flow to or from either plate.

(a) Find the electron concentration immediately adjacent to each plate, using the facts that the electrons there have a Maxwellian distribution of velocities, and that the rate of electron reentrance must equal the electron emission.

(b) If the potential midway between the plates is 2 volts below that of the plates, what is the electron concentration at the midpoint (the Boltzmann relation applies here).

(c) Derive an equation for the potential distribution curve between the electrodes.

10. A gaseous discharge ordinarily contains electrons, ions, and neutral gas particles. Suppose that in a certain mercury-vapor arc there are 10^{12} electrons per cc, the electron temperature being 30,000° K. There are the same number of ions as electrons per cubic centimeter. The temperature of the mercury atoms and ions is 1000° K. There are 4×10^{13} mercury atoms per cc.

(a) Find the random ion current density, and the random electron current density.

(b) Find the mean free path for electrons having the average energy.   (See Table XV.)

(c) Find the mean free path for electrons having 10.4 electron volts of energy (this is just enough energy to ionize a mercury atom).

11. (a)   What must be the concentration of argon gas in a cathode-ray tube in order to permit 80 per cent of the electrons to travel from the electron gun to a fluorescent screen 50 cm away without hitting an argon atom?   The gun delivers 100-volt electrons.   (See Section 132 and Table XV.)

(b)   Write an equation for the number of argon ions produced per centimeter of path per second, at distance z along the path, assuming that one-fifth of all the collisions of electrons with argon atoms produce argon ions.   The beam represents a current of 20 microamperes.

CHAPTER XII

AMPLIFIER CIRCUIT PRINCIPLES*

133. Plate Resistance and Grid-Plate Transconductance of High-Vacuum Thermionic Tubes. The current-voltage relationships for high-vacuum thermionic tubes are easily measurable and reproducible. They can be satisfactorily represented by families of characteristic curves of the nature discussed in Chapter VI. The most important families of curves are sets of *mutual characteristics* and *plate characteristics*, illustrated respectively by Figs. 50a and 50b, page 123, for an RCA Type 6C5 metal-clad triode.

Any thermionic vacuum tube is an example of a *nonlinear* circuit element. An ordinary resistance which follows Ohm's law is called a *linear* circuit element, because the current is directly proportional to the voltage. Fig. 100 contrasts the current-voltage relationship for an ordinary linear resistance (the straight line through the origin) with that for an illustrative *nonlinear* circuit element.[E, XXI]

The term " resistance," as applied to a nonlinear circuit element, may have a variety of meanings. E/I (volts divided by amperes) may have one value,
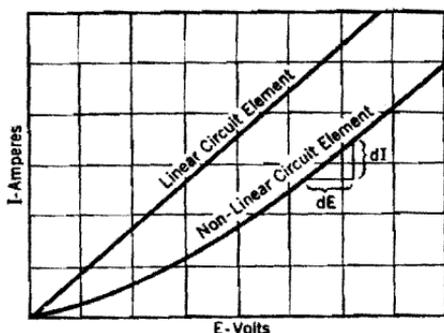
FIG. 100. Contrast between linear and non-linear circuit elements.

watts $/I^2$ another, and dE/dI still another, although in an ordinary resistance all three have the same value. In many nonlinear circuit elements, and especially in vacuum tubes, the currents and voltages are usually *pulsating*, consisting of alternating-current components superposed on direct-current ones. When the alternating components of voltage and current are relatively small, the ratio between them is dE/dI, which can therefore be called the alternating-current resistance of the element.

In a triode, tetrode, pentode, etc., the alternating-current resistance

---

* For collateral reading in connection with this and the next two chapters, the reader is referred to books by Chaffee,[E] Glasgow,[F] McIlwain and Brainerd,[G] Everitt,[H] Terman,[I] Peters,[AA] and to Radio Handbooks. [BB, CC]

to flow of plate current, relating the alternating components of plate current and plate voltage, is called the *plate resistance, R_p*. It is defined mathematically as

$$R_p = \frac{\partial E_p}{\partial I_p}, \text{ that is, } R_p = \left[\frac{dE_p}{dI_p}\right]_{E_g \text{ constant}} \tag{488 p}$$

This of course also describes the reciprocal of the slope of a plate characteristic curve. Since this slope is by no means the same for all values of plate and grid voltages (see Fig. 50 and other similar sets of characteristic curves in Chapter VI), it follows that the value of R_p depends on the circuit assembly in which a tube is used.

The equation

$$s_m = \frac{\partial I_p}{\partial E_g}, \text{ that is, } s_m = \left[\frac{dI_p}{dE_g}\right]_{E_p \text{ constant}} \tag{489 p}$$

defines the *grid-plate transconductance s_m*. The grid-plate transconductance is therefore the slope of a mutual characteristic curve and, like R_p, may have a variety of values for a given tube. Transconductance is sometimes stated in mhos, but more often in micromhos.

134. The Amplification Factor. Equation (282) describes in a general way the type of volt-ampere relationship to be expected in a triode, the essential feature being dependence of plate current on an equivalent voltage, the dependence being nonlinear. Equation (282) is approximately valid if potentials are such that all the electrons go to the plate and the range of space-charge-limited current is not exceeded. The dependence of B, μ, n on the geometry of the tube has been analyzed in preceding chapters, but for the purposes of circuit applications it is satisfactory to treat these quantities as empirical constants. Only one of them, the amplification factor μ, commonly appears directly in circuit equations.
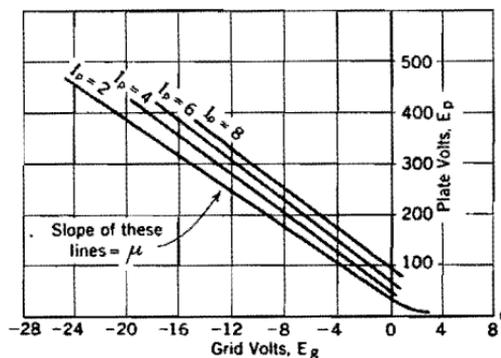
FIG. 101. Plate voltage-grid voltage curves corresponding to Fig. 50. Slope of these curves is the amplification factor μ.

The *amplification factor μ* is, *for purposes of circuit analysis*, defined as follows, quite independently of its significance relative to internal tube geometry:

$$\mu = -\frac{\partial E_p}{\partial E_g}, \text{ that is, } \mu = -\left[\frac{dE_p}{dE_g}\right]_{I_p \text{ constant}} \tag{490}$$

μ is therefore the slope of the various straight parallel slant lines in Fig. 101, for along each line the plate current is constant.

Suppose a small change dE_g occurs in grid voltage, the plate voltage remaining constant, then

$$dI_p = \frac{\partial I_p}{\partial E_g} dE_g = s_m \, dE_g$$

Now suppose a subsequent change dE_p in plate voltage occurs, grid voltage remaining constant at its new value; a second change $\overline{dI_p{}'}$ in current must result. Mathematically,

$$\overline{dI_p{}'} = \frac{\partial I_p}{\partial E_p} dE_p = \frac{dE_p}{R_p}$$

Only if $dI_p + \overline{dI_p{}'} = 0$ does the current return to its original value. Therefore, for constant current

$$s_m \, dE_g + \frac{dE_p}{R_p} = 0$$

or, after rearrangement,

$$-\left[\frac{dE_p}{dE_g}\right]_{I_p \text{ constant}} = s_m R_p$$

Comparison of this last form with Equation (490) shows that

$$\mu = s_m R_p \qquad\qquad\qquad (491 \text{ p})$$

This result is in accordance with a general mathematical principle that is illustrated by the following equation:

$$\frac{\partial I_p}{\partial E_g} \times \frac{\partial E_p}{\partial I_p} = -\frac{\partial E_p}{\partial E_g} \qquad\qquad (492 \text{ p})$$

The relation expressed by Equation (492) seems fairly obvious except for the minus sign; the primary reason for following the mathematical details leading up to Equation (491) is to demonstrate the necessity for that sign. If R_p and s_m are both positive, $\partial E_p/\partial E_g$ must be negative, and μ positive.

The plate characteristic curves of Fig. 3b, also of Fig. 50b, are essentially parallel to one another; that is, each one is like its neighbor except for a horizontal shift. This means that any change ΔE_g in grid voltage can be compensated for by a change ΔE_p in plate voltage, provided that $\Delta E_g + \frac{\Delta E_p}{\mu} = 0$, μ being constant and independent of the magnitude of the grid voltage change. I_p is a function of the equivalent voltage.

Adjacent curves are not parallel to one another in families of tetrode and pentode characteristics, for in such devices the current is not even

approximately a function of an equivalent voltage; see Figs. 55 and 60, pages 128 and 136.   When used in connection with these multigrid tubes the quantity μ, as defined by Equation (490), loses entirely the significance relative to interelectrode geometry that it has in connection with triode characteristics.   It becomes, like R_p and s_m, simply a quantitative measure of a certain decidedly variable property of a family of curves.

135. Evaluation of Tube Constants.   The quantities μ, R_p, and s_m are sometimes called " tube constants " in spite of the fact that only one of them, μ, is constant, and that one so only for triodes.   The values of these " constants " can be obtained by direct measurement, using a vacuum-tube bridge, or graphically from the characteristic curves.

The value of R_p at any point on a set of plate characteristics can be obtained by measuring the slope at that point.   It can be obtained from a set of mutual characteristics by determining from the graph the increments ΔE_p and ΔI_p along a line of constant grid voltage (vertical in this case).   Then

$$R_p = \left[\frac{\Delta E_p}{\Delta I_p}\right]_{E_g \text{ constant}} \tag{493 p}$$

The quantities s_m and μ can be similarly evaluated from the slopes of properly prepared curves, or by incremental methods.

136. Simple Amplifier Circuits; the Load Line.   Triodes, tetrodes, and pentodes are used extensively in amplifier circuits.   The time variations of the voltage between the output terminals of a perfect amplifier follow the pattern set by the input voltage, but are of greater magnitude by a constant factor called the *voltage amplification* or *voltage*
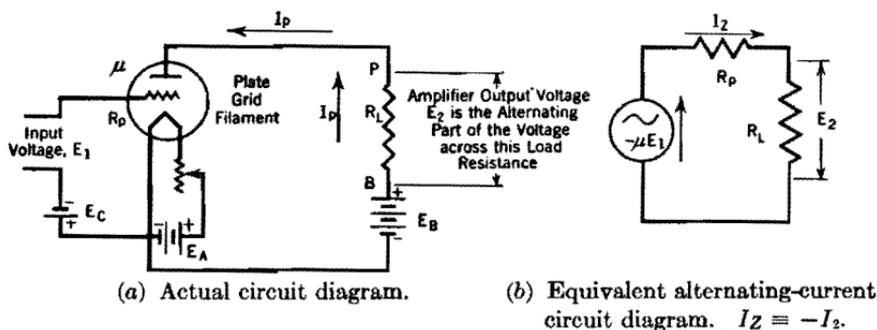
(a) Actual circuit diagram.   (b) Equivalent alternating-current
                                       circuit diagram. $I_Z \equiv -I_2$.
        FIG. 102.   Amplifier circuit with resistance load.

gain of the amplifier.   Most actual amplifiers introduce some distortion of the voltage pattern, and the kind and magnitude of the distortion is frequently a controlling factor in the selection of the type of amplifier to be used for a particular purpose.

There is in general use a classification of commercial amplifiers into

Class A, Class B, and Class C types.[F 145, I 120]   The two latter types are easily understood once the essentials of Class A operation are grasped. Discussion of Class B amplifiers, and of the distinction between the types, is postponed to Section 151.   No analysis of Class C operation is given in this text, because the use of Class C amplifiers is confined almost exclusively to rather specialized radio transmitter applications. [F 289, I 244]

Fig. 102a is a diagram of a simple but typical form of Class A amplifier circuit.   The filament (cathode) of the tube is heated by current from the " A " battery, of voltage E_A.   The plate and grid circuits contain respectively a " B " battery, of voltage E_B, and a " C bias " battery, of voltage E_C.   The plate circuit contains also an *output resistance* or *load resistance* R_L.   The alternating input voltage, of root-mean-square value E_1, instantaneous value e_1, is introduced into the grid circuit in series with the bias battery.   The useful output voltage, of root-mean-square value E_2, instantaneous value e_2, is the alternating component of the voltage across R_L.   The voltage amplification or gain of this amplifier is the ratio

$$\frac{E_2}{E_1} = \text{gain}$$

The input voltage is sometimes called the *excitation voltage*, sometimes the *signal voltage*.

In the following paragraphs the symbols e_g, e_p, and i_p represent respectively the instantaneous values of grid voltage (i.e., grid-to-cathode voltage), plate voltage, and plate current.   These are all usually pulsating quantities.   The grid potential is normally negative relative to the cathode, so that the symbols E_C (bias voltage) and e_g usually stand for negative numerical values.

In a circuit like that of Fig. 102a the instantaneous plate voltage is less than the plate battery voltage by an amount dependent on the plate current, for obviously

$$e_p = E_B - i_p R_L \qquad \text{(494 p)}$$

This equation describes the plate
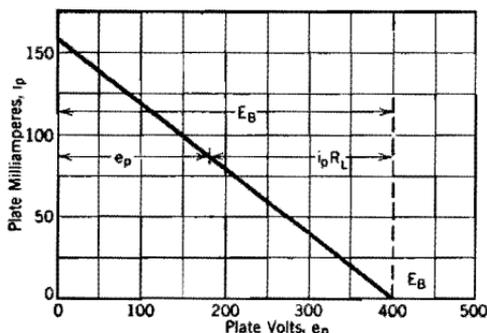
Fig. 103.   Load line, for the circuit of Fig. 102, when $E_B = 400$, $R_L = 2{,}500$.

voltage in terms of the circuit external to the tube.   Its graphical representation on a set of e_p, i_p coordinates is called the *load line*, and is drawn, on Fig. 103, for $R_L = 2{,}500$ ohms and $E_B = 400$ volts. The location of the load line for a given external circuit is entirely independent of the kind of tube used.   To emphasize this fact no

plate characteristic curves are shown in Fig. 103. Actual placement of the load line for a given set of circuit constants is most easily accomplished by locating the two extreme points. Thus in this figure, when $i_p = 0$, $e_p = E_B = 400$ volts; when $e_p = 0$, $i_p = E_B/R_L = 160$ milliamperes. The load line describes graphically the concept that the plate voltage must fall as the plate current rises, and vice versa.

137. Point Q of Zero Excitation; Current-Voltage Locus; Dynamic or Tube-and-Circuit Characteristic. Fig. 104b represents the load line of Fig. 103 combined with a particular set of plate characteristic curves. Suppose the input voltage, E_1, also variously called signal voltage and excitation voltage, for the moment is zero. Then $e_g = E_C$, and both plate voltage and plate current are constant at values corresponding to the intersection of the $e_g = E_C$ plate characteristic curve with the load
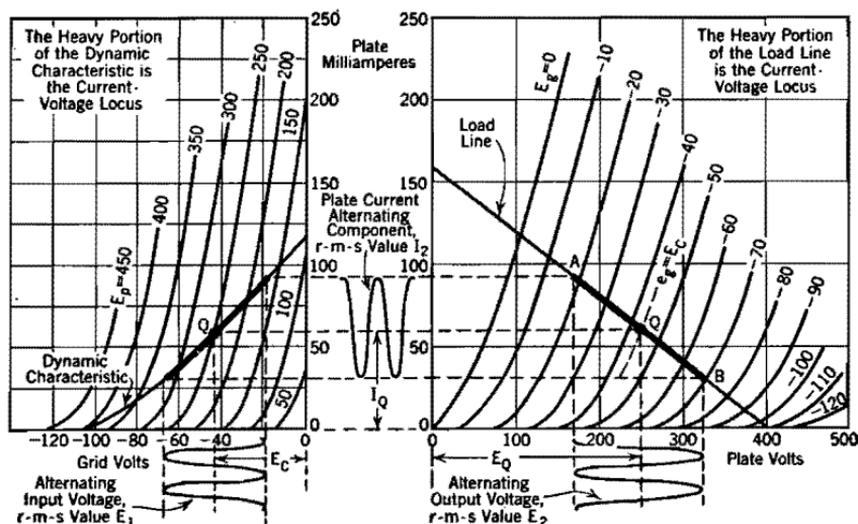
(a) On mutual characteristics.   (b) On plate characteristics.

FIG. 104.   Current and voltage variations in a Class A amplifier with resistance load.

line. This intersection will be called the "point of zero excitation" and identified by the letter Q ("quiescence").[E 193] Corresponding plate current and plate voltage values for this point are symbolized as I_Q and E_Q; the grid voltage at Q is of course E_C.

Suppose that there is introduced into the grid circuit, in series with the grid bias voltage E_C, an alternating voltage of instantaneous value e_1. Then

$$e_g = E_C + e_1 \qquad (495)$$

If e_1 varies sinusoidally, with root-mean-square value E_1 and fre-

quency f, Equation (495) becomes

$$e_g = E_C + \sqrt{2}\ E_1 \cos 2\pi ft \tag{496}$$

Each successive momentary value of grid voltage now corresponds to a new plate characteristic curve, with a new intersection point on the load line. The point describing instantaneous values of e_p and i_p oscillates along the load line with the frequency f. The range of the oscillation for a given value of E_1 is illustrated in Fig. 104b by the heavy portion AB of the load line, extending from $e_g = E_C + \sqrt{2}\ E_1$ to $e_g = E_C - \sqrt{2}\ E_1$. This portion of the load line is the *current-voltage locus* for this particular circuit and excitation voltage.

According to Fig. 103 and Equation (494) any variation in e_p must be accompanied by an equally large variation in the voltage across the load resistance, for E_B cannot change. Hence, as indicated in Fig. 104b, the peak value of the alternating-current output voltage (root-mean-square value E_2) is half the total extent of the e_p variations of the locus. Similarly the peak value of the plate current's alternating-current part (root-mean-square value I_2) is half the total extent of the i_p variations of the locus.

Graphical representation of the operating behavior on a set of mutual characteristics is often more useful than that on the plate characteristic graph just discussed. The load line can be directly transferred to the E_g, I_p graph by plotting thereon the values of e_g and i_p that correspond to the intersections of the various plate characteristic curves with the load line. The curve resulting from such a transfer process is called the *dynamic characteristic*, not the load line; it is shown in Fig. 104a.

The dynamic characteristic differs in nature from the load line in that the dynamic characteristic is not in general straight, and depends for its placement on properties both of the tube and the external circuit. Its flexion results directly from the flexion of the tube characteristic curves, as is evident from the manner of spotting points along it. If the tube characteristic curves are straight or nearly straight within a given range of plate currents, the dynamic characteristic is likewise straight or nearly so within the same current range.

The use of the name " dynamic characteristic " is a little misleading. It is likely to suggest to a beginning student of tube circuits a curve describing response only to alternating excitation voltages. Actually the dynamic characteristic represents the response of the circuit of Fig. 102a both to direct-current and to alternating-current excitation voltages.

The heavy portion of the dynamic characteristic in Fig. 104a is the current-voltage locus that corresponds to the similar heavy portion of the load line in Fig. 104b. The value of e_g at the point Q of zero excitation

is of course E_C. Therefore Q lies just half-way between the extreme points of the " swing " of the grid voltage, on Fig. 104a.

138. Relations between Alternating-Current and Direct-Current Components of Voltage and Current. Suppose that, as in the preceding section, a sinusoidal excitation voltage of root-mean-square value E_1 is used in the circuit illustrated in Fig. 102a. Equations (495) and (496) then describe the relations between the direct-current and alter-

FIG. 105.   Relations between direct-current and alternating-current components of plate voltage and plate current in a Class A amplifier with resistance load.

nating-current components of grid voltage. These relations are illustrated graphically in Fig. 105, in which T is any point along the current-voltage locus.

The plate current has a direct-current component I_Q, and an alternating-current component of instantaneous value i_2. The relation

$$i_p = I_Q + i_2 \qquad (497)$$

defines i_2. This definition is illustrated in Fig. 105. If I_2 is the root-mean-square value of i_2, and i_2 is sinusoidal,

$$i_p = I_Q + \sqrt{2}\, I_2 \cos 2\pi f t \qquad (498)$$

An inspection of Fig. 105 indicates that as the grid voltage swings in the positive direction (toward zero) the plate current increases (i.e., moves to the left and *upward* along the load line). Thus e_1 and i_2 rise and fall together, and reach maximum positive values at the same instant. They may be said to be in phase with one another. The

t's in Equations (496) and (498) are therefore identical; that is, they are measured from a common zero-instant.

Equations (495) and (497) describe the relations between the alternating and direct-current components of e_g and of i_p respectively, The relation between the alternating and direct-current components of plate voltage e_p is similarly described as follows:

$$e_p = E_Q + e_2 \tag{499}$$

The total load voltage e_{BP} (see Figs. 102a and 106) is of course the voltage drop in R_L due to the flow of plate current. Fig. 106 illustrates the fact that e_{BP}, measured from B to P, Fig. 102a, is negative. That is, the potential of P is below that of B.

Fig. 106 also illustrates the fact that as e_p increases, e_{BP} becomes less negative by an equal amount; that is, e_{BP} and e_p experience identical positive increments simultaneously. Therefore e_2 in Equation (499) is the *alternating component of load voltage as well as of plate voltage. Thus e_2 is in fact the useful output voltage* of the amplifier illustrated in Fig. 102a. The point T in Fig. 105 is chosen so as to make e_2 negative, while P in Fig. 106 corresponds to a positive value of e_2.

It is evident from Fig. 105 that:

$$\text{Slope of the load line} = \frac{i_p - I_Q}{e_p - E_Q} = \frac{i_2}{e_2} \tag{499.1}$$

But this slope is $-1/R_L$; therefore

$$e_2 = -i_2 R_L \tag{500 p}$$

An inspection of Fig. 105 shows that, along the load line, i_2 increases in a positive sense as e_2 swings negative, just as is to be expected from Equation (500).

A relation similar to Equation (500) exists between the corresponding root-mean-square values, that is

$$E_2 = -I_2 R_L \tag{501 p}$$

The negative sign in Equations (500) and (501) is of importance only when phase angles are being considered.

Potential at B, Fig. 102
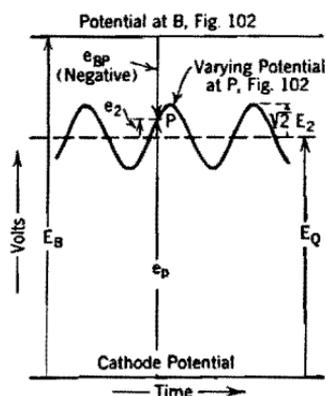
Fig. 106.   Voltage relations in Class A amplifier plate circuit.

The counterpart, for plate voltage variations, of Equation (498) is of course

$$e_p = E_Q + \sqrt{2}\, E_2 \cos 2\pi ft \qquad (502)$$

The effect of the negative sign is made evident by a comparison between Equation (498) and the following, obtained by using $E_2 = -I_2 R_L$ in Equation (502)

$$e_p = E_Q - \sqrt{2}\, I_2 R_L \cos 2\pi ft \qquad (502.1\ p)$$

Thus as i_p increases, e_p decreases; i_2 and e_2 are 180° out of phase with one another.

In summary:

(a) e_1 and i_2 rise and fall together

(b) e_2 falls when i_2 rises, therefore, necessarily,

(c) e_2 falls when e_1 rises.

The foregoing explains why e_2 and e_1 are said to be 180° out of phase with one another in an amplifier that has a purely resistive load.

139. Elliptical Current-Voltage Locus with Reactive Load. Often the load is reactive; for example, it may contain an inductance of L henrys, having reactance X_L ohms. A circuit with a load of this nature is illustrated in Fig. 107a. If the *plate current* is a sine wave of a single frequency, uncomplicated by the presence of harmonics, the

(a) Actual circuit.   (b) Equivalent alternating-current circuit.
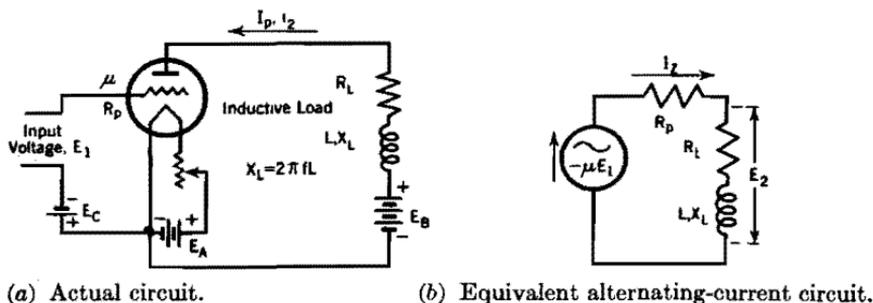
Fig. 107. Class A amplifier circuit with inductive load.

corresponding current-voltage locus on a set of plate characteristics is a true ellipse, of the type illustrated in Fig. 108b. Furthermore, the shape of the ellipse is, for sine wave i_2, not at all dependent on the shapes of the tube characteristic curves. The plate battery voltage is fixed when the battery is chosen, and the voltage drop between the plate battery and the plate depends only on the amplitude of the sine wave current variation and on the R_L, X_L of the load.

Current and voltage variations in the plate circuit of Fig. 107a must

obey the following circuit equation:

$$e_p = E_B - i_p R_L - L\frac{di_p}{dt} \tag{503 p}$$

Certain general features of the ellipse in Fig. 108b can be determined by a study of this equation.

The upper and lower extremes of the ellipse, points A, B, Fig. 108b, correspond to the positive and negative peak values of the alternating-

(a) On mutual characteristics.   (b) On plate characteristics.
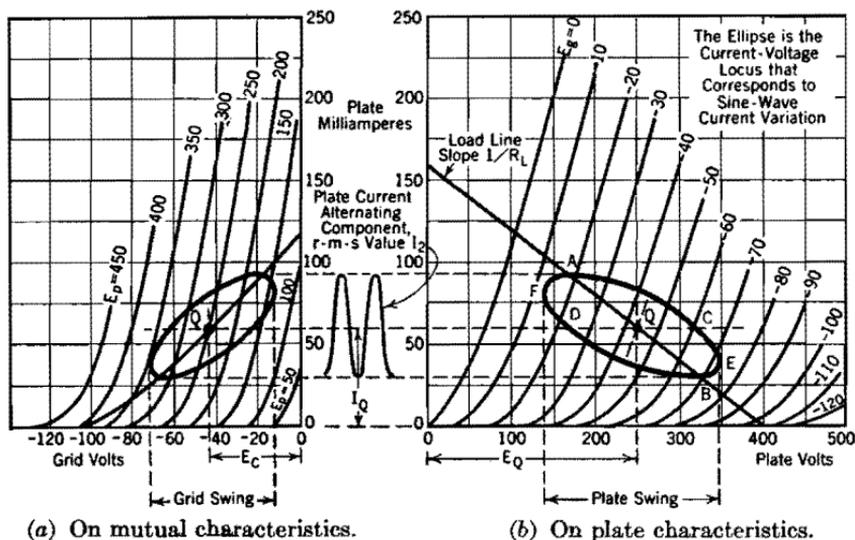
FIG. 108.   Current and voltage variations in Class A amplifier with inductive load, sine-wave current variation.

current component i_2.   At both points the rate of change of i_2, so also of i_p, is zero.   Therefore $L\,di_p/dt$ is zero in Equation (503), and the expression for plate voltage reduces to

$$e_p = E_B - i_p R_L$$

But according to Equation (494) any point whose coordinates are i_p and $E_B - i_p R_L$ lies on a load line drawn for load resistance R_L and plate battery voltage E_B.   This reasoning therefore demonstrates that the upper and lower extremes of the ellipse must lie on such a load line. This is equivalent to saying that *the ellipse must have horizontal tangents at both intersections with the load line.*

Equation (497) may be introduced into Equation (503).   If it is then recognized that $dI_Q/dt = 0$ and that $E_B - I_Q R_L = E_Q$, the following expression results:

$$e_p = E_Q - i_2 R_L - L\frac{di_2}{dt} \tag{504 p}$$

The two points C, D, at which the current is just I_Q, correspond to passage of i_2 through zero. When i_2 is zero, di_2/dt is a maximum. The inductive voltage in the load is therefore at its maximum value, which is $\sqrt{2}\,I_2X_L$. At this moment i_2R_L is zero. Thus at C and D the plate voltage is respectively just $\sqrt{2}\,I_2X_L$ volts less and greater than E_Q. These two points are of course horizontally in line with Q.

Comparison of Equations (499) and (504) shows that in this circuit (Fig. 107a)

$$e_2 = -\left(i_2R_L + L\frac{di_2}{dt}\right) \tag{505 p}$$

If i_2 is sinusoidal it can be expressed as

$$i_2 = \sqrt{2}\,I_2 \cos 2\pi ft \tag{506 p}$$

Equation (506) may be introduced into Equation (505), and the resulting expression handled as in elementary alternating-current circuit theory. Such treatment leads to a sinusoidal expression for e_2, of the form

$$e_2 = -\sqrt{2}\,I_2\sqrt{R_L^2 + X_L^2} \cos\,(2\pi ft + \theta) \tag{507 p}$$

E_2 is the root-mean-square value of e_2, so that

$$E_2 = -I_2\sqrt{R_L^2 + X_L^2} = -I_2Z_L \tag{508 p}$$

This reduces to Equation (502) if $L = 0$.

The angle θ is that between E_2 and I_2; it is such that

$$\cos\theta = \sqrt{\frac{R_L}{R_L^2 + X_L^2}} \tag{509 p}$$

The extreme right and left points E, F, in Fig. 108b, correspond to peak values of the alternating component of load voltage, so that they must be $\sqrt{2}\,E_2$ volts to the right and left respectively of Q, and above and below it by an amount depending on the value of θ.

The parametric equations for the ellipse can be written by using Equations (506) and (507) in Equations (497) and (499). They are

$$i_p = I_Q + \sqrt{2}\,I_2 \cos 2\pi ft \tag{510 p}$$

$$e_p = E_Q - \sqrt{2}\,I_2Z_L \cos\,(2\pi ft + \theta) \tag{510.1 p}$$

By assuming arbitrary values for $2\pi ft$ (in radians or degrees) corresponding values of i_p and e_p can be determined from these parametric equations if E_Q, I_Q, R_L, X_L, and I_2 are known. Points around the ellipse can then be spotted.

The elliptical current-voltage locus of Fig. 108b transfers to a similar figure on the mutual characteristic graph, Fig. 108a, in exactly the way that the load line of Fig. 104b transfers to the dynamic charac-

teristic of Fig. 104a.   If the tube characteristics are straight or nearly so within the area enclosed by the ellipse on Fig. 108b, the locus on Fig. 108a derived from it is also a true ellipse.   The grid excitation necessary to produce a sine wave plate current is then itself sinusoidal.   If there is appreciable curvature of the plate characteristic curves within the locus first discussed, the mutual-characteristic locus is only approximately elliptical.

The analysis just given of elliptical loci has been based specifically on a simple R, L load circuit.   The manner of approach used can, however, easily be adapted to load circuits consisting of much more complicated combinations of resistances and reactances.   A load circuit may contain condensers if a parallel path is provided for the passage of the direct-current part of the plate current.   The $\sqrt{R_L{}^2 + X_L{}^2}$ of the simple circuit just treated must be replaced by the impedance Z_L of the more complex circuit, and phase angles calculated from vectors in the usual way.

140. The Alternating-Current Equivalent Circuit; Phase Reversal in an Amplifier.   It is convenient to be able to express the gain of an amplifier quickly and easily in terms of tube constants evaluated at the operating point Q.   This is customarily accomplished by the aid of an imaginary equivalent alternating-current circuit.   For example, the equivalent circuit shown in Fig. 102b is the one to be used in predicting the behavior of the actual circuit diagrammed in Fig. 102a.   The equivalent circuit consists in general of the load in series with an imaginary resistance R_p and an imaginary alternator generating a voltage $-\mu E_1$.   The negative sign in $-\mu E_1$ is a helpful reminder of the phase shift introduced by an amplifier, besides being mathematically correct.

The instantaneous and root-mean-square values of current in the equivalent circuit are i_z and I_z, where

$$i_z = -i_2$$
$$I_z = -I_2 \tag{511}$$

i_2 and I_2 have the meanings assigned in Section 138.   The quantities μ and R_p have the values that exist at Q in the actual circuit.   The use of this equivalent circuit is justified in the following paragraphs.

Horizontal distances in Fig. 105, page 270, can be measured in both plate-voltage and grid-voltage units.   The conversion ratio between results of the two kinds of measurement is $-\mu$; see Equation (490). For example, the difference between the plate voltage at C and that at Q is $-\mu$ times the corresponding difference in grid voltage.   This must be true in order to make the equivalent voltage and therefore the plate current the same at Q as at C.

The grid voltage, therefore also e_1, is the same at T as at C, because these two points lie on a common plate characteristic curve. Using this fact, and measuring distances in plate voltage units:

$\overline{CQ} = \mu e_1$ (positive ˉat C, for $\overline{CQ}$ is a positive distance; ˉfurthermore μ, also e_1 at T, are positive).

$\overline{CM} = i_2 R_p$ (for R_p is the reciprocal of the slope of the plate characteristic curve)

$\overline{MQ} = i_2 R_L$

But of course

$$\overline{CQ} = \overline{CM} + \overline{MQ}$$

Therefore

$$\mu e_1 = i_2 R_p + i_2 R_L \tag{512 p}$$

The following form of Equation (512) is often useful:

$$i_2 = \frac{\mu}{R_p + R_L} e_1 \tag{512.1 p}$$

As e_1 and i_2 rise and fall together, that is, are in phase with one another, no negative signs appear in Equations (512) or (512.1).

The most useful relation between the various alternating components is that between e_2 and e_1, obtained by multiplying Equation (512.1) through by R_L, then using Equation (500), with the following result:

$$e_2 = -\mu e_1 \frac{R_L}{R_p + R_L} \tag{513}$$

Voltages in a series resistive circuit divide in proportion to the various resistances. Therefore in the equivalent circuit, Fig. 102b, the relation between E_2 and $-\mu E_1$ is in exact accordance with Equation (513). That is, according to both the real and equivalent circuits, the voltage gain of the amplifier is

$$\frac{E_2}{E_1} = -\mu \frac{R_L}{R_p + R_L} \tag{513.1}$$

It is not convenient in equivalent circuit calculations to deal with a current I_2 that is negatively related to E_2, as in Equation (501). It is convenient instead to define and use a current i_z, of root-mean-square value I_z, such that

$$E_2 = I_z R_L, \; e_2 = i_z R_L \tag{514}$$

It is evident that $I_z = -I_2$, as stated earlier, in Equations (511).

I_Z will be found to be the current in a generalized load impedance Z_L.

On using $-i_z$ instead of i_2 in Equation (512.1) it appears that

$$i_z = \frac{-\mu e_1}{R_L + R_p} \tag{515 p}$$

also that

$$I_Z = \frac{-\mu E_1}{R_L + R_p} \tag{515.1 p}$$

These expressions are of course in entire agreement with the equivalent circuit, Fig. 102b.
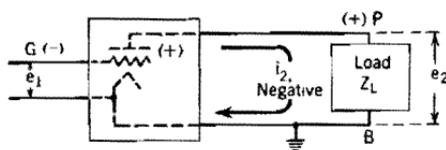
(a) *Electronic amplifier.* To make P momentarily positive as here shown, the alternating load current must flow in the direction indicated by the arrow. However, this direction of current flow corresponds to a *negative* value of i_2, so also to a negative value of e_1. Thus for the instant here shown G is negative. (See Section 138.)
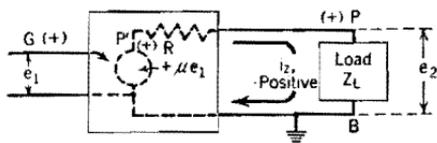
(b) *Ideal amplifier.* To make P momentarily positive as shown, the alternating load current i_z flows in the direction indicated by the arrow. This is the *positive* direction for i_z. When P is positive, P' and G are also positive. No such amplifier as this exists; however, the polarities here shown are the same as in alternating–current transformer circuit diagrams as usually drawn.
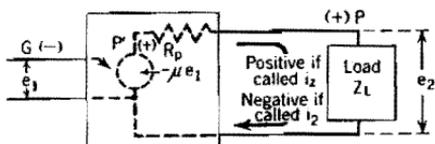
(c) *Equivalent circuit used for predicting electronic amplifier behavior.* To make P momentarily positive as shown, the alternating load current must flow in the direction indicated by the arrow. This is the *negative* direction for i_2, the *positive* direction for i_z. When P is positive, P' is still more positive, as in (b), but G is negative, as in (a). Thus when e_2 and i_z have positive values, e_1 has a negative value.

Fig. 109. Alternating–current equivalent amplifier circuits.

An actual amplifier may be skeletonized as in Fig. 109a. All direct-current elements have been omitted from this diagram, and only alternating-current voltages and currents are considered in connection with it.

The polarity relationships stated underneath it are in accordance with Fig. 105 and the discussion in Section 138. These polarity relationships exist because the actual effect of driving the grid potential negative is to increase the obstruction offered by the tube to the flow of plate current. The resulting decrease of current through the load to the plate is accompanied by a rise of plate voltage to a more positive value.

An ideal type of amplifier would have polarity relationships like those illustrated in Fig. 109b. Such an amplifier would contain an alternating-current generator with internal voltage proportional to e_1. The proportionality factor might well be called μ. The amplifier might also contain a constant internal resistance, denoted by R. It is apparent that the behavior of such an ideal amplifier could be very easily predicted by the methods of elementary alternating-current circuit theory. No confusion in regard to polarities would ever arise. The current i_z in the load of impedance Z_L would normally be called positive when P is positive, as shown in the figure.

The equivalent circuit actually used in predicting the behavior of electronic amplifiers is illustrated in Fig. 109c. It is an adaptation of the ideal amplifier concept to actual conditions. The polarity relationship between i_z and e_2 is made just the same as in the ideal amplifier, in order to regularize circuit calculations. However, in the true amplifier, Fig. 109a, i_2 is *negative* when P swings positive, whereas, in Figs. 109b and 109c, i_z is *positive* when P swings positive. Thus i_z is just another name for $-i_2$.

$-\mu e_1$, rather than $+\mu e_1$, is used in Fig. 109c in order to account for the fact that the grid G must in reality swing negative to make the plate P swing positive, as is required if i_z is to be positive.

A vector diagram for the equivalent circuit of Fig. 109c is shown in
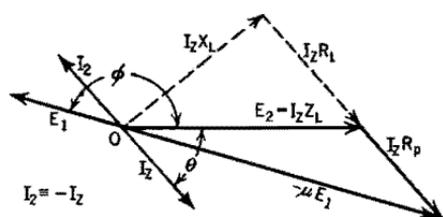
FIG. 110. Vector diagram for amplifier equivalent circuit, Fig. 109c. Dotted vectors indicate components of E_2 if Z_L consists of R_L, X_L in series, as in Fig. 107.

Fig. 110. The vector E_1 has a direction opposite to that of $-\mu E_1$, and is smaller in the ratio $1/\mu$. The angle between E_2 and I_Z is evidently that between E_2 and $-I_2$; it was assigned the symbol θ in Section 138. The angle ϕ between E_2 and E_1 is usually greater than 90°, and is 180° in an amplifier with resistance load (Fig. 102).

Analysis of amplifier circuit behavior by means of the equivalent circuit is justifiable only if both μ and R_p are reasonably constant throughout the range of the current-voltage locus. Values of μ and

R_p at Q are used because such values are presumably representative of true values throughout the range of the locus.

If the load contains an inductance, the point used (as T, Fig. 105 has just been used) for graphical analysis of instantaneous relations, will lie along an elliptical locus, of the type illustrated in Fig. 108b. It will therefore in general lie to the right or left of the load line by an amount $L \frac{di_2}{dt}$, as indicated by Equation (504). Then for a point M that is still, as in Fig. 105, directly below T,

$$\overline{MQ} = i_2 R_L + L \frac{di_2}{dt}$$

giving

$$\mu e_1 = i_2 (R_L + R_p) + L \frac{di_2}{dt}$$

Use of the relation $i_2 = -i_Z$ converts this to the following

$$-\mu e_1 = i_Z (R_L + R_p) + L\left(\frac{di_Z}{dt}\right) \tag{516 p}$$

Equation (516) is of course the circuit equation for an equivalent circuit having an inductance L in series with the load resistance (Fig. 107b). The method used in obtaining it can be generalized to satisfy any complicated assembly of resistances and reactances in the load. The general vector relation

$$-\mu E_1 = I_Z R_p \oplus I_Z Z_L \tag{517 p}$$

involving a generalized load impedance Z_L, follows directly once the instantaneous relation is established. Fig. 110 illustrates this vector relation.

The result expressed by Equation (517) can be arrived at entirely analytically. For a triode, at least, the plate current is known to be a function of the equivalent voltage $e_g + (e_p/\mu)$. Suppose attention be confined to a range of current values within which all the tube characteristics are straight lines. Within this range the dependence of current on the equivalent voltage must be *linear*, and can be expressed mathematically as follows:[*]

$$i_p = I_0 + a\left(e_g + \frac{e_p}{\mu}\right) \tag{518 p}$$

Here I_0 and a are constants. At the point Q this relation becomes

$$I_Q = I_0 + a\left(E_C + \frac{E_Q}{\mu}\right) \tag{519 p}$$

---

[*] For the treatment of the plate current as an infinite series of terms containing e_g and e_p, see Glasgow, [E 337] also McIlwain and Brainerd.[G 112]

the constants being unchanged. Subtraction of Equations (518) and (519) and substitution from Equations (495), (497), and (499) lead to the following result:

$$i_2 = a\left(e_1 + \frac{e_2}{\mu}\right) \qquad (520\ \text{p})$$

This result might have been written to begin with, for it is only good sense to expect that if total quantities are linearly related their differences are also, and in a similar way.

Partial differentiation of Equation (520) relative to excitation voltage shows that

$$\frac{\partial i_2}{\partial e_1} = a \qquad (521\ \text{p})$$

But since i_2 and e_1 differ from i_p and e_g only by constant amounts,

$$\frac{\partial i_2}{\partial e_1} = \frac{\partial i_p}{\partial e_g}$$

Therefore, using Equation (489) and the general relation $\mu = s_m R_p$,

$$a = \frac{\mu}{R_p} \qquad \qquad (522\ \text{p})$$

If this value for a is used in Equation (520), rearrangement gives

$$\mu e_1 = i_2 R_p - e_2 \qquad (523\ \text{p})$$

When $-i_z$ is substituted for i_2, Equation (523) becomes

$$-\mu e_1 = i_z R_p + e_2 \qquad (524\ \text{p})$$

This relation is in accordance with the equivalent circuit for any type of load; see Fig. 109c. Equation (517) follows directly.

141. Slope of the Dynamic Characteristic. It will be convenient in later discussions of harmonic distortion to know what determines the shape of the dynamic characteristic curve. The *slope* of the dynamic characteristic is of course uniform within the range of straight-line tube characteristics. Its value is di_p/de_g, therefore also di_2/de_2.

A mathematical expression for this slope is obtained by differentiating Equation (512.1) with respect to e_1. The result, expressed in terms of i_p and e_g rather than i_2 and e_1, is

$$\frac{di_p}{de_g} = \frac{\mu}{R_p + R_L} \qquad (525\ \text{p})$$

The derivation, in the preceding section, on which Equation (512.1) is based can be applied to any local portion of a curved dynamic charac-

teristic by making T and Q approach sufficiently closely to one another. Hence Equation (525) expresses the slope at *any* point along a curved dynamic characteristic in terms of the values of μ and R_p effective at that point.

It should be noted that di_p/de_g is distinctly different from $\partial i_p/\partial e_g$, which is the slope of a mutual characteristic curve. The total derivative expresses the rate of change of i_p with respect to e_g when e_p changes as demanded by the amplifier circuit assembly. The partial derivative expresses such a rate when e_p as well as all other quantities except e_g and i_p remain constant.

142. Uses and Limitations of the Equivalent Circuit; Harmonic and Frequency Distortion. It is apparent from the derivations in Section 140 that the equivalent circuit must be thought of in terms of alternating-current quantities exclusively. The total current in it is I_Z, which is equal and opposite to the alternating part of the plate current in the actual circuit. The voltage across the load in the equivalent circuit is just E_2, the alternating component of plate voltage in the actual circuit. Batteries are essentially direct-current devices, therefore do not appear in the equivalent circuit. Battery internal resistances should be included, in series with the load, if they are appreciable relative to tube and load resistances. This is rarely the case.

The current in the equivalent circuit of a Class A amplifier with resistance load, Fig. 102b, is evidently related to the other circuit quantities as follows:

$$I_Z = \frac{-\mu E_1}{R_p + R_L} \qquad (526\ p)$$

The output voltage E_2 is of course $I_Z R_L$, so that the gain E_2/E_1 of the amplifier is given by the expression

$$\text{Voltage Gain} = \frac{E_2}{E_1} = -\mu \frac{R_L}{R_p + R_L} \qquad (527\ p)$$

An alternate form of this relation is

$$\frac{E_2}{E_1} = -\mu \frac{1}{1 + \dfrac{R_p}{R_L}} \qquad (528\ p)$$

In subsequent expressions for gain similar to these the negative sign will often be omitted for simplicity.

The expressions just given and other similar ones for more complicated load circuits are valid *to the extent that all the tube characteristic curves have the same and a constant slope throughout the extent of the current-voltage locus.* Of course strictly speaking the characteristic curves are

straight lines only for infinitesimal current variations. Practically, however, their slopes in the operating range are ordinarily nearly uniform. Equivalent circuit calculations may in most cases be expected to give very satisfactory approximations to the actual circuit behavior.

The usual practical criterion, for resistive loads, of the range within which the equivalent circuit calculations are valid is the *uniformity of slope of the dynamic characteristics* (see Section 141). The dynamic characteristic is often practically straight over a considerably wider range than are the mutual characteristic curves.

If the load is reactive, the current-voltage locus is not confined to the dynamic characteristic. In this case the degree of uniformity of the quantity $\mu/(R_L + R_p)$ within the entire area of the elliptical current-voltage locus is the criterion of freedom from harmonic distortion.

If the tube characteristic curves do not have the same and constant slopes throughout the extent of the current-voltage locus, there appear in the load circuit *harmonics*. These are alternating currents and voltages whose frequencies are twice, three times, four times, etc., that of the *fundamental* or excitation-voltage frequency. Operation under such conditions is said to result in *harmonic distortion*. The magnitudes of the various harmonic voltages and currents depend partly on the nature of the flexion of the tube characteristics, and partly on the reactive properties of the load circuit. Only rarely is it necessary to consider harmonics higher than the third.

It is customary to measure the amount of harmonic distortion by means of percentages. Thus there may be a 15 per cent second harmonic and an 8 per cent third harmonic. This means that the root-mean-square second and third harmonic voltages in the output are respectively 15 per cent and 8 per cent of the fundamental-frequency voltage. The root-mean-square value of a combination of voltages of different frequencies is the square root of the sum of the squares of the contributing root-mean-square values. Therefore the total harmonic distortion in such a case would be $\sqrt{15^2 + 8^2} = 17$ per cent. See Sections 148, 149, and 150 for further discussions of harmonic distortion.

The alternating-current equivalent circuit is especially useful in the study of *frequency distortion*, which results from nonuniformity of gain at different excitation frequencies. In many cases the gain may be expected to be reasonably constant over a definite range of frequencies but to drop off rapidly outside of that range. This is illustrated in Fig. 128, page 122. Speech or music contains many frequencies. If some lie within the range of uniform gain of an amplifier used for voice reproduction and some outside of it, not all are amplified alike. Such nonuniform amplification is said to introduce frequency distortion.

143. Choice of Tube and Load Resistances. The general principles involved in the selection of proper relative magnitudes of tube and load resistances can be satisfactorily illustrated by reference to the simple amplifier circuit of Fig. 102. It should first be noted that according to Equation (528) the greatest possible voltage gain is μ. That value cannot be reached except by making R_L infinite, which is absurd as it would require an infinite plate battery voltage to supply the direct-current part of the plate current. Practically, a compromise must be reached between the desirability of high voltage gain and the expense and power waste inherent in a high plate battery voltage. Fig. 111 illustrates graphically the fact that the gain increases very rapidly with rising load resistance until R_L is several times R_p, but that continued increase to tens or hundreds of times R_p gives very little further improvement. Hence there arises a general working principle that R_L should be large enough to make the gain between two-thirds and nine-tenths of μ, if the primary function of the amplifier is to provide voltage amplification.
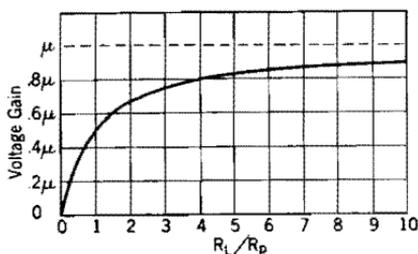
Fig. 111. Dependence of gain on the ratio of load resistance R_L to plate resistance R_p.

A *power amplifier* is one used to deliver considerable alternating-current power, rather than a high output voltage, at the load. The proper ratio of tube to load resistance for a power amplifier depends on which one of the three following sets of circumstances exists:

(1) The excitation voltage is small, and it is desired to choose a load resistance that will make power delivered to the load as great as possible for a specified tube. In this case μ and R_p are fixed. This problem is identical with that of determining the load resistance into which a separately excited generator delivers maximum power. The power delivered to the load of the equivalent circuit, Fig. 102b, is of course $I_Z^2 R_L$; using Equation (526)

$$\text{A-c power delivered to load} = (\mu E_1)^2 \frac{R_L}{(R_p + R_L)^2} \qquad (529\ p)$$

The value of R_L needed to make this power a maximum is obtained by differentiating the whole expression with respect to R_L and equating the result to zero. It is found that the required condition is met when

$$R_L = R_p \qquad (530)$$

(2) The excitation voltage is small, and it is desired to deliver maximum power to a load of fixed resistance R_L. The power delivered to the

load is as before given by Equation (529), but now R_p is the variable. It is apparent from an inspection of this relation that with all other quantities fixed the power is greatest when R_p is least. However, very little can be gained by reduction of R_p below $\frac{1}{4}$ or $\frac{1}{5}$ of R_L. This requirement is therefore satisfied in a practical sense when R_p is a moderately small fraction of R_L. This analysis explains the occasional need for tubes of very low plate resistance.

(3) The excitation voltage may be made as large as desired, but the tube and its properties are specified. It is desired to choose a load resistance that will give maximum undistorted power output. A graphical study of this problem, carried out in the next section, shows that the stated requirement is met when

$$R_L = 2R_p \tag{531}$$

144. Maximum Undistorted Power Output. Fig. 112 is a diagram of the plate characteristic curves of a tube that is to be used for power amplification in a circuit like that of Fig. 102a. Preliminary estimates and manufacturers' specifications suggest the use of an operating voltage E_Q. That is, the operating point Q is to lie along a vertical line at E_Q
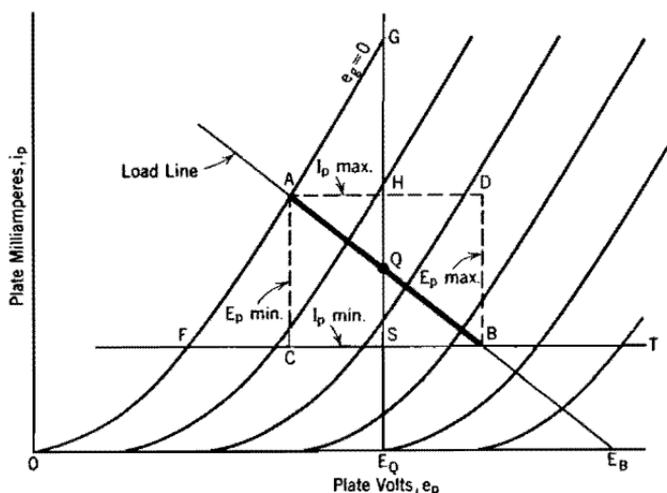
Fig. 112.   Current-voltage locus for maximum undistorted power, output, Class A amplifier with resistance load. $R_L = 2R_p$.

plate volts. It is desired to select the values of load resistance and excitation voltage that will result in maximum undistorted power output.

In order to avoid distortion, (I) the current-voltage locus must remain above a horizontal line FT, Fig. 112, placed so that above it the plate characteristic curves are practically straight, and (II) the current-voltage locus must remain below the $e_g = 0$ plate characteristic curve. If the

grid swings positive, grid current flows during part of each cycle, with resulting distortion of the excitation voltage wave form, and so of the output.

In order to make maximum use of the region available, the current-voltage locus should extend from some point B on the line FT to some point A on FG, the $e_g = 0$ plate characteristic. Furthermore, points A and B must lie equally to the right and left of the vertical line for which $e_p = E_Q$, in order to locate Q properly. They must also lie equally above and below Q.

Let $I_{p\,max}$, $I_{p\,min}$, $E_{p\,max}$, $E_{p\,min}$ symbolize the respective maximum and minimum values of plate current and plate voltage. These are evidently related to rms alternating-current values as follows:

$$E_2 = \frac{E_{p\,max} - E_{p\,min}}{2\sqrt{2}} \tag{532}$$

$$I_Z = \frac{I_{p\,max} - I_{p\,min}}{2\sqrt{2}} \tag{533}$$

The alternating-current power P_L in the load is

$$P_L = E_2 I_Z = \tfrac{1}{8}(E_{p\,max} - E_{p\,min})(I_{p\,max} - I_{p\,min}) \tag{534 p}$$

Graphically, in Fig. 112:

$$\overline{AH} = \tfrac{1}{2}(E_{p\,max} - E_{p\,min}) \tag{535}$$

also

$$\frac{\overline{AH}}{\overline{HG}} = R_p \tag{536}$$

So that

$$E_{p\,max} - E_{p\,min} = 2R_p \cdot \overline{HG} \tag{537}$$

Also

$$I_{p\,max} - I_{p\,min} = \overline{SH} \tag{538}$$

Therefore power can be expressed as

$$P_L = \tfrac{1}{4}R_p \cdot \overline{HG} \cdot \overline{SH} \tag{539}$$

Since R_p is constant above FT, power must be a maximum when the product $\overline{HG} \cdot \overline{SH}$ in this equation is greatest. Now the lines FG, FT, $G\hat{S}$, hence the intersection G and S, are located by imposed conditions, so that the value of $\overline{SH} + \overline{HG}$ is fixed. Thus for maximum power:

(a) $\overline{SH} + \overline{HG}$ must be constant.

(b) $\overline{SH} \cdot \overline{HG}$ must be a maximum.

It will be shown in the following paragraph that these conditions are satisfied only when

$$\overline{SH} = \overline{HG} \tag{540}$$

If $\overline{SH}$ and $\overline{HG}$ are equal, Q must be located at one-fourth of the distance from S to G. Now $R_L = \overline{AH}/\overline{QH}$, but $\overline{QH} = \tfrac{1}{2}\overline{SH} = \tfrac{1}{2}\overline{HG}$; therefore, using Equation (536)

$$R_L = \frac{\overline{AH}}{\tfrac{1}{2}\overline{HG}} = 2\left(\frac{\overline{AH}}{\overline{HG}}\right) = 2R_p \tag{541}$$

This result is independent of the value of E_Q. Therefore *for maximum undistorted power output $R_L = 2R_p$ for any value of E_Q.* The excitation voltage and grid bias required do depend on the choice of operating voltage, and can be read directly from a diagram like Fig. 112, drawn for the particular tube used.

The proof that, if $\overline{SH} + \overline{HG} =$ a constant, and $\overline{SH} \cdot \overline{HG} =$ a maximum, $\overline{SH} = \overline{HG}$ is as follows:

Let $\overline{HG} = f\,\overline{SG}$, where f is any fraction. $\tag{542}$

Then $\overline{SH} = (1 - f)\,\overline{SG}$ $\tag{543}$

Also $\overline{SH} \cdot \overline{HG} = f(1 - f)\,\overline{SG}^2$ $\tag{544}$

Since $\overline{SG}^2$ is a constant, this product is a maximum when $f(1 - f)$ a maximum. Upon differentiating $f(1 - f)$ with respect to f and equating the result to zero f is found to be $\tfrac{1}{2}$, so that

$$\overline{SH} = \overline{HG} = \tfrac{1}{2}\overline{SG} \tag{545}$$

145. Plate Circuit Efficiency and Power Dissipation. Fig. 113 illustrates graphically the various power relationships associated with the operation of a circuit like that of Fig. 102a. Two sets of conditions will be contrasted: (1) that when the excitation voltage is zero, (2) that when there is an excitation voltage, but the current-voltage locus AB does not extend beyond the range in which the plate characteristic curves are straight or nearly so. With this limitation the effect of harmonics need not be considered.

(1) When the excitation voltage is zero the plate current is I_Q, plate voltage E_Q, plate battery voltage E_B, all perfectly steady. Areas in Fig. 113 measure power in watts, as follows:

$OCDG =$ Power introduced into the circuit by the plate battery. (546)

$HQDG =$ Power dissipated as heat in the load resistance, owing to the passage of the direct current I_Q. (547)

$OCQH$ = Power brought to the plate of the tube by the raining on it of electrons which have been accelerated through the plate voltage field.   This power must be removed from the plate by heat radiation, and is called the *plate dissipation*.   (548)
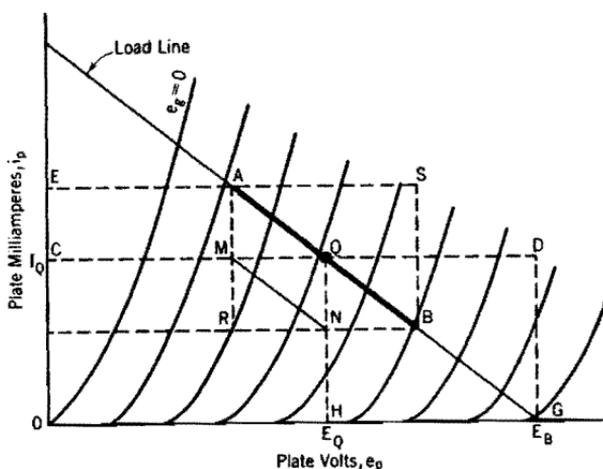
FIG. 113.   Power relationships in a Class A amplifier with resistance load.

(2) When there is an excitation voltage, giving rise to the current-voltage locus AB, the tube and the load resistance carry both direct current and alternating current.   The *average value* of plate current is still I_Q, and the *average value* of the voltage across the load resistance is still $E_B - E_Q$.   As before

$OCDG$ = Power introduced into the circuit by the plate battery. (549)

$HQDG$ = Power introduced into the load resistance *owing to the passage of direct current*.   (550)

However, there is an additional input of power to the load resistance because of the alternating component of current.   As pointed out in the previous section, its value is $\frac{1}{8}(E_{p\,max} - E_{p\,min})(I_{p\,max} - I_{p\,min})$.   This quantity is one-eighth of the area $ASBR$, and can be conveniently indicated by the area QMN.   Thus

QMN = Power introduced into the load resistance *by the passage of alternating current*.   (551)

Plate dissipation is the remainder, that is

$$OCMNH = \text{Plate dissipation} \qquad (552)$$

The interesting feature of this comparison is the indication that the plate dissipation decreases to just the extent that the power output

increases.   The greater the useful output, the less severe is the dissipation burden on the plate of the tube.

The plate circuit efficiency is of course

$$\frac{\text{Alternating-current power in the load}}{\text{Power input from the plate battery}} = \frac{QMN}{OCDG} \qquad (553)$$

It is apparent that ordinary Class A power amplifiers such as illustrated here are not highly efficient devices.   They are nevertheless satisfactory for many purposes because of compensating merits of simplicity and freedom from distortion.

In reviewing analytically the graphical relations just presented it is helpful to recall that the rms value of the total current in a circuit element passing currents of various frequencies is the square root of the sum of the squares of the contributing rms values.   For example, in the load resistance of Fig. 102a there are two currents, I_Q and I_Z, one of zero frequency, the other of excitation-voltage frequency.   Therefore

$$\text{rms value of total current in load resistance} = \sqrt{I_Q^2 + I_Z^2} \quad (554)$$

Total power in any resistance is proportional to the square of the total rms current;   therefore

$$\left.\begin{array}{c}\text{Total power in} \\ \text{load resistance}\end{array}\right\} = (\sqrt{I_Q^2 + I_Z^2})^2 R_L = I_Q^2 R_L + I_Z^2 R_L \qquad (555\ \text{p})$$

Equation (555) demonstrates that the power in the load of the equivalent circuit is the same as the alternating-current power in the load of the actual circuit.

146. Use of Chokes and Condensers to Provide " Parallel Feed " of Direct-Current Power to the Plate.   The discussion in connection with Fig. 113 indicates a rather low efficiency for the type of amplifier illustrated in Fig. 102.   The efficiency can be somewhat increased, and the investment in batteries or equivalent power source reduced, by introducing into the plate circuit a " by-pass " condenser and a large inductance, called a *choke*.   The proper circuit arrangement is indicated in Fig. 114a, page 291.

It is helpful to think of the choke as primarily a device which blocks or " chokes off " alternating current, but offers no obstruction to direct current.   Conversely the condenser blocks the passage of direct current, but offers little or no obstruction to the passage of alternating current. The inductance of the choke, and the capacitance of the by-pass condenser, must be chosen large enough to block and pass respectively currents of the frequency range within which the excitation voltage is to lie.

The complete equivalent circuit for a parallel-feed amplifier is shown in Fig. 114b. However, within the frequency range for which the choke and condenser are designed their reactances approach respectively infinity and zero, so that they may be omitted from the equivalent circuit diagram. It then becomes identical with that of a " series feed " amplifier, Fig. 102b. This section contains a discussion of conditions existing *when the excitation frequency is within this range.*

In the actual circuit the direct-current part of the plate current passes through the choke, with negligible voltage drop, so that *the · plate battery voltage need only be the desired operating-point voltage, E_Q.* The power introduced by the plate battery becomes $E_Q I_Q$, indicated by the area *OCQH* in Fig. 113. The power required to operate the amplifier is therefore less than for the similar " series feed " circuit, Fig. 102a, by the amount of the area *QDGH*. This area represents the power loss in the load resistor due to the passage of direct current, and no such power loss takes place in a parallel-feed amplifier.

The alternating-current part of the plate current passes through the resistance-condenser path. The reactance of the condenser is made small enough so that the $I_Z X_C$ across it is insignificant. Therefore $E_3 = E_2$. That is, the alternating voltage across the resistance-condenser path as a whole is just the load voltage. This same alternating voltage must of course exist across the parallel choke-and-battery path. The choke permits this alternating voltage to be established by an extremely small alternating current. Thus *the total instantaneous plate voltage is the algebraic sum of the steady plate battery voltage and the instantaneous alternating load voltage,* the latter appearing across the choke.

It has just been pointed out that the alternating component of voltage appears across the load resistance in one path, and across the choke in the other. Similarly the direct-current potential difference established by the battery in one path exists across the condenser in the other path. The overall voltage for both paths is the algebraic sum of E_Q and the alternating load voltage.

As in the series feed circuit, an *increase of plate current* above I_Q results in a *decrease of plate voltage* below E_Q. Therefore the summation of alternating- and direct-current quantities is properly described by Equations (497) and (499), for the parallel-feed as well as for the series-feed circuit.

This suggests that *the load line for a parallel-feed amplifier is the same as that for a similar series feed amplifier, provided the point Q of zero excitation and the load resistance are the same;* this is in fact true. The alternating component of the plate current passes through the load resist-

ance in both cases, and in both cases the alternating voltage so produced is algebraically added to the direct-current part, E_Q, of plate voltage.

It is important to note that *the zero-current and zero-voltage intercepts of the load line of a parallel-feed amplifier have no particular significance.* The load line is located by passing a line of slope $1/R_L$ through the chosen point Q. And of course $E_B = E_Q$.

147. Frequency Limitations of Parallel-Feed Amplifiers. The frequency of the excitation voltage of a parallel-feed amplifier may fall below the range for which the choke and condenser are designed. The variation of voltage across the output resistance must then be calculated by means of the complete equivalent circuit, Fig. 114b. For very small frequencies the reactance of the inductance is so small as practically to short-circuit the condenser-resistance circuit, so that as the frequency approaches zero the gain vanishes. A series-feed amplifier is not subject to this limitation; its output voltage follows the pattern of the excitation voltage faithfully no matter how low the frequency.

Often it is required to amplify transient voltage impulses that recur irregularly, as in physiological research into the electrical effects associated with heart-beats. An impulse of this type may rise very rapidly, then decay very slowly. A parallel-feed amplifier would record the rapid rise correctly, but would indicate a more rapid decay than actually occurs in the excitation impulse.

This is an illustration of the general principle that transient impulses must ordinarily be amplified by means of circuits containing only batteries, resistors, and tubes. Circuits containing chokes and condensers can be used for this purpose only if a half-cycle of some low frequency for which they provide satisfactory steady state amplification is longer than the entire duration of each transient impulse.

X_L and X_C are the reactances of choke and condenser respectively in Fig. 114. In order to make the circuit separate the direct from the alternating current effectively, the ratios R_p/X_L and X_C/R_L must be small. Hence the lower the frequency at which effective action is desired, the larger must be the physical dimensions and hence cost of both choke and condenser. Equations (560) and (561) below show that chokes and condensers large enough to result in satisfactory response as far as the magnitude of the output voltage is concerned may yet be small enough to introduce considerable deviation from the 180° phase shift of the series feed circuit.

Fig. 114c is a vector diagram for the complete parallel-feed amplifier equivalent circuit. The current I_Z divides into I_L in the choke and I_C in the resistance-condenser path. These two paths are in parallel, therefore

$$I_L X_L = I_C \sqrt{R_L^2 + X_C^2} \tag{556 p}$$

(a) Actual circuit diagram.   (b) Complete equivalent alternating-current circuit diagram.
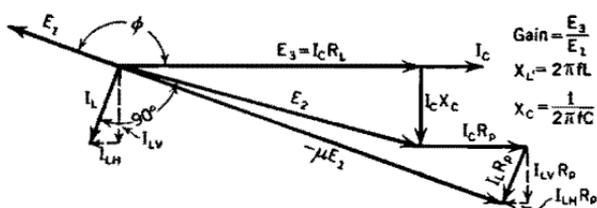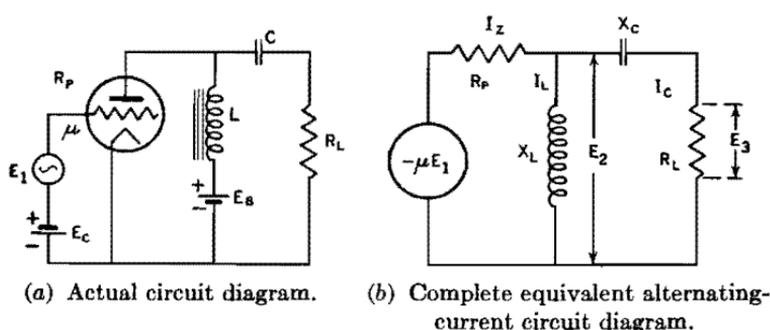
(c) Vector diagram for complete equivalent circuit. Within the normal frequency range the vectors E_2, E_3, $-uE_1$, E_1, all lie along a common line, and $E_3 \equiv E_2$. Then $\phi = 180°$.

Fig. 114.   Parallel-feed Class A amplifier with resistance load.

In the vector diagram:

$$I_{LH}(\text{part of } I_L \text{ in phase with } I_C) = I_C \frac{X_C}{X_L} \qquad (557 \text{ p})$$

$$I_{LV}(\text{part of } I_L \text{ at right angles with } I_C) = I_C \frac{R_L}{X_L} \qquad (558 \text{ p})$$

$$(\mu E_1)^2 = \left[I_C R_L + I_C R_p - I_C R_p \frac{X_C}{X_L} \right]^2 + \left[I_C X_C + I_C R_p \frac{R_L}{X_L} \right]^2 \qquad (559 \text{ p})$$

Equation (559) rearranges into the following expression for gain:

$$\text{Gain} = \frac{E_3}{E_1} = \frac{\mu}{\sqrt{\left[1 + \frac{R_p}{R_L} - \frac{R_p}{X_L}\frac{X_C}{R_L} \right]^2 + \left[\frac{X_C}{R_L} + \frac{R_p}{X_L} \right]^2}} \qquad (560 \text{ p})$$

Equation (560) reduces to Equation (528) when R_p/X_L and X_C/R_L vanish.

The phase angle ϕ between the output voltage E_3 and the input voltage E_1 can be calculated from its sine, which is expressible as follows:

$$\sin\phi = \sin(180° - \phi) = \frac{\dfrac{X_C}{R_L} + \dfrac{R_p}{X_L}}{\sqrt{\left[1 + \frac{R_p}{R_L} - \frac{R_p}{X_L}\frac{X_C}{R_L} \right]^2 + \left[\frac{X_C}{R_L} + \frac{R_p}{X_L} \right]^2}} \qquad (561 \text{ p})$$

The ratios R_p/X_L and X_C/R_L need be only moderately small in order to make the gain approach very closely to the high-frequency value given by Equation (528). For example, if both ratios are 0.2, and $R_L = 2R_p$, the actual gain is 0.661μ, as compared with a value 0.667μ if both ratios are zero. This is only a difference of three-fourths of one per cent. There is, however, a phase shift of 15.3° in this case, that is, E_3 makes an angle of 164.7° with E_1.

PROBLEMS

CHAPTER XII

In order to work properly many of the problems for Chapters XII, XIII, and XIV, it is desirable to have blue-printed, or otherwise reproduced for class use, sets of plate characteristic curves of the following tubes: Western Electric 101-F (Fig. 3b), RCA-6C5 (Fig. 50b), RCA-24 (Fig. 55), RCA-59 (Fig. 60b); also a set of the plate characteristics illustrated in Fig. 104b.

1. A Western Electric Type 101-F triode, characteristics as in Fig. 3, is used in a circuit like that diagrammed in Fig. 102. The grid bias E_c is −8 volts, the plate battery voltage 240, and the load resistance 13,500 ohms.

(a) On a set of plate characteristic curves for this tube draw the load line, and determine from it the values E_Q and I_Q of plate voltage and current at the point Q of zero excitation.

(b) On a separate sheet of graph paper plot the dynamic characteristic that corresponds to the load line.

(c) Locate the ends of the current-voltage locus on both load line and dynamic characteristic, for an excitation voltage whose root-mean-square value is 4 volts.

(d) Determine the voltage gain of this amplifier by inspection of the current-voltage loci. (See also Problem 2, Chapter XIII.)

2. (a) Tube and circuit as in Problem 1. Determine R_p and μ from the plate characteristic curves by graphical methods, and calculate the gain from Equation (527) or (528). Compare with the results obtained in Problem 1.

(b) Measure the slope, at Q, of the dynamic characteristic constructed in (b), Problem 1, and compare with the value obtained by using your values of μ and R_p in Equation (525).

(c) Plot the mutual characteristic curve for $E_p = E_Q$ on the graph that contains the dynamic characteristic. Measure its slope, which is the grid-plate transconductance, and see how nearly you check the result obtained by using your values of μ and R_p in Equation (491).

3. Circuit similar to that in Problem 1, tube the same as there, but the load now contains a 10,000-ohm reactance in series with the 13,500-ohm load resistance:

(a) On your plate characteristic curve set for this tube spot the following six points on the elliptical I_p, E_p current-voltage locus:

> The upper and lower extreme values of I_p
> Two points for which $I_p = I_Q$
> Two points for which $E_p = E_Q$

Also draw vertical lines to which the two ends of the ellipse must be tangent. Sketch in the ellipse.

(b) Transfer this ellipse to an I_p, E_g diagram, in just the way a load line is transferred to become a dynamic characteristic.

4. The plate current of a certain triode is satisfactorily expressed by the following empirical equation:

$$I_p = 0.000170 \left(E_g + \frac{E_p}{8} \right)^{1.7}$$

(a) Find the plate resistance R_p and the grid-plate transconductance s_m, when $E_p = 200$ volts, $E_g = -15$ volts.

(b) Find the plate current under the stated conditions.

(c) If this tube is used in the circuit of Fig. 102, and $E_Q = 200$, $E_C = -15$, I_Q having the value found in (b), what must the plate battery voltage be if $R_L = 10,000$ ohms?

(d) What is the voltage gain of the amplifier represented by this arrangement?

5. Suppose that, in the circuit of Fig. 107, $R_L = 1,500$ ohms, $X_L = 2,000$ ohms at 500 cycles, $R_p = 2,500$ ohms, $\mu = 6$.

(a) Find the voltage gain at 500 cycles, using a vector diagram of the voltages and currents in the corresponding equivalent circuit.

(b) What is the voltage gain at zero frequency?

(c) At a very high frequency?

6. Suppose that R_p is 12,000 ohms for the tube used in an amplifier like that of Fig. 107. The power factor of the load is to be 60 per cent. Choose a load impedance so that the gain will be 0.8μ. Show the vector diagram used in your analysis. Find the phase angle between output voltage and excitation voltage.

7. An RCA-6C5 triode (Fig. 50) is to be used in a Class A amplifier circuit as in Fig. 102. $E_C = -8$ volts, $I_Q = 6$ milliamperes.

(a) State the value E_Q of plate voltage at Q, and determine by graphical means the values of μ and of R_p at Q.

(b) What value of output resistance R_L, in ohms, should be used to give maximum alternating-current power in the output resistance, if the root-mean-square value of the grid excitation is 2 volts?

(c) Using the value of R_L just chosen, determine the voltage gain of the amplifier.

(d) Determine the output power in watts, and the plate circuit efficiency.

8. Tube and circuit diagram as in the preceding problem.

(a) What value of R_L should be used to give "maximum undistorted power output"?

(b) State values of E_C, I_Q, E_Q which you believe will locate Q so as to give "maximum undistorted power output" without having E_p ever exceed 400 volts. Prepare diagrams or graphs which clearly indicate your reasoning.

(c) For this circuit and the circuit constants just chosen, determine the necessary excitation voltage, the voltage gain, the output in watts, the plate circuit efficiency, and the plate dissipation. (See also Problem 3, Chapter XIII.)

9. An RCA-6C5 triode (Fig. 50) is employed in a circuit like that diagrammed in Fig. 102. $R_L = 25,000$ ohms, $E_C = -8$ volts, $E_B = 450$ volts.

(a) On a set of plate characteristic curves for this tube draw in the load line and locate the point Q of zero excitation. What are the values of E_Q and I_Q?

(b) $E_1 = 5$ (root-mean-square value of the excitation voltage). Determine the plate circuit efficiency, and the plate dissipation.

(c) Circuit arrangement is now changed to that shown in Fig. 114. The frequency of the excitation is high enough so that all the alternating current may be assumed to pass through R_L. What value of plate battery voltage must be used to produce operation along the same load line as with the series-feed arrangement, grid bias being unchanged?

(d) Determine the new plate circuit efficiency and plate dissipation.

10. For the parallel-feed amplifier of part (c) of the preceding problem, choose values of C and L such that at 25 cycles per sec:
(1) the gain is 60 per cent of its value at high frequencies, and (2) $X_C/R_L = R_p/X_L$. Also determine the phase angle between output and excitation voltages at 25 cycles.

11. An RCA-24 screen-grid tube (Fig. 55) is to be used in a Class A amplifier.
(a) Draw the circuit diagram, showing proper connections of control grid, screen, and plate. Use parallel feed to the plate.
(b) Grid bias E_C is to be -1.5 volts. Select a value for E_Q that is about in the middle of the useful operating range of the plate characteristic curves. State the corresponding plate battery voltage.
(c) On a set of plate characteristic curves for this tube draw a load line, through the Q you have selected, that reaches the extremes of the useful range of plate voltage at about $E_g = 0$ and at about $E_g = 3$. What load resistance is required to produce operation along such a load line?
(d) Identify on this load line the extremes of the current-voltage locus that correspond to $E_1 = 1$ volt (root-mean-square). Find the voltage gain for such operation by inspection of your diagram.

12. An RCA-59 tube, connected for use as a pentode (Fig. 60) is employed in a Class A amplifier, with a load resistance of 5,000 ohms and a 400-volt plate battery. Grid bias E_C is -15 volts.
(a) Draw a circuit diagram showing connections to control grid, screen grid, suppressor grid, and plate, as well as all batteries, resistors, etc. Use parallel feed to the plate.
(b) Draw the load line on a set of plate characteristic curves for this tube.
(c) Determine by inspection of your graph the values of I_Q and E_Q.
(d) The root-mean-square value of the excitation voltage (sinusoidal) is 7.07 volts. Spot extremes of the current-voltage locus, and determine the voltage gain by inspection of your diagram.
(e) Determine the alternating-current power output, the plate circuit efficiency, and the plate dissipation.

13. Tube and circuit properties as in the preceding problem.
(a) Plot points for and draw carefully, on a sheet of graph paper, the dynamic characteristic for this amplifier. Determine its slope at Q graphically.
(b) Plot points for and draw carefully, on the same sheet of graph paper, the mutual characteristic curve for $E_p = E_Q$, as derived from your plate characteristic set. Determine graphically the slope s_m (grid-plate transconductance) of this mutual characteristic at Q.
(c) Using Equations (525) and (491), determine, from the slopes measured in (b) and (c), values of R_p and μ for use in the equivalent circuit of this amplifier.
(d) Predict the voltage gain by the use of Equation (528), and compare with the result obtained in part (d) of the previous problem.

14. Tube and circuit properties as in the two preceding problems.
(a) Determine R_p at Q graphically, from your plate characteristic set.
(b) Plot points for and draw carefully, on a sheet of graph paper, an E_p vs. E_g curve like those in Fig. 101, for $I_p = I_Q$. Obtain μ graphically as the slope of this curve at Q.
(c) Predict the gain by the use of Equation (528), employing the results of parts (a) and (b), and compare with values for gain obtained in the two previous problems.

15. Tube and circuit as in the three preceding problems, except that an inductance whose reactance is half the value of the load resistance is introduced into the circuit.

Assume that plate current variations are sinusoidal, and that the elliptical current-voltage locus passes through the two points that were formerly the extreme points on the straight current-voltage locus.

(a) Locate on your plate characteristic sheet the following four other points on the elliptical current-voltage locus:

$$\text{Two for which } i_p = I_Q$$
$$\text{Two for which } e_p = E_Q$$

Also, draw two vertical tangents to the ellipse.
Sketch in the ellipse.

(b) Plot points for and draw carefully on a sheet of graph paper the current-voltage locus on an I_p vs. E_g graph that corresponds to the ellipse drawn in (a) for the I_p vs. E_p graph.

16. A certain amplifier's circuit is like Fig. 114a except that the positions of C and R_L are reversed, and the output voltage is that across the condenser.

(a) Draw the corresponding equivalent circuit.

(b) Draw a vector diagram for the equivalent circuit, and from it derive a complete expression for the voltage gain in terms of frequency.

CHAPTER XIII

HARMONICS; CLASS B AND PUSH-PULL AMPLIFIERS

148. Straightness of the Dynamic Characteristic a Criterion of Freedom from Harmonic Distortion. Harmonic distortion was briefly mentioned in Section 142. The requirements for freedom from it are illustrated in Fig. 115.

The grid swing in Fig. 115 does not extend beyond the straight-line portion of the dynamic characteristic. It is apparent from the graphical construction used that because the dynamic characteristic is straight the sine wave excitation voltage produces a true sine wave plate current. Furthermore, if the dynamic characteristic is straight, its slope, Equation (525), must be uniform. Therefore introduction of the sine wave relation $e_1 = \sqrt{2}E_1 \cos 2\pi ft$ into Equation (512.1) leads to a single-frequency sine wave expression for i_2. It is thus clearly evident that an amplifier is free from harmonic distortion if the current-voltage locus remains on and within the straight portion of the dynamic characteristic.

The manner in which harmonics are introduced into the output by flexion of the dynamic characteristic is illustrated in Figs. 116 and 117 and described in the associated discussions. For the present the important fact is that such flexion does produce harmonics.

The mutual characteristic curves in Fig. 104a are by no means free from flexion within the range of plate current swing. Yet the dynamic characteristic is nearly straight within that range, and the harmonic distortion is correspondingly small. Straightness of the dynamic characteristic is evidently a better criterion of freedom from harmonic distortion than is straightness of the tube characteristics.

It is easy to understand why this is true. The slope of the dynamic characteristic is $\mu/(R_L + R_p)$; that of a mutual characteristic is μ/R_p. In a triode μ is practically constant, but R_p changes, as indicated by the changing slopes of the tube characteristic curves. If, as is usually true, R_L is considerably larger than R_p, the latter may change considerably without producing a marked change in the slope of the dynamic characteristic. This explains why it is generally true in circuits employing triodes that harmonic distortion can be minimized by the use of a load resistance that is large relative to R_p.

The utility of dynamic-characteristic straightness as a criterion can be illustrated by a power amplifier circuit that contains a pentode instead

of a triode. Except for the tube used, the circuit is similar to Fig. 102a.
The pentode's plate characteristics are as illustrated in Fig. 60, page 136.

Values of μ and of R_p vary widely in Fig. 60. Yet by proper location
of the load line a dynamic characteristic that is nearly straight can be
obtained. The horizontal and vertical intercepts for such a load line
would occur at about 400 volts and 80 milliamperes respectively. An
amplifier designed to operate along such a load line could deliver con-
siderable power to the load, yet be reasonably free from harmonic
distortion.
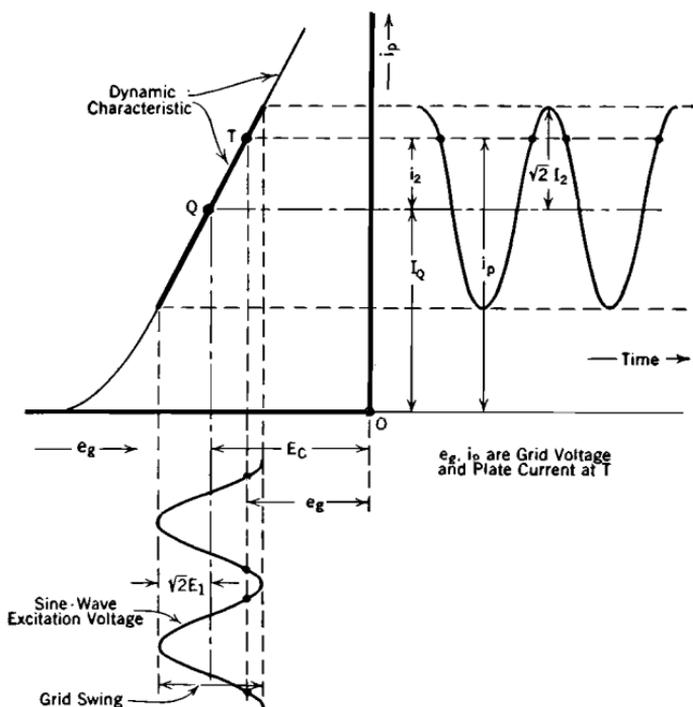
FIG. 115.   Freedom from harmonic distortion with straight-line dynamic character-
istic.   Sine-wave excitation, sine-wave i_2

If an amplifier's load circuit contains an inductance, the current-
voltage locus opens up into a loop, as illustrated in Fig. 108a. Perfect
freedom from harmonic distortion would require the mutual character-
istics to be parallel, straight, evenly spaced lines throughout the extent
of the loop. This would correspond to uniformity of both μ and R_p
throughout the loop.

However, just as when there is no loop, and for similar reasons, the
practical criterion of freedom from distortion is the degree of uniformity
of the quantity $\mu/(R_p + R_L)$ throughout the extent of the locus.   Magni-

tudes of this quantity can be studied in any particular case by drawing straight lines parallel to the load line but spaced at regular intervals from it, then drawing the corresponding lines on the mutual characteristic graph. The slope at any point on any such line is $\mu/(R_p + R_L)$.

149. Parabolic Dynamic Characteristic Introduces a Second Harmonic. Fig. 116 illustrates the response to sinusoidal excitation when the dynamic characteristic is convex downward and the load
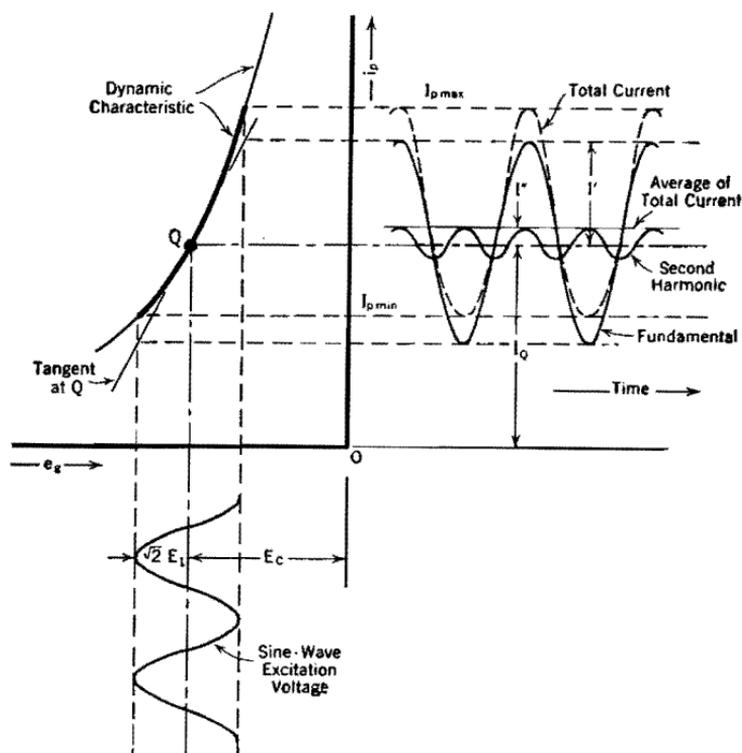
Fig. 116.   Production of second harmonic and shift of average current when dynamic characteristic is parabolic.

purely resistive.   This is the usual condition in a Class A amplifier circuit employing a triode, although often the flexion is scarcely discernible.

A curve of this type can usually be represented with reasonable accuracy over a considerable range by a parabola (second degree curve).* If the point Q of zero excitation is taken as the origin of reference, the equation of the parabola is

$$i_2 = Ae_1 + Be_1{}^2 \qquad\qquad (562\ \text{p})$$

* For the method of using a Taylor's series for i_p in the study of distortion (" plate current expansion "), see Glasgow's book.[F 337]

The slope of this parabola at Q is

$$\left(\frac{di_2}{de_1}\right)_{i_2=0} = A \qquad (563 \text{ p})$$

But the parabola is the dynamic characteristic. Therefore, according to Equation (525)

$$A = \frac{\mu}{R_p + R_L} \qquad (564 \text{ p})$$

in which μ and R_p are evaluated at Q. Thus the first term in Equation '(562) corresponds exactly to the equivalent-circuit prediction as stated by Equation (512.1). Since B is usually small, calculations based on the equivalent circuit may be expected to predict close approximations to the true values of alternating current and voltage in the load.

The current due to the second term in Equation (562) is evidently proportional to B. Differentiation shows that the flexion d^2i_2/de_1^2 of a truly parabolic characteristic is constant and equal to $2B$. This fact gives a graphical significance to B.

In order to determine the pattern of the total current when both terms are taken into account the excitation voltage must be expressed as a sinusoidal variant with time, that is, as

$$e_1 = \sqrt{2}E_1 \cos \omega t \qquad (565)$$

As usual the frequency is involved according to the relation

$$\omega = 2\pi f \qquad (566)$$

Upon entering the sinusoidal expression for e_1 into Equation (562) the following equation for the variation in plate current is obtained:

$$i_2 = \sqrt{2}AE_1 \cos \omega t + 2BE_1^2 \cos^2 \omega t \qquad (567 \text{ p})$$

If now the relation

$$\cos^2 \omega t = \tfrac{1}{2} + \tfrac{1}{2}\cos 2\omega t \qquad (568)$$

is used, the cosine-squared term in Equation (567) expands into a constant (direct-current) term plus a second harmonic (double-frequency) cosine term. The quantity A can be expressed in terms of tube and circuit constants. These operations convert Equation (567) to the following:

$$i_2 = \sqrt{2}\frac{\mu E_1}{R_p + R_L} \cos \omega t + BE_1^2 + BE_1^2 \cos 2\omega t \qquad (569 \text{ p})$$

Thus the overall results of the use of a sinusoidal excitation voltage with a parabolic dynamic characteristic are:

(1) To produce in the load a fundamental-frequency current that is exactly in accordance with the prediction of the equivalent circuit.

(2) To increase the direct current in tube and load from I_Q to $I_Q + BE_1^2$. This is of course the new average value of the total current.

(3) To introduce a second harmonic alternating current whose crest value is just the same as the increase in the direct current, that is, BE_1^2.

The various current components are shown in correct phase relationship in Fig. 116. The dotted curve represents the total current.

Let $I_{p\,max}$ and $I_{p\,min}$ represent the maximum and minimum instantaneous values of total plate current. These can be determined graphically from a diagram like Fig. 116. Let I' and I'' represent the amplitudes (crest values) of the fundamental and second harmonic currents respectively. Then, from the figure,

$$I_{p\,max} = I_Q + I' + I'' + I'' \qquad (570)$$

$$I_{p\,min} = I_Q - I' + I'' + I'' \qquad (571)$$

I'' occurs once for the second harmonic and once for the direct-current increment, in each equation. Subtraction of these equations gives

$$I' = \frac{I_{p\,max} - I_{p\,min}}{2} \qquad (572)$$

Addition of them gives

$$I'' = \frac{I_{p\,max} + I_{p\,min} - 2I_Q}{4} \qquad (573)$$

The *per cent second harmonic distortion* is of course $100 \times \dfrac{I''}{I'}$.

150. Third Harmonic Introduced by Dynamic Characteristic of Cubic Form. When a power amplifier pentode is used in a Class A amplifier, the dynamic characteristic may have a point of inflection in the working range. It cannot then be represented, even approximately, by a parabola. In such a case the true shape can usually be reasonably well approximated by a cubic equation; see Fig. 117. The equation of a dynamic characteristic of this type, with the origin taken at the operating point Q, is

$$i_2 = Ae_1 + Be_1^2 + Ce_1^3 \qquad (574\ p)$$

As with the second degree curve, A is the slope at the operating point, and is related to tube and circuit constants according to Equation (564). A is most satisfactorily evaluated in any particular case as the slope of a

tangent, at Q, to the graphically constructed dynamic characteristic curve.

When the sinusoidal expression for the excitation voltage is introduced, the first two terms of Equation (574) produce exactly the three terms

FIG. 117.   Production of second and third harmonics, also shift of average current, when dynamic characteristic is of cubic form.

I_0' is amplitude of equivalent circuit fundamental-frequency prediction. I' is actual fundamental-frequency amplitude.

that appear in Equation (569).   The term in $e_1{}^3$ becomes $2\sqrt{2}CE_1{}^3\cos^3\omega t$. This term can be expanded by means of the trigonometric relation

$$\cos^3\omega t = -\tfrac{3}{4}\cos\omega t + \tfrac{1}{4}\cos 3\omega t \qquad (575)$$

Thus the third-degree term contributes a subtraction from the fundamental, and a third harmonic term, as follows:

$$\left.\begin{array}{l}\text{Term in } e_1{}^3 \\ \text{contributes}\end{array}\right\} -\frac{3}{\sqrt{2}}CE_1{}^3\cos\omega t + \frac{1}{\sqrt{2}}CE_1{}^3\cos 3\omega t \qquad (576)$$

Thus the overall results of the use of a sinusoidal excitation voltage with a dynamic characteristic of cubic form are:

(1) To produce a fundamental-frequency current in the load that is less than that predicted by the equivalent circuit, by three times the value of the third

harmonic; see Equation (576). Therefore

$$\left.\begin{array}{l} \text{rms value of fundamental} \\ \text{component of load current} \end{array}\right\} = \frac{\mu E_1}{R_p + R_L} - \frac{3}{2} C E_1{}^3 \qquad (577\text{ p})$$

(2) To increase the direct current from I_Q to $I_Q + B E_1{}^2$.

(3) To introduce a second harmonic current component whose crest value is the same as the increase in direct current, that is, $B E_1{}^2$. If the dynamic characteristic is symmetrical with respect to Q, $B = 0$, and there is no second harmonic component.

(4) To introduce a third harmonic current component whose crest value is $C E_1{}^3 / \sqrt{2}$, root-mean-square value $\frac{1}{2} C E_1{}^3$.

The fundamental, both harmonics, direct-current increase, and total current (dotted) are all illustrated in Fig. 117. The zero axis for all components is at $i_p = I_Q$.

At the maximum value of plate current the fundamental, direct-current increment, and second harmonic are all positive, the third harmonic negative, so that

$$I_{p \max} = I_Q + I' + I'' + I'' - I''' \qquad (578)$$

At the minimum value of plate current only the fundamental is negative, the direct-current increment and both harmonics being positive, so that

$$I_{p \min} = I_Q - I' + I'' + I'' + I''' \qquad (579)$$

Addition and rearrangement of these two equations gives

$$I'' = \frac{I_{p \max} + I_{p \min} - 2 I_Q}{4} \qquad (580)$$

Thus the second harmonic and the direct-current increment are evaluated exactly as in the previous section. Subtraction of the same equations gives

$$I_{p \max} - I_{p \min} = 2 I' - 2 I''' \qquad (581)$$

This can be rearranged into the following:

$$I' = \frac{I_{p \max} - I_{p \min}}{2} + I''' \qquad (582)$$

Equation (582) shows that the fundamental component of the load current has an amplitude greater than half the plate current swing by just the amplitude of the third harmonic.

The third harmonic can be evaluated, and the true crest value of the fundamental then obtained, by use of a tangent drawn to the dynamic characteristic at the operating point. At the point H in Fig. 117 this tangent reaches the extreme positive value of grid voltage. The current I_0', *measurable graphically*, is the crest value of the fundamental that is predicted by the equivalent circuit. That is, $I_0' = \sqrt{2} A E_1$.

It has been pointed out that the actual fundamental is less than the equivalent circuit prediction by three times the third-harmonic amplitude. Therefore

$$I' = I_0' - 3I''' \tag{583}$$

This relation can be introduced into Equation (581), and a solution for I''' obtained in the following form:

$$I''' = \frac{2I_0' - (I_{p\,max} - I_{p\,min})}{8} \tag{584}$$

On using Equation (584) in Equation (582) it appears that

$$I' = \frac{3(I_{p\,max} - I_{p\,min}) + 2I_0'}{8} \tag{585}$$

It has just been shown how to determine entirely by graphical means the amplitudes of the fundamental, direct-current increment, and second and third harmonics for a dynamic characteristic that is best approximated by a cubic curve. Ordinarily the existence of the third harmonic is of more consequence than the modification in value of fundamental that goes along with it. Considerable reverse curvature is required to produce an appreciable third harmonic; thus the very pronounced inflection in the curve of Fig. 117 produces only a $6\frac{1}{4}$-per cent third harmonic.

It is sometimes useful to express the combined effects of the various harmonics as follows:

$$\left.\begin{array}{l}\text{Per cent total}\\\text{harmonic distortion}\end{array}\right\} = \frac{\sqrt{I''^2 + I'''^2 + I''''^2 + \cdots}}{I'} \times 100 \tag{586}$$

Ordinarily harmonics higher than the second and third are so small as to escape attention.

151. Class B Amplifiers. Up to this point in the text attention has been confined to the Class A type of amplifier operation. As pointed out in Section 145, this type of operation results in low plate circuit efficiency. The name " Class B amplifier " is applied to circuits in which the grid bias and excitation are handled in a particular way that permits reasonably high plate circuit efficiency.

The essential distinguishing feature of a Class B amplifier is that it is *biased to cut-off*, or approximately so. The meaning of this phrase is illustrated by Fig. 118. The grid bias E_C in this figure extends approximately to the lower end or cut-off point of the dynamic characteristic.

In contrast with this, any Class A amplifier is biased to some voltage considerably above cut-off. Very often the bias for Class A operation

is chosen so as to locate the point Q about half-way between zero grid volts and cut-off. (See Figs. 104*a*, 108*a*, 115, 116, and 117.) A Class C amplifier is one that is biased to considerably *below* cut-off. In Class C operation plate current flows only during portions of the positive half-cycles of the grid excitation voltage.[F 289, I 244, H·XVII]

In typical Class B operation, illustrated by Fig. 118:

(*a*) The plate current consists of half-wave impulses. To the extent that the dynamic characteristic approaches a straight line with a sharp turn at cut-off, the plate current impulses approach true half sine waves, sometimes called *sine loops*.
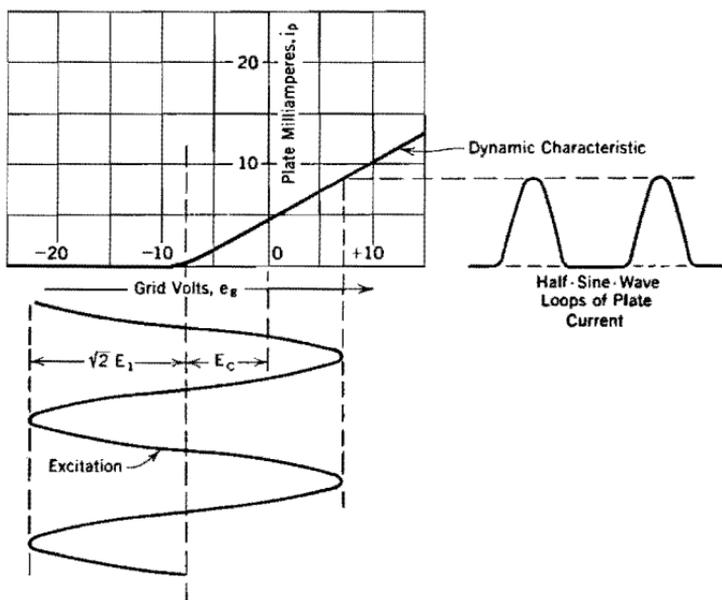
Fig. 118. Class B amplifier operation. Grid is biased to cut-off.

(*b*) Grid current is to be expected during a part of each positive grid swing. For this reason the source of grid excitation must have an internal impedance low relative to that between filament and positive grid. If the source of excitation voltage has a high internal impedance, the flow of grid current causes an internal voltage drop which flattens the positive wave crests of the excitation voltage. This distortion of excitation voltage is of course amplified and appears in the output.

(*c*) The crest values, also the root-mean-square and average values, of the plate current loops are proportional to the excitation voltage if the dynamic characteristic is a straight line. Because of this *linear* relation between the excitation voltage and the plate current, such an amplifier is sometimes called a *linear* amplifier.

(*d*) The variations in plate current are much larger in proportion to its average steady value than is true for Class A amplifiers. Since these variations provide the useful output, and the average current measures the plate circuit power demand, the plate circuit efficiency is much higher than for Class A amplifiers. Furthermore, the existence of a large alternating component of plate current tends to *reduce* the plate dissipation (see Section 145).

(*e*) Plate current variations of the type indicated in Fig. 118, though possessing the merit of linearity, are not directly serviceable because they are very far from being full sine wave variations. The sine loop type of plate current contains very pronounced harmonics.

These sine-loop impulses can be made useful either by (1) employing a *push-pull* circuit which matches each half sine wave with its mate of opposite polarity, thus producing a full sine wave reasonably free from harmonics, or by (2) using a tuned plate circuit which eliminates the harmonics. Class B amplifiers of both types are in common use. The push-pull type is usually employed with audio-frequency excitation, the tuned type with modulated radio-frequency excitation. The push-pull type is discussed in the next section. Discussions of Class B tuned ("linear") amplifiers are given by Terman,[I 210] Glasgow,[F 287] and Everitt.[H XVII]

152. Class B Push-Pull Amplifiers. Fig. 119 is a circuit diagram, and Fig. 120 a plate current-grid voltage diagram, for a push-pull Class B amplifier.[F 195] The plate circuits are coupled to the load by means of a transformer. An analysis of transformer-coupling principles appears later, in Sections 159 to 161. In the present discussion it is only necessary to bear in mind that a transformer's voltage ratio is the same as its turn ratio, and that primary ampere-turns must equal secondary ampere-turns.

The excitation voltage is introduced through a center-tapped transformer so that one grid swings positive while the other swings negative. In Fig. 120 grid voltages for one tube are plotted positive toward the right, those for the other positive toward the left, so that a common sine wave can be used for illustrating the grid excitations. The plate-current coordinate directions are similarly reversed. The grid bias is the same for both tubes.
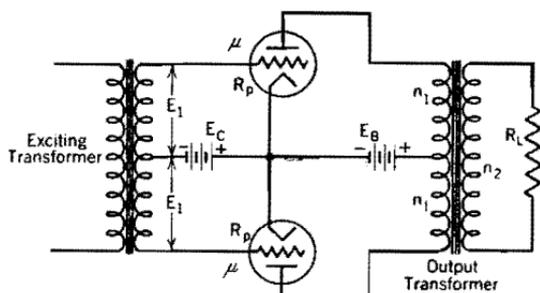
Fig. 119.   Push-pull amplifier circuit.

The pronounced fillets in the dynamic characteristics near cut-off make each tube produce current loops that are decidedly widened out at the bottom. However, the transformer coupling into the load gives the two tube currents opposite effects on current in the load, so that the fillet currents tend to cancel one another. A resultant or composite dynamic characteristic, whose plate current is everywhere the algebraic sum of the plate currents in the two tubes, can be used to predict the load-current wave form. In Fig. 120 this composite characteristic consists of the straight slant portions of the two individual characteristics, plus a short dotted extension across the gap between them. If the composite dynamic characteristic is a straight line, the current in the load is free from harmonics.
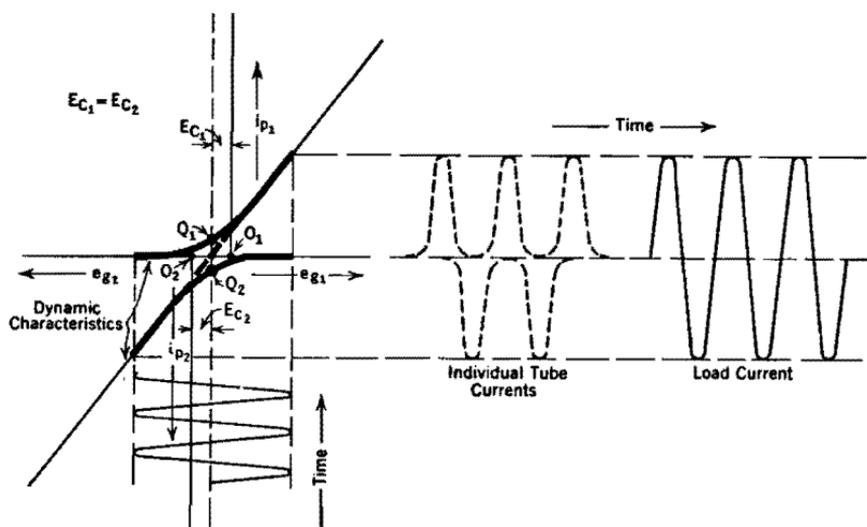
FIG. 120. Dynamic characteristics for Class B push-pull amplifier operation.

In Figs. 115, 116, and 117 the zero reference axis for the various alternating-current components of load current is a horizontal line through Q. The origin of the equations used for approximating the shapes of the dynamic characteristics in these figures was taken at Q, which lies on this reference axis.

In Fig. 120 the corresponding reference axis is the common zero-plate-current axis. The origin of an approximating equation should therefore be taken along this axis, just midway between Q_1 and Q_2. The composite dynamic characteristic, whether straight or curved, is necessarily symmetrical about that point, by virtue of the manner of its construction. Therefore it can be approximated by a cubic equation that has no second-degree term, so introduces no second-harmonic current term. This

explains̄why the output of a push-pull amplifier is free from second-harmonic distortion. In fact, all even harmonics are eliminated in this way by such a circuit. Only odd harmonics (third, fifth, seventh, etc.) can appear in the load of a circuit of the type diagrammed in Fig. 119.

Proper choice of the grid bias is essential in order to make the third harmonic a minimum. Tubes for push-pull Class B service are often designed so that with normal load resistances the straight portion of each one's dynamic characteristic points directly toward the origin, making it unnecessary to use any grid bias at all. The correct grid bias is then zero, and grid current flows all the time either in one tube or the other.

153. Dynamic Characteristics and Equivalent Circuits for Push-Pull Amplifiers. The push-pull circuit illustrated in Fig. 119 can be used for either Class A[F 184] or Class B operation, or for an intermediate type of service called Class AB operation.[F 200, H XVII] In each case even harmonics are eliminated. Furthermore, the steady components of plate current oppose each other magnetically in the transformer core. This eliminates any tendency toward core saturation and consequent frequency distortion due to low primary inductance (see Section 160).

A plate current-grid voltage diagram for a Class A push-pull amplifier appears in Fig. 121. The Class A nature of the operation is indicated by the facts that the tubes are not biased to cut-off and the grids do not swing positive. The composite dynamic characteristic represents, as before, the algebraic sum of the two plate currents for a given grid voltage. As before, its symmetry permits only odd harmonics to exist in the load circuit. In this case the individual-tube plate current variations do not extend downward to the zero current axis.

Class A push-pull amplifier operation can be satisfactorily analyzed by means of the equivalent circuit diagram shown in Fig. 122a. The load resistance R_L in this diagram is of course a perfectly definite quantity, being just that used in the actual circuit. However, the two tubes feed through a common transformer core into a common load resistance. For this reason a little study is necessary to determine just how the load affects tube operation, that is, what value of load resistance should be used in Equation (525) to determine the slope of the dynamic characteristics at the Q's.[85]

In a Class A push-pull amplifier neither tube's current-voltage locus extends appreciably beyond the region in which the tube characteristic curves are approximately straight. Therefore the plate resistances of the two tubes are equal and approximately uniform throughout each cycle. The primary circuit in Fig. 122a, therefore, has identical values of R_p in its upper and lower halves.

The middle connection in this diagram is shown dotted because,

although it actually exists, the tubes produce equal and opposite funda-mental-frequency currents through it; hence its net fundamental-frequency current is zero. It may therefore be omitted in the analysis of fundamental-frequency Class A operation.
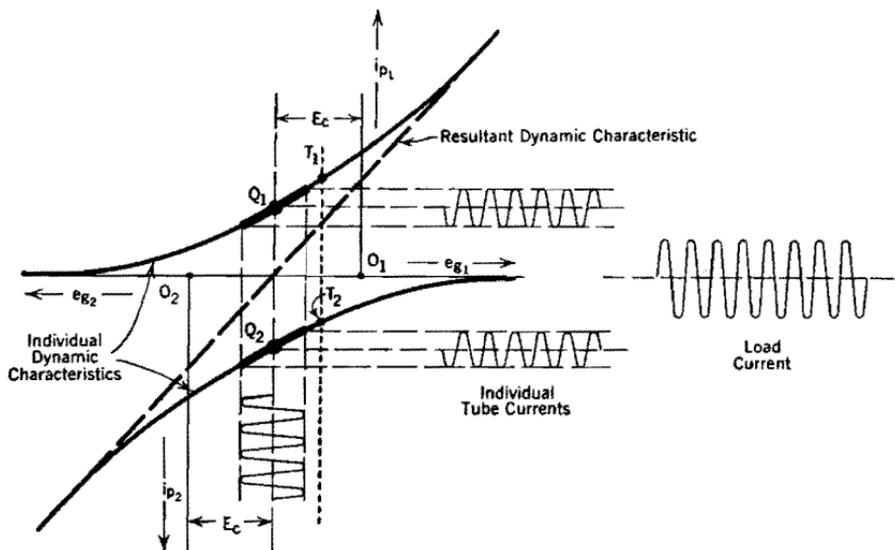
Fig. 121.   Dynamic characteristics for push-pull Class A amplifier operation.

(a) Class A operation.   (b) Class B operation.
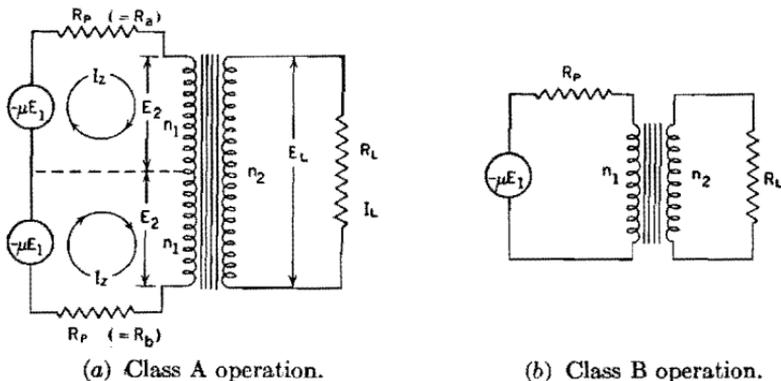
Fig. 122.   Equivalent circuit diagrams for push-pull amplifiers.

With the middle (dotted-line) connection omitted, and the two excita-tion voltages, as well as values of μ and R_p, alike, it is evident that

$$-2\mu E_1 = 2I_Z R_p + 2E_2 \qquad (587\ p)$$

The usual simple transformer relations require that

$$2E_2 = \frac{2n_1}{n_2} I_L R_L \quad \text{and} \quad n_2 I_L = 2n_1 I_Z \qquad (588\ p)$$

An expression for the gain E_L/E_1 is easily derived by using these in Equation (587). For the purpose of determining how the load affects tube operation Equations (588) are combined to give an expression for E_2, as follows:

$$2E_2 = I_Z\left(\frac{2n_1}{n_2}\right)^2 R_L \tag{589 p}$$

The use of Equation (589) converts Equation (587) into

$$-2\mu E_1 = 2I_Z R_p + I_Z\left(\frac{2n_1}{n_2}\right)^2 R_L \tag{590 p}$$

After substituting $-I_2$ for I_Z, this can be written

$$\frac{I_2}{E_1} = \frac{\mu}{R_p + 2\left(\dfrac{n_1}{n_2}\right)^2 R_L} \tag{591 p}$$

Since both dynamic characteristics are straight or nearly so within the range of grid swing, their slopes are uniform and equal to i_2/e_1, so to I_2/E_1. This fact is expressed mathematically as follows:

$$\frac{di_p}{de_g} = \frac{\mu}{R_p + 2\left(\dfrac{n_1}{n_2}\right)^2 R_L} \tag{592 p}$$

Comparison of Equation (592) with Equation (525) shows that the dynamic characteristics must have at the Q's the same slope as though each tube were in a single-tube Class A amplifier, with a load resistance of $2R_L(n_1/n_2)^2$ ohms. Thus in push-pull *Class A operation* each tube may be said to be "working into" a resistance of the magnitude just stated, while the "plate-to-plate" resistance is twice this value, that is, $4R_L(n_1/n_2)^2$ ohms.

In a *Class B* push-pull amplifier current flows through the two tubes alternately, each one being an open circuit while the other carries current. The two circuit halves in Fig. 122a then carry current in alternate half-cycles, the middle dotted connection completing the circuit for both. Since the two halves are identical and work alternately, they can be satisfactorily represented by either one working alone continuously, as in the Class B push-pull equivalent circuit diagram shown in Fig. 122b. Calculations based on this circuit can be used to predict the gain.

The use in connection with Fig. 122b of relations similar to Equations (588) shows that in Class B operation each tube is "working into" a resistance of $R_L(n_1/n_2)^2$ ohms, just half the value found for Class A operation. The "plate-to-plate" resistance is still $4R_L(n_1/n_2)^2$.

The major (straight slant) parts of the individual dynamic characteristics in Fig. 120 have slopes that are described by using $R_L(n_1/n_2)^2$ for the load resistance in Equation (525). Neither the load lines nor the dynamic characteristics are, over these major straight portions, directly in line with the points Q on their respective diagrams.*

Fig. 121 may satisfactorily be thought of as a detail, to an enlarged scale, of a small portion of Fig. 120. Thus in the immediate neighborhood of the Q's the two diagrams are identical in nature. For example, Equation (592) gives the slopes of the dynamic characteristics at the Q's for both Class A and Class B operation; but in one case this slope is important, in the other case quite unimportant.

According to this viewpoint an essential difference between the two kinds of operation is that one employs a relatively large, the other a relatively small, grid swing. Obviously the grid swing can be increased gradually. These questions arise: how rapidly does the two-to-one change in effective load resistance take place, and how can conditions intermediate between the extremes be analyzed? These questions are especially pertinent for Class AB operation.

The problems thus presented can be satisfactorily approached by studying the response of the equivalent circuit of Fig. 122a to small-amplitude, local operation about a steady grid position such as indicated by the dotted line $T_1 T_2$, Fig. 121. Imagine that the normal excitation voltage is arrested for a brief interval at the $T_1 T_2$ value, and an extremely small investigational excitation E_1' imposed locally. It will be presumed that the tube characteristics are curved, so that the local values of R_p for the two tubes are unlike. The plate resistances will be symbolized as R_a and R_b, located in the upper and lower halves of the equivalent circuit, Fig. 122a, respectively.

Evidently the two circuit halves are not now symmetrical, and the middle connection must carry the unbalance current. Let I_a' represent the current in the upper half, I_b' that in the lower half, and I_L' that in the load resistance. Magnetic coupling forces the two halves of the transformer primary to have identical voltages, called E_2'. Voltage across the load resistance is E_L'. Note that E_1' is very small indeed, and that it swings about the grid voltage at $T_1 T_2$, not about that at the Q's. Transformer action in Fig. 122a requires that

$$n_2 I_L' = n_1 I_a' + n_1 I_b' \qquad (593\ \mathrm{p})$$

* The load line in Thompson's useful method of graphical analysis of push-pull operation[35] is drawn with a slope corresponding to $R_L(n_1/n_2)^2$, the resistance that each tube works into in Class B operation. His curve descriptive of effective operating conditions for one tube has the slope $1/2R_L(n_1/n_2)^2$ at passage through the Class A operating point, and is asymptotic to the straight Class B load line of slope $1/R_L(n_1/n_2)^2$.

Therefore

$$E_L' = R_L I_L' = R_L \frac{n_1}{n_2} (I_a' + I_b') \qquad (594 \text{ p})$$

Also

$$E_2' = \frac{n_1}{n_2} E_L' = (I_a' + I_b') \left(\frac{n_1}{n_2}\right)^2 R_L \qquad (595 \text{ p})$$

and

$$-\mu E_1' = I_a' R_a + E_2' \qquad (596 \text{ p})$$

Combination of the last two equations leads to the following result:

$$-\mu_1 E' = I_a' \left[R_a + \left(\frac{n_1}{n_2}\right)^2 R_L \right] + I_b' \left(\frac{n_1}{n_2}\right)^2 R_L \qquad (597 \text{ p})$$

The mate to this, obtained from the lower half of the circuit, is

$$-\mu E_1 = I_a' \left(\frac{n_1}{n_2}\right)^2 R_L + I_b' \left[R_b + \left(\frac{n_1}{n_2}\right)^2 R_L \right] \qquad (598 \text{ p})$$

These equations can be simultaneously solved for either I_a' or I_b'. The solution for I_a' can be expressed in terms of the ratio I_a'/E_1'. This is also di_p/de_g for the proper tube, because of the smallness of the investigating-current amplitude. The local slope of the dynamic characteristic at T_1 is found in this way to be

$$\frac{di_p}{de_g} = \frac{\mu}{R_a + \left(1 + \frac{R_a}{R_b}\right)\left(\frac{n_1}{n_2}\right)^2 R_L} \qquad (599 \text{ p})$$

This indicates that at each momentary value of grid voltage the upper tube operates as though working into a sort of apparent load resistance, of magnitude

$$\left(1 + \frac{R_a}{R_b}\right) R_L \frac{n_1^2}{n_2^2} \qquad (600 \text{ p})$$

The corresponding quantity for the lower tube contains R_b/R_a instead of R_a/R_b.

In a Class A push-pull amplifier, $R_a = R_b$ over the entire working range, so that the apparent load resistance is constant and equal to $2R_L (n_1/n_2)^2$, as already shown in Equation (591). Over almost all the working range of a Class B amplifier the plate resistance of either one tube or the other is infinite. While the R_a circuit is working, R_b is infinite, reducing the apparent load to $R_L (n_1/n_2)^2$, which is in accordance with the Class B analysis already given.

PROBLEMS

Chapter XIII

1. The dynamic characteristic of a certain Class A amplifier is represented satisfactorily within the operating range by a parabola whose vertex is at $I_p = 2$ milliamperes, $E_g = -18$ volts, and passes through the point $I_p = 10$ milliamperes, $E_g = 0$. Grid bias E_C is -9 volts.

(a) Write the equation of this parabola, using the point Q of zero excitation as the origin of your i_2, e_2 coordinates.

(b) An excitation voltage $E_1 = 4$ volts (root-mean-square) and frequency f is used. Find the per cent second-harmonic distortion, the steady component of current in the plate circuit, and the root-mean-square value of the fundamental-frequency component of plate current.

2. Using the I_p, E_g, current-voltage locus of part (c), Problem 1, Chapter XII, determine the per cent second-harmonic distortion for the amplifier there described.

3. Find the per cent second-harmonic distortion for operation as in part (c), Problem 8, Chapter XII.

4. Find the per cent second-harmonic distortion when $E_1 = 1$ volt, in Problem 11, Chapter XII.

5. Suppose that in a Class A amplifier employing an RCA-59 pentode (Fig. 60) the following circuit values apply: $E_C = -20$ volts, $E_Q = 160$ volts, $R_L = 8{,}000$ ohms.

(a) Determine the root-mean-square value of the fundamental component of the output voltage, and of the second- and third-harmonic components, by the method discussed in connection with Fig. 117.

6. Tube and circuit as in Problems 12, 13, and 14 of Chapter XII.

(a) Find the amplitude of the fundamental component of plate current, also the per cent second-harmonic distortion and the per cent third-harmonic distortion, by the method outlined in connection with Fig. 117.

(b) Using the amplitude of the fundamental component of current obtained in (a), predict the voltage gain, and compare with the values predicted in the three earlier problems.

7. Two tubes with characteristics as illustrated in Fig. 104 are used in a circuit like that diagrammed in Fig. 119. As in Fig. 104, $E_B = 400$ volts; however, $E_C = -95$ volts. $R_L = 400$ ohms, $n_2/n_1 = 0.4$; $E_1 = 60$ volts, root-mean-square.

(a) Does this choice of magnitudes result in Class A or in Class B operation?

(b) Draw the proper equivalent circuit, labeling correctly the magnitudes of the various circuit elements.

(c) Draw the composite dynamic characteristic, and determine from it the per cent second-harmonic distortion, also the per cent third-harmonic distortion.

(d) Determine the voltage gain, the plate circuit efficiency, and the plate dissipation.

8. Conditions as in the preceding problem, except that $E_C = -40$ volts, and $E_1 = 10$ volts, root-mean-square.

(a) Does this choice of magnitudes result in Class A or in Class B operation?

(b) Draw the equivalent circuit and label magnitudes of all circuit elements in it.

(c) Draw the composite dynamic characteristic.

(d) Determine the voltage gain, the plate circuit efficiency, and the plate dissipation.

9. Fig. *A* illustrates a *modulated* alternating voltage (modulation is not discussed in this book). A modulated voltage has the following properties: it alternates sinusoidally at a high frequency f_1, called the *carrier frequency*, but the peak values of the alternations vary as follows:

$E_{max} = E_a + E_b \cos 2\pi f_2 t$

Here f_2 is the frequency of modulation, and is very much less than f_1.

FIG. *A*.   Audio modulation of carrier wave.

(*a*) Write the complete equation describing the variation, with time, of the instantaneous value of the modulated voltage.

(*b*) Expand this equation trigonometrically into three cosine terms, which will be: (1) a carrier-frequency term in $\cos 2\pi f_1 t$, (2) a lower side-band term in $\cos 2\pi (f_1 - f_2)$ and (3) an upper side-band term in $\cos 2\pi (f_1 + f_2)t$.

(*c*) A modulated voltage, for which $E_a = 16$, $E_b = 10$, $f_1 = 2,000,000$, $f_2 = 1,500$, is the excitation for a Class B amplifier that is used as a detector, as follows: the circuit is arranged as in Fig. 102*a*, and the dynamic characteristic is practically a straight line, as in Fig. 118. Along this straight line

$E_g = -30$ volts when $I_p = 0$
$I_p = 12$ milliamperes when $E_g = 0$

A grid bias of -30 volts is used. Write the equation for the trace that would be made by a magnetic oscillograph carrying the plate current, if the oscillograph is only capable of following current variations at frequencies less than 5,000 cycles per sec. (The oscillograph serves as a mechanical filter.)

AMPLIFIER COUPLING; OSCILLATORS

154. Cascading of Amplifiers; Voltage Gain and Decibel Gain. It is frequently necessary to use more than one amplifier, or, according to the more common phraseology, more than one stage of amplification. In multistage amplifiers the output voltage of the first stage is the input voltage (excitation) of the second stage, the output of the second stage the input of the third, and so on. The overall voltage gain of a multistage amplifier is of course the *product* of the gains of the individual stages.

The term "decibel gain," often abbreviated to "db gain," is frequently used in communication engineering work. It is a measure of the *logarithm* of the gain, so that the overall decibel gain of a multistage amplifier is the *sum* of the decibel gains of the individual stages. The basic definition of the decibel gain of an amplifier is

$$\text{db gain} = 10 \log_{10} \frac{\text{output watts}}{\text{input watts}} \tag{601}$$

db gain is thus originally defined as a measure of *power gain* rather than of voltage gain.

This seems a strange statement, for one of the useful properties of an electronic amplifier is that no power is required to vary the grid potential. Thus the power input to the grid is zero; yet according to the definition just stated the decibel gain of an amplifier measures the ratio of power output to power input.

The explanation of this apparent contradiction is to be found in current telephone engineering practice. Long distance telephone conversations are made possible by amplifiers, called "repeaters," located at appropriate distances. Each section of telephone transmission line is connected at] its far end to ("terminates in") a resistance. The line current of course flows through this resistance. The resulting voltage drop across the resistance is the excitation voltage for a repeater. This repeater's load is the next section of transmission line, which terminates in the *input resistance* of another repeater, whose load is the next section of transmission line, and so on.

This practice has given rise to the telephone engineer's habit of

specifying the input and output resistances R_1 and R_2 (the latter heretofore called R_L) for which any given amplifier is suited. The watts input and watts output in Equation (601) refer to the power in these resistances.

Thus decibel gain can also be expressed as follows:

$$\text{db gain} = 10 \log_{10} \frac{\dfrac{E_2{}^2}{R_2}}{\dfrac{E_1{}^2}{R_1}} = 20 \log_{10} \frac{E_2}{E_1} \sqrt{\frac{R_1}{R_2}} \qquad (602 \text{ p})$$

If $R_1 = R_2$, as is often true, Equation (602) simplifies to

$$\text{db (voltage) gain} = 20 \log_{10} \frac{E_2}{E_1} \qquad (603)$$

Only if the input and output resistances are equal does Equation (603) give the gain *in decibels as defined by Equation* (601). Equation (603) is, however, sometimes used as a definition of the *decibel voltage gain* of an amplifier, *without regard to the input or output resistances or the power in them.* This usage has grown out of the need for some standard logarithmic measure of voltage gain.

It is important to note that Equations (601) and (603) may in general give distinctly different results as applied to a given amplifier, each result having a useful meaning if properly interpreted. One is a logarithmic measure of a power ratio, the other of the voltage gain.

155. Direct Coupling between Stages. Each stage of amplification must in general be treated as a distinct unit, with its own real and its own equivalent circuit. However, some provision must always be made for *coupling* the alternating output voltage of one stage into the grid circuit of the next. It is furthermore desirable to do this in such a way as to permit use of a common direct-current source of plate current for all tubes, and of a common source of filament heating current.

Fig. 123 illustrates a multistage amplifier circuit which answers these requirements. Inspection of the circuit shows that the alternating part of the voltage across each load resistor is also the alternating component of the grid voltage for the succeeding tube. This amplifier contains neither chokes nor condensers; only tubes, batteries, and resistors occur in it. It is consequently suitable for amplifying transient voltage impulses as well as steady alternating-current voltages, in accordance with the discussion in Section 147.

The circuit illustrated in Fig. 123 has serious disadvantages. In order to establish proper grid bias voltages, between-stage grid batteries must be used whose voltages are of the order of magnitude of the plate

battery supply.   In spite of the fact that these batteries can be of an inexpensive type because they need carry no current, they add undesirable bulk and expense.   They may also sometimes have harmful capacitance effects.

Furthermore, each grid bias voltage is dependent on the steady plate current of the preceding stage, which depends in turn on the preceding grid bias, which depends on the preceding plate current, and so on. Readjustment of bias, plate voltage, or load resistance of any stage necessitates readjustments of grid battery voltages in all succeeding
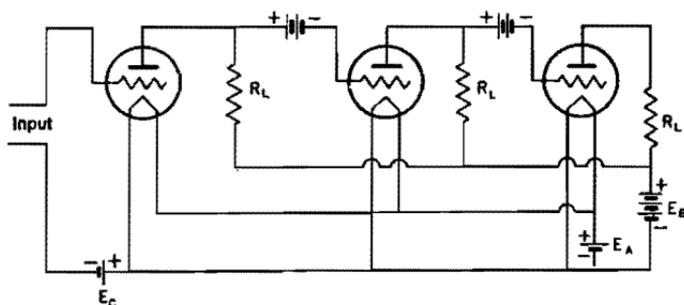
Fig. 123.   Three-stage amplifier without condensers or chokes, suitable for amplifying direct-current, transient, or alternating-current voltages.

stages.   The bias voltages are very likely to change slowly because of various minor effects such as slow variations in the temperatures of batteries, resistors, etc.   Any such changes·in the earlier stages are amplified and produce undesirable " drift " of the grid bias voltages of the later stages and of the voltage across the load resistor.

The foregoing discussion has illustrated the general nature of the circuit design problem of determining the best way to obtain proper coupling, yet maintain the proper grid bias at each stage with reasonable economy of equipment.

156. Transformer and Condenser Coupling.   The coupling problem just referred to becomes a relatively simple one when the excitation voltage is alternating.   Then the output voltage from one stage can be transmitted to the succeeding grid through a condenser or a transformer.   Either device provides the necessary direct-current insulation, so that the " bucking-out " grid batteries employed in the circuit of Fig. 123 are not required.   Fig. 124 illustrates the use of transformers, Fig. 125 that of condensers, for coupling between stages.   Filament supply batteries have been omitted from the diagrams for the sake of simplicity.   Both arrangements illustrated permit the use of common plate, grid, and filament batteries.

Load resistors are not needed in the plate circuits of the first two

stages in Fig. 124 because each transformer has a high alternating-current impedance, which serves the purpose of a load resistance. In fact, at normal frequencies these first two stages have the equivalent of infinite load resistances. Their load lines are horizontal lines through the operating points.

When coupling is accomplished by means of condensers, as in Fig. 125, the resistances marked R_L may be replaced by chokes. Then the R_G's serve as load resistances. The circuit for each stage is then identical with the parallel-feed amplifier arrangement illustrated in Fig. 114a.
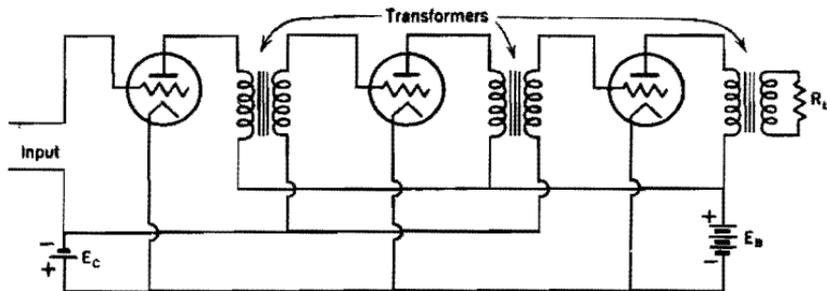
FIG. 124. Three-stage amplifier with transformer coupling, suitable for amplifying alternating-current voltages.
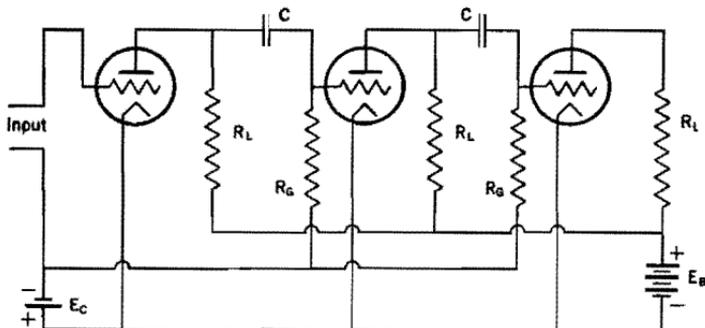
FIG. 125. Three-stage resistance-coupled amplifier with blocking condensers, suitable for amplifying alternating-current voltages.

In an actual multistage amplifier an open circuit exists between grid and filament in each tube. Each amplifier circuit is separated from those adjacent to it by these open circuits. This fact necessitates completely separate handling of the successive equivalent circuits. The gain for each is calculated independently of the others, the total gain being the product of the individual gains. The total phase angle shift is the sum of the individual phase shifts.

At very high frequencies the grid-filament impedance is not infinite,

because of interelectrode tube capacitances. Furthermore, the capaci-
tances between the plate and other tube electrodes in a triode introduce
disturbing capacitance coupling between adjacent stages. By the use
of screen-grid tubes, and of electrostatic shielding between circuits out-
side the tubes, the alternating electric fields of the various stages can be
successfully separated from one another even at high frequencies.

157. Voltage vs. Power Amplification. All stages of a multistage
amplifier except the final one serve primarily as voltage amplifiers.
However, the final stage must in many cases be a power amplifier (see
Section 143 for the distinction between the two kinds).

Of course every amplifier exists to serve some definite purpose. That
purpose determines the nature and to a considerable extent the resistance
and reactance of the load or " burden " of the final stage.   The burden
may be a voice-reproducing device, an electromagnetic relay, an oscillo-
graph element, or any one of many other devices.   Some burdens require
alternating current with but little voltage, some voltage with but little
current; and some require both alternating current and voltage, with
consequent power consumption.

A power amplifier serving a low-impedance burden must employ
a tube having a low plate resistance, in accordance with the discussion
in Section 143.   For this reason power amplifier tubes having rela-
tively low plate resistances are commonly used in final-stage power
amplifiers.   But low plate resistance is usually not compatible with a
high amplification factor, so that the voltage gain of the last stage may
be very moderate, possibly even less than one.   This is not a serious
fault, because voltage gain may be obtained in preceding stages of
voltage amplification.

158. Resistance-Condenser Coupling. Figs. 126a and 126b are
diagrams of the actual and equivalent circuits of a stage of amplification
that is coupled to the grid of the next stage by means of a condenser
and resistances.   The resistance R_G in the grid circuit of the succeeding
stage is large relative to R_L.   It is therefore reasonable to assume
that at all frequencies the alternating current through R_G and C is neg-
ligible relative to that through R_p.   Subject to this approximation the
following familiar equation expresses the relation between E_2 and E_1,
for any frequency:

$$\frac{E_2}{E_1} = \frac{\mu}{1 + \dfrac{R_p}{R_L}} \qquad (604 \text{ p})$$

The actual gain E_3/E_1 is sharply affected by frequency, because the
voltage E_2 splits vectorially between the resistance and the condenser

in proportion to their impedances, so that

$$\frac{E_3}{E_2} = \frac{R_G}{\sqrt{R_G^2 + X_C^2}} = \frac{1}{\sqrt{1 + \dfrac{X_C^2}{R_G^2}}} \qquad (605 \text{ p})$$

The overall gain is therefore approximately

$$\frac{E_3}{E_1} = \frac{E_3}{E_2}\frac{E_2}{E_1} \cong \frac{\mu}{\left(1 + \dfrac{R_p}{R_L}\right)} \frac{1}{\sqrt{1 + \dfrac{X_C^2}{R_G^2}}} \qquad (606 \text{ p})$$

The complete vector diagram for the equivalent circuit is shown in Fig. 126c. An exact expression for the gain can be obtained from this

(a) Circuit diagram.   (b) Equivalent circuit diagram.
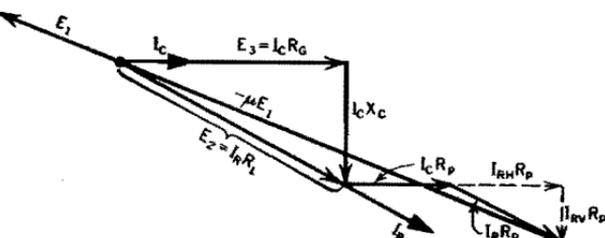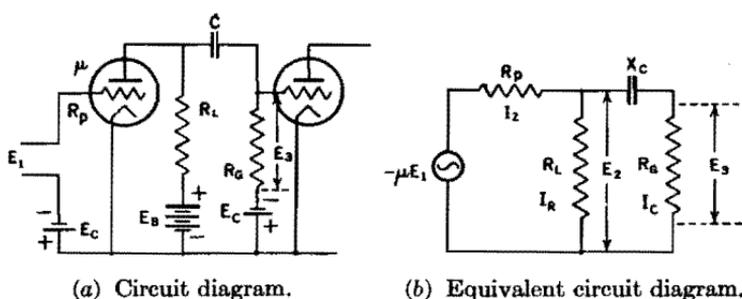
(c) Vector diagram for equivalent circuit.

FIG. 126.   Resistance coupling, with blocking condenser.

vector diagram by a process that closely parallels the derivation of Equation (560) for the parallel-feed amplifier. The final result is

$$\frac{E_3}{E_1} = \frac{\mu}{\sqrt{\left[1 + \dfrac{R_p}{R_L} + \dfrac{R_p}{R_G}\right]^2 + \dfrac{X_C^2}{R_G^2}\left[1 + \dfrac{R_p}{R_L}\right]^2}} \qquad (607 \text{ p})$$

The two terms in the radical are proportional to the components of $-\mu E_1$ that are respectively in phase with and at right angles to E_3.

This permits ready determination of either the sine or the cosine of the angle ϕ between E_3 and E_1.

If $R_G = 20R_L$ and $R_L = 2R_p$ the approximation, Equation (606), predicts a gain that is too small by $1\frac{2}{3}$ per cent at the higher frequencies. The error is less at low frequencies. Of course it becomes still less if the ratio R_G/R_p is made larger.

159. Transformer Coupling, Infinite Output Resistance. Fig. 127a and the solid lines of Fig. 127b illustrate the actual and equivalent circuits of a stage of amplification which is coupled to the grid of the next tube by means of a transformer. n symbolizes the transformer's turn ratio. The direct-current part of the plate current passes through the primary winding of the transformer, which has a negligible direct-current resistance. The grid bias voltage for the next stage is applied through the secondary winding.

The secondary of the transformer is open-circuited, so that the primary winding affects the plate behavior exactly as a simple inductance would. It is therefore represented in the equivalent circuit by the inductance L_P, having reactance X_P ohms. It should be noted that X_P is the "primary reactance," corresponding to the ratio

$$\frac{\text{primary volts}}{\text{magnetizing-current amperes}}$$

of an ordinary transformer. It must not be confused with the transformer's leakage reactance. In electric power supply circuits leakage reactance is important because of its effect on a transformer's voltage regulation, but in the usual amplifier circuit leakage reactance exerts only a minor influence.

(a) Actual circuit diagram.   (b) Equivalent circuit diagram.
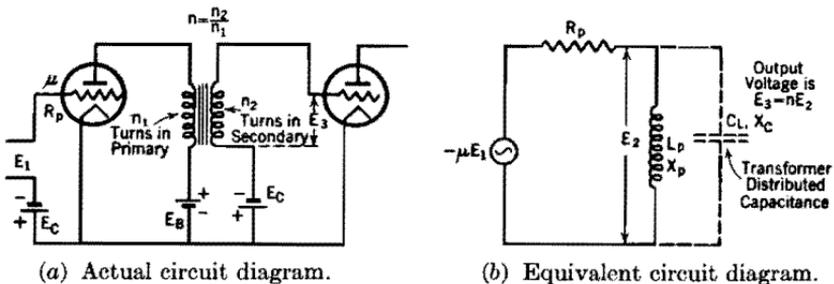
Fig. 127.   Transformer-coupled amplifier with infinite load impedance.   Transformer turn ratio $= n = n_2/n_1$.

The statement $E_3 = nE_2$ in Fig. 127b is an essential part of the equivalent circuit concept. The actual output voltage is taken from the transformer secondary, which is omitted from the diagram for the sake of simplicity.

The expression for gain corresponding to Fig. 127b is of course very simply derived, thus

$$\frac{E_3}{E_1} = \frac{nE_2}{E_1} = \frac{n\mu X_P}{\sqrt{R_p{}^2 + X_P{}^2}} = \frac{n\mu}{\sqrt{1 + \dfrac{R_p{}^2}{X_P{}^2}}} \tag{608 p}$$

If the frequency becomes large enough so that $R_p/X_P \ll 1$, the gain approaches $n\mu$; if the frequency approaches zero, X_P and the gain both vanish.

Equation (608) apparently indicates that the gain can be indefinitely magnified by increasing the turn ratio $n = n_2/n_1$ of the transformer, but this indication is false, for reasons that will soon be apparent. There is also, however, a correct indication that in order to permit satisfactory operation at low musical frequencies (down toward 20 cycles per second) the primary inductance L_P must be made rather large, often 20 to 50 henrys. The primary inductance is related to the geometry and material of the transformer as follows:

$$L_P = n_1{}^2 \left(\frac{d\Phi}{dn_1 i_p} \right) \times 10^{-8} \tag{609 p}$$

Here $d\Phi/dn_1 i_p$ is the slope of the magnetization curve for the iron core of the transformer. The primary inductance can be made large either by making $d\Phi/dn_1 i_p$ large, which requires bulk, weight, and cost in the iron core, or by using a large number of primary turns.

However, the use of many primary turns necessitates the use of a proportionately large number of secondary turns, and with many secondary turns the *distributed capacitance* of the secondary winding causes loss of amplification at the upper end of the frequency range. This distributed capacitance, which grows rapidly as n_2 increases, is represented in the equivalent circuit by a condenser C_L (dotted in Fig. 127b) in parallel with L_P. At a sufficiently high frequency the reactance of this condenser becomes so small that it practically short-circuits the transformer secondary. When this happens the gain vanishes entirely. The gain equation including the effect of distributed capacitance is:

$$\frac{E_3}{E_1} = \frac{n\mu}{\sqrt{1 + R_p{}^2 \left[\dfrac{1}{X_P} - \dfrac{1}{X_C} \right]^2}} \tag{610 p}$$

X_L vanishes at low frequencies and X_C at high frequencies, so that the gain vanishes for both high and low frequencies. In the intermediate range, where $R_p/X_P \ll 1$ because L_P is large, and $R_p/X_C \ll 1$ because C is small, the gain approximates $n\mu$. This is illustrated by the solid

curve in Fig. 128. The intermediate range must include all musical-tone frequencies in an amplifier that is to be used to reproduce music without frequency distortion. With the values of R_p, L_P, and C_L usually employed, Equation (610) *does not predict a resonant peak in the gain*, despite the appearance of the denominator.
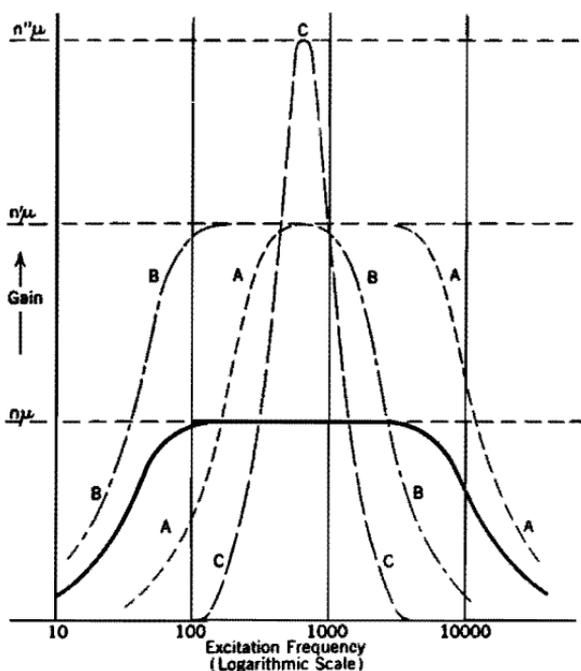
Fig. 128. The heavy solid line illustrates the variation of gain with frequency in a transformer-coupled amplifier with infinite load resistance, turn ratio n, no leakage reactance. The effects of changes in turn ratio on frequency response are illustrated as follows:

Curve $AAAA$: Turn ratio increased from n to n' by using fewer primary turns; as a result the lower frequency limit is raised.

Curve $BBBB$: Turn ratio increased from n to n' by using more secondary turns; as a result the upper frequency limit is lowered.

Curve CCC: Turn ratio increased by both means; as a result the frequency range is narrowed to a resonant peak at the frequency for which $X_P = X_C$.

Fig. 129 illustrates the experimentally measured response to frequency variation in a circuit like that of Fig. 127a. There is evidently agreement between the experimental results and the solid-line curve of Fig. 128, except for the sharp rise to a peak just before the high-frequency drop. This peak is caused by resonance between the distributed capacitance in the secondary and the *leakage reactance* of the transformer windings. It will be recalled that leakage reactance has not been

indicated in any of the circuit diagrams, and that it has not been taken into account in any of the expressions for gain. The simple circuit shown in Fig. 127b could never give a gain greater than $n\mu$. The effect of leakage reactance can be accounted for in the equivalent circuit analysis by introducing in Fig. 127b a series reactance X_L between R_p and the point where the C_L circuit branches off.

160. Transformer Size and Turn Ratio. If in an amplifier of the type illustrated in Fig. 127 the turn ratio is to be made larger in order to obtain a larger gain, and if the dimensions, material, and therefore magnetization characteristics of the core are to remain constant, either n_2 must be made larger or n_1 smaller. If n_1 is made smaller, L_P becomes less, and the range of frequencies over which constant gain exists becomes narrower because the lower frequency limit is raised. If n_2 is made larger, C_L is enlarged and the upper frequency limit lowered. The effects of both kinds of changes are illustrated in Fig. 128. Either increases the likelihood of frequency distortion.

Increase of the turn ratio thus reduces the frequency range at one end or the other, unless accompanied by an increase in physical bulk, weight, and therefore cost of the magnetic circuit. In radio receivers turn ratios rarely exceed 5 or 6 (only partly for reasons just discussed). Even with turn
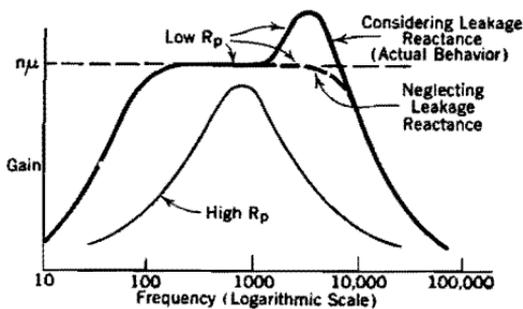
FIG. 129. Frequency response of a transformer-coupled amplifier with infinite load resistance, illustrating the effect of leakage reactance.

ratios no greater than this, very cheap transformers, having little iron and many turns, have an unsatisfactory frequency range. Reduction of the range at both ends occurs if a tube with very large R_p is used, or if the turn ratio is enlarged by increasing n_2 and at the same time decreasing n_1.

Equation (609) expresses a proportionality between the primary inductance of a transformer and the slope of the magnetization curve of the transformer's core. This is of interest because in real circuits like that shown in Fig. 127a the transformer primary winding carries a steady component of plate current which may be considerably larger than the alternating-current component. If the transformer core is not liberally designed partial magnetic saturation occurs owing to the flow of direct current. The actual primary inductance is then proportional to the average value of $d\Phi/dn_1 i_p$ for the range of values of $n_1 i_p$ represented in the

alternating-current " swing." If the direct-current point is well up on the magnetization curve, the slope of the curve in the alternating range is small; therefore the primary inductance is small and the low-frequency limit undesirably high.

Thus a transformer may give satisfactory service as long as it does not have to pass a steady component of current, yet introduce serious frequency distortion when required to carry such a current.

For some purposes it is not only acceptable but desirable to use a transformer coupling that responds best to one definite frequency. Then the turn ratio is made large enough so that the gain is nearly always limited by either X_P or X_C. The gain then approaches the value $n\mu$ only at resonant frequency, as illustrated by the high peaked curve in Fig. 128.

161. Transformer Coupling to a Finite Load Resistance. Figs. 130a and 130b illustrate the actual and equivalent circuits if a load resistance is coupled to the plate circuit through a transformer. X_P is the primary reactance, as in the previous section.   The load resistance is represented in the equivalent circuit by the magnitude R_L/n^2, rather than by its actual value R_L. (In the equivalent circuit diagram of a transformer all resistances and reactances of the secondary circuit may be treated as though in the primary, provided they are divided by the square of the turn ratio.   The derivation of Equation (590) illustrates the reasons why this is true.)

The output voltage E_3 has as before the value nE_2.   The complete equivalent circuit diagram would of course include representations of the resistances of the primary and secondary windings and of the leakage reactances, the primary elements in series with R_p and the secondary elements in series with the load.   These series elements are usually so small relative to R_p and R_L/n^2 that their omission introduces no appreciable error.

Throughout the major part of the normal frequency range of an amplifier like that of Fig. 130a the reactances X_P and X_C are both so large relative to R_L/n^2 that they may safely be neglected.   The circuit can then be analyzed as though X_P and X_C were removed entirely, corresponding to complete idealization of the transformer.   By analogy with the treatment of the simple equivalent amplifier circuit in Section 142, the expression for gain can then be written as follows:

$$\text{Gain} = \frac{E_3}{E_1} = \frac{nE_2}{E_1} = \frac{n\mu}{1 + \dfrac{R_p}{R_L/n^2}} \qquad (611\ \text{p})$$

Equation (611) shows that although the transformer magnifies the gain

by the factor n it also shrinks the load by the factor n^2. This latter effect tends to reduce the gain.

If R_p and R_L are both fixed, the gain is roughly proportional to the turn ratio, for moderate values of the turn ratio. For very large turn ratios the gain vanishes. This is illustrated in Fig. 130c. The solid line represents the gain for various values of n. The dotted line is a tangent to the solid line at the origin; its slope is $n\mu$. The maximum
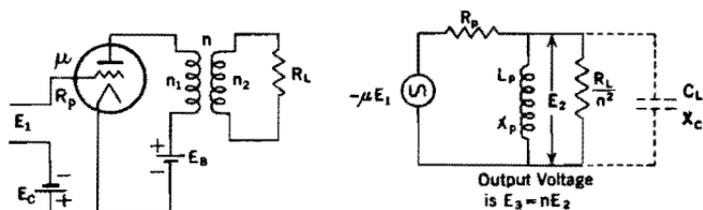
(a) Actual circuit diagram; $n = n_2/n_1$ (b) Equivalent circuit diagram.
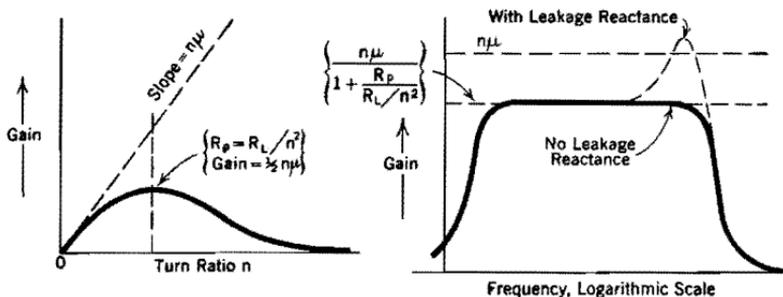
(c) Gain as a function of turn ratio.   (d) Dependence of gain on frequency.

FIG. 130.   Transformer coupling to a finite load resistance.

gain occurs at the value of n for which $R_p = R_L/n^2$, that is, at the turn ratio for which the resistances in the equivalent circuit are " matched." The gain at this point is $\frac{1}{2}n\mu$, E_2 being just half of $-n\mu E_1$. No advantage is to be gained by using a turn ratio greater than $\sqrt{R_p/R_L}$. If, as sometimes happens, R_L is less than R_p, it may be desirable to use a transformer having more primary than secondary turns, making n fractional.

At low frequencies the effect of X_P, and at high frequencies that of X_C, is important. If, in the equivalent circuit of the parallel-feed amplifier, Fig. 114b, the condenser is short-circuited (equivalent to making $X_C = 0$), the circuit of Fig. 130b is obtained. It is therefore possible to obtain, for Fig. 130, the proper expressions for gain and phase shift in terms of frequency by making $X_C = 0$ in Equations (560) and (561), and at the same time using R_L/n^2 for R_L, $n\mu$ for μ, and X_P for X_L.

The resulting expression can be adapted for use at frequencies high

enough to involve the distributed capacitance by employing

$$R_p{}^2\left(\frac{1}{X_P} - \frac{1}{X_C}\right)^2 \text{ in place of } \frac{R_p{}^2}{X_P{}^2}$$

Fig. 130d illustrates the general nature of the frequency response so predicted.

162. Regeneration. It is possible to couple the output of an amplifier back into the grid circuit, in such a way that the excitation voltage is the original input plus some definite fraction of the similarly patterned output voltage. Such an arrangement provides regeneration,[E XIII, E XIV, I VII] also called *feed-back*.[F 212, F X] The coupling may be made relatively " loose " so that the effect of the output voltage is only slightly apparent in the grid circuit. A stable *regenerative amplifier*, in which the gain is markedly increased by regenerative action, may be obtained in this way. The theory of stable regenerative action is rather involved, but has been fairly completely worked out.[F 217] It is not presented in this text.

If there is " close coupling " between output and input, so that the output affects the excitation strongly, any initial voltage impulse in either grid or plate circuit tends to magnify itself without limit; such behavior is termed unstable regenerative action. Most of the electronic circuits that are used to produce sustained electrical oscillations employ unstable regenerative action.

163. Tuned Plate Oscillators. A tuned plate oscillator is one of the many circuit arrangements[G 588, F X, I VII, H XVIII] that can be used to produce sustained oscillations by regenerative action. As illustrated in Fig. 131, it is very similar to an amplifier in which the load is a resonant circuit. The load is closely coupled to the grid circuit by transformer action due to the mutual inductance M between coils L_P and L_G. Any initial voltage impulse, such as that due to closing a switch that completes the plate circuit, sets up resonant-frequency electrical oscillations (an alternating current) in the plate circuit. But for the coupling back into the grid circuit, these oscillations would die out in a few cycles, just as a pendulum soon stops swinging unless given cyclic encouragement. As a result of the coupling into the grid circuit, the grid receives oscillatory excitation of just the proper frequency and phase relation to magnify the oscillations already started in the plate circuit, so that they continue indefinitely.

The vector diagram for the equivalent circuit of a tuned plate oscillator is shown in Fig. 131c. In preparing this diagram resonant frequency in the " tank circuit " is assumed. Therefore $X_C = X_P$, and $I_C = I_L$. Also, R_L is assumed to be small relative to X_P. From the

diagrams:

$$-\mu E_1 = I_C X_C + I_L R_p \frac{R_L}{\sqrt{R_L{}^2 + X_P{}^2}} \qquad \text{(612 p)}$$

The coupling between L_P and L_G is symbolized by the mutual inductance M. The corresponding mutual reactance $2\pi f M$ will be called X_M. The excitation voltage can be expressed as

$$E_1 = -I_L X_M \qquad \text{(613 p)}$$
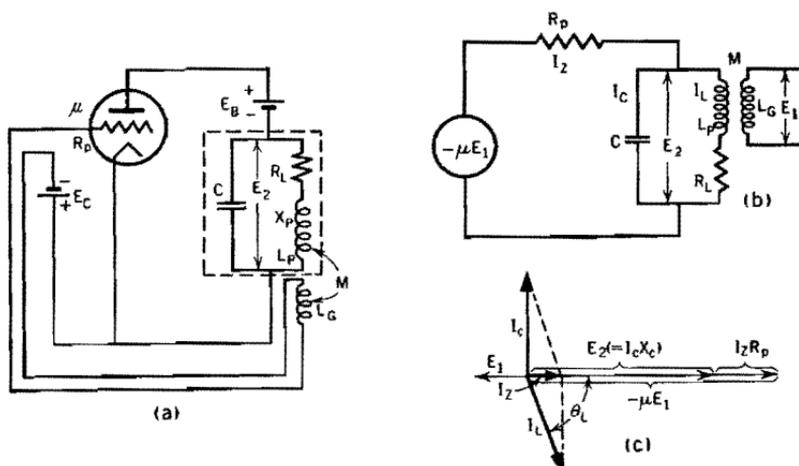
(a) Circuit diagram. The dotted line encloses the "tank circuit." Polarity is correct if L_P and L_G are separated sections of a continuously wound coil.

(b) Equivalent circuit diagram.

$$R_p = \frac{\mu M - L_P}{R_L C}$$

(c) Vector diagram. X_C symbolizes reactance of the condenser C, X_P the reactance of the inductance L_P.

Fig. 131.   Tuned plate oscillator.

The polarity of the coupling must be such as to call for the negative sign used here. Use of this equation and of the fact that $R_L \ll X_L$ convert Equation (612) into

$$\mu I_L X_M = I_C X_C + I_L \frac{R_p R_L}{X_P} \qquad \text{(614 p)}$$

The currents can be canceled. It is convenient to interchange X_C and X_P; this interchange is permissible because X_C and X_P are equal. There results

$$\mu X_M = X_P + \frac{R_p R_L}{X_C} \qquad \text{(615 p)}$$

If the X's are expressed in terms of M, L_P, and C, the $2\pi f$'s cancel, leaving

$$\mu M = L_P + R_p R_L C \qquad \text{(616 p)}$$

The usual form of Equation (616) is

$$R_p = \frac{\mu M - L_P}{R_L C} \tag{617 p}$$

Equation (617) seems to indicate that sustained oscillations can result only when the circuit constants are such as to make the right-hand side have a definite value, corresponding to the R_p of the particular tube used. However, the experimentally observed fact is that the circuit oscillates whenever the right-hand side of this equation exceeds a definite minimum value.

The reason for this behavior can be discovered by a graphical study of the current-voltage locus, shown in Fig. 132, for such an oscillating circuit. This locus lies along a straight " load line " because the " tank circuit " of Fig. 131 acts like a resistance, in that the total current I_Z through it is in phase with and proportional to the voltage E_2. An expression for the apparent resistance R_T of the tank circuit is obtained by observing from the circuit and vector diagrams in Fig. 131 that

$$I_Z = \frac{E_2}{\sqrt{R_L{}^2 + X_P{}^2}} \frac{R_L}{\sqrt{R_L{}^2 + X_P{}^2}} \cong E_2 \frac{R_L}{X_P{}^2} \tag{618 p}$$

The ratio of E_2 to I_Z is of course R_T, so that

$$R_T = \frac{X_P{}^2}{R_L} \tag{619 p}$$

The load line in Fig. 132 has the slope $1/R_T$. The tube characteristic curves, so also the locus, level off at a temperature-limited value of current. The existence of this upper current limit accounts partly for the existence of an upper limit to the continued growth of oscillation amplitude.

It is evident that the value of R_p is not constant at all points along the load line. The point Q lies, in Fig. 132, at about the height at which the greatest slope of the characteristic curves occurs, so at the height corresponding to the least value of R_p. This establishes the minimum value which the right-hand side of Equation (617) can have if oscillations are to persist. At just this least value the oscillations are of very small amplitude. Oscillation over a large voltage range along the load line brings the current-voltage locus into regions in which the intersecting plate characteristics have smaller slopes, corresponding to greater local values of R_p than exist near Q. Hence for large-amplitude oscillations the *average* value of R_p for a complete cycle of oscillation is much greater than for small-amplitude oscillation. Thus if, in Equation (617), M is increased, or R_L or C decreased, making the whole quantity larger, the amplitude

of oscillation grows, and with it the average value of R_p. Thus Equation (617) remains approximately true even for wide current swings.

As long as the current-voltage locus stays within the straight-line range BD, Fig. 132, the current I_Z and the grid voltage variations are approximately sinusoidal. However, the harmonic components of I_Z grow rapidly as operation extends out over the flat horizontal portions to the left of A and right of B, or even into regions to the left of D where the grid-voltage wave shape is affected by the flow of grid current. The tank circuit's impedance to these harmonics is very small, but to the fundamental-frequency current very large. Therefore the harmonic components of I_Z produce a very small, the fundamental component a very large, tank-circuit voltage. For this reason the tank-circuit voltage
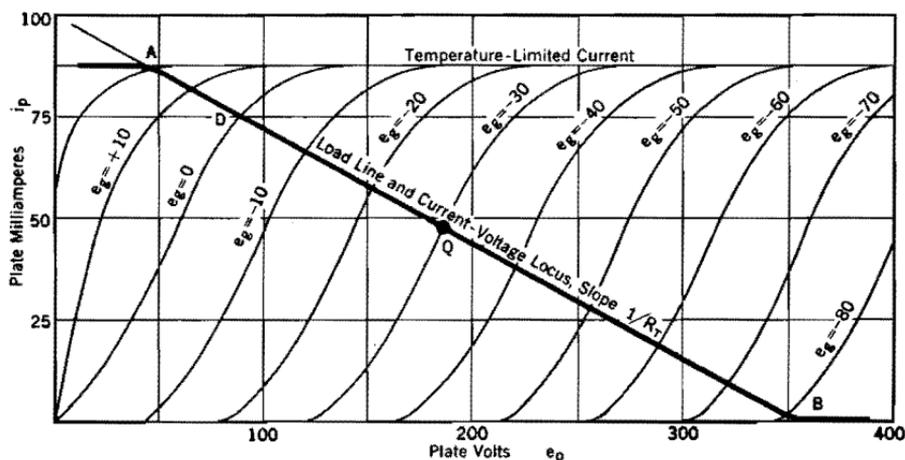
FIG. 132. Current-voltage locus for a tuned plate oscillator.

E_2 is very nearly sinusoidal, even at large oscillation amplitudes that make I_Z contain very pronounced harmonic components. These harmonic components flow through the tank circuit chiefly by way of the condenser.

In oscillators of this and most other types employing regeneration the grid swings positive during each cycle, so that grid current exists normally during a portion of each cycle. This permits self-biasing, illustrated in Fig. 133a. In a self-biased circuit the grid battery is replaced by a small resistance in parallel with a condenser. The voltage drop in the resistance, caused by flow of grid current, provides the grid bias voltage. The condenser is charged during grid-current peaks. By discharging through the resistance during grid-current valleys it maintains the grid bias voltage at a reasonably constant value throughout each cycle.

164. Other Regenerative Oscillator Circuits. For purposes of reference the essential features of three other types of oscillator circuits that employ the regenerative principle are shown in Fig. 133. Equations can be derived that describe the relations between R_p and circuit constants for each. Those for the tuned grid and Hartley oscillators appear below the figure. The proper interpretations of such expressions are all similar to that given for Equation (617).
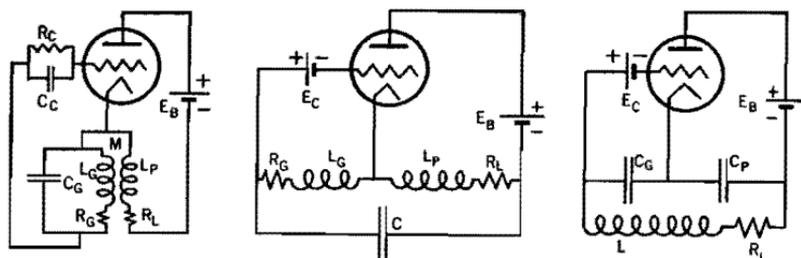
(a) Tuned grid oscillator.   (b) Hartley oscillator.   (c) Colpitt oscillator.

FIG. 133.   Circuit diagrams for regenerative oscillators. See also Fig. 131.

For the tuned grid circuit:

$$R_p = \frac{\mu M L_G - M^2}{R_G L_G C_G}$$

For the Hartley circuit:

$$R_p = \frac{(L_P + M)\,[\mu(L_G + M) - (L_P + M)]}{C\,(R_L + R_G)\,(L_P + L_G + 2\,M)}$$

165. Tuned Amplifiers. In Fig. 131a the inductance L_G is connected back into the grid circuit. If, instead, it is connected in series with the grid of a succeeding stage of amplification, the circuit becomes that of a *tuned amplifier*, which has a large voltage gain for frequencies very close to resonance, but a small voltage gain for all other frequencies. The equivalent circuit shown in Fig. 131b is similarly that of a tuned amplifier as well as of a tuned plate oscillator. When used as a tuned amplifier the voltage that appears across the secondary terminals of L_G is the excitation voltage of the next stage; when used as an oscillator it provides the oscillator's own excitation.

It is not difficult to analyze this equivalent circuit vectorially for a general frequency, for which the two reactances are not equal. Construction of the vector diagram and the derivation from it of the general equation for voltage gain are left as exercises for the reader. A study of the equation for gain, and of a graph of gain vs. logarithm of the frequency, shows that the use of a coil with a value of R_L small relative to X_P favors sharp "tuning." That is, with such a coil large gain occurs only over a very narrow band of frequencies. It is possible to design tuned amplifiers to give almost any desired sharpness of tuning.

166. Dynatron Oscillators.[85.5] The screen-grid plate characteristic curve shown in Fig. 56, page 130, has a negative slope, therefore a *negative plate resistance*, for a considerable range of plate voltages beyond *CC*. The volt-ampere curves for electric arcs and glow discharges, also the grid current-grid voltage curves for high-vacuum triodes, exhibit in whole or in part similar negative resistance properties.

Ordinary resistances, which are *positive*, absorb energy from alternating-current circuits. A device for which the alternating-current resistance is *negative* can introduce alternating-current energy into a resonant load circuit like that shown in Fig. 134*a*. The source of energy is of course the plate battery. A circuit which employs a negative-resistance volt-ampere curve to produce an alternating current is called a "dynatron oscillator."

Dynatron operation can be analyzed by reference to Fig. 134*b*, in which there is shown a negative-resistance portion of a tube characteristic curve. The particular type of device used to produce this curve is unimportant. However, if a screen grid or other tube with more than two electrodes is used, no variations in screen or control grid voltages are required to produce the oscillatory behavior. Operation occurs along *one tube characteristic curve only*. The present study does *not* involve a *family* of tube characteristic curves.

The tank circuit in Fig. 134*a*, like the tank circuits in tuned plate oscillators, behaves at resonant frequency as a resistance of $R_T = X_L{}^2/R_L$ ohms. During normal operation as a dynatron oscillator, resonant-frequency current flows through this apparently resistive tank circuit, so that the current-voltage locus must lie along a straight "load line" passing through the operating point and having a slope $1/R_T$. The plate voltage at Q is the plate battery voltage, because the resistance R_L is so small that the direct-current drop through it is negligible. Three possible load lines, I, II, and III, are illustrated in Fig. 134*b*.

The load circuit thus requires that the current-voltage locus must lie along a straight load line. However, the current-voltage locus must also lie along the tube characteristic curve. Both of these requirements are simultaneously satisfied when the load line coincides with the tube characteristic, as along a portion of Load Line II, Fig. 134*b*. For this load line $R_T = -R_p$ (evaluated at Q). Oscillation can take place at almost any desired frequency, for there is a wide range of values that L and C can have, yet give to R_T the required magnitude.

The current in this type of circuit is not *necessarily* oscillating; it may be steady, with voltage and current values as at Q. In order to start oscillation it may be necessary to give the circuit an initial electrical "kick." Making the final connection that completes the plate battery

circuit usually accomplishes this purpose satisfactorily. An alternate method is to complete the circuit except for the condenser, then to give the condenser a charge, and subsequently connect it into the circuit.

Suppose this latter method is used in a dynatron circuit having resonant-frequency Load Line II, Fig. 134b. The condenser is initially charged to 15 volts or less. Load Line II coincides exactly with the tube characteristic curve for about 15 volts each way from Q. Since the negative resistance of the tube is exactly equal numerically to the positive apparent resistance of the tank circuit, the net alternating-current resistance of the oscillating circuit is zero. Oscillations start with an

(a) Circuit diagram.   (b) Current-voltage variations.
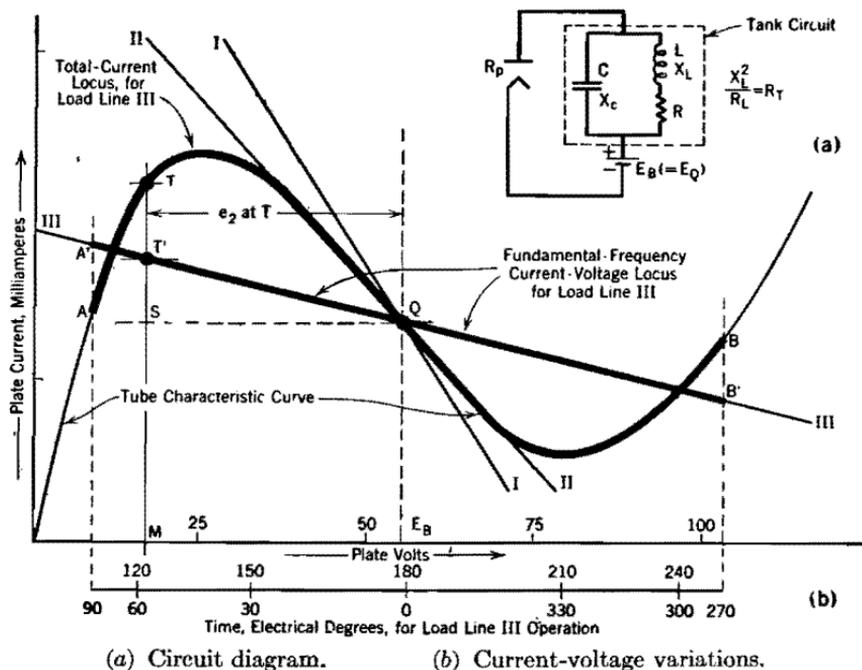
Fig. 134.   Dynatron oscillator.

amplitude dependent on the initial charge on the condenser, and continue without change in amplitude. This is analogous to the behavior of a frictionless pendulum, which swings forever with an amplitude dependent on the initial displacement given it. If the condenser is initially charged to 15 volts, the plate voltage will swing to 15 volts above and below Q.

If the tank circuit constants are such as to produce resonant-frequency Load Line I, the apparent resistance of the tank circuit is greater numerically than the negative plate resistance. Hence the oscillating circuit has a net positive resistance. Behavior is similar to that of a

pendulum whose swing involves friction.   Oscillations start, but decay at a rate dependent on the magnitude of the net positive resistance.

Suppose the charged condenser is switched into a circuit having resonant-frequency Load Line III.   Oscillations start as usual with an initial amplitude dependent on the charge on the condenser.   However, the amplitude increases at a rate dependent on the net alternating-current resistance, which is now negative.   If the volt-ampere characteristic of the tube were an infinitely long straight slant line, the oscillation amplitude would increase indefinitely.

Actually the tube characteristic curve cannot continue straight indefinitely; it must eventually bend.   The actual current-voltage locus must follow the bends.   This at first makes the average plate resistance for the entire cycle have a still greater value than along the straight-line portion, so increasing the rate of growth of amplitude.   Soon the swing extends beyond the maximum and minimum points on the curve, so entering regions in which the local contributions to the average plate resistance are *positive*.   Extension of amplitude farther into these regions is accompanied by a change to a progressively less negative plate resistance.   Operation stabilizes ultimately at an amplitude for which some average of the plate resistance is $-R_T$.

When dynatron oscillations are of sufficient amplitude to extend into curved parts of the tube characteristic curve, the current in the tube contains harmonics of the fundamental resonant frequency.   The impedance of the tank circuit is very high to the fundamental frequency, but very low to all its harmonics.   For example, if at resonance $X_L = X_C = 10$ ohms, and $R_L = 0.1$ ohm, the apparent resistance of the tank circuit at fundamental frequency is $100/0.1 = 1,000$ ohms; the reactance to the second harmonic is approximately 6.7 ohms, and to higher harmonics still less than this.

All the harmonic components of current, therefore, pass through the tank circuit with inappreciable voltage drop.   In contrast with this, the fundamental-frequency component of plate current produces a large sinusoidally varying fundamental-frequency component of plate voltage. Thus, as in a tuned plate oscillator, the alternating component e_2 of plate voltage is almost a pure sine wave in spite of the presence of pronounced harmonics in the plate current.   A time scale, in electrical degrees, for the variation of plate voltage for Load Line III is shown in the figure.

Two current-voltage loci are needed properly to describe operation in connection with Load Line III.   Current values along one of them, $A'B'$, follow the fundamental-frequency current variations.   The corresponding voltage variation is sinusoidal, for the reasons stated in

the previous paragraph, and the tank circuit is resistive. Therefore this locus is a straight line along Load Line III. The other locus, AB, is a total-current locus, and must therefore follow the tube characteristic curve.

The points T, T', describing voltages and currents at about the $57°$ instant, may be used to contrast these two loci. Here the sinusoidal alternating component e_2 of plate voltage has the instantaneous value $\overline{QS}$, as shown in the figure. The total plate current is $\overline{MT}$, the fundamental-frequency component of plate current is $\overline{ST'}$, and the harmonic components total $\overline{TT'}$.

The only alternating-current power consumption in such a circuit is that due to the fundamental-frequency current, because the harmonics flow chiefly through the condenser, which has a very small power factor.

PROBLEMS

Chapter XIV

1. (a) Derive Equation (607), using the vector diagram that appears in Fig. 126.

(b) Derive also an equation for the phase angle between E_3 and E_1 for this circuit.

2. Circuit as in Fig. 126a. $R_p = 12,500$ ohms, $C = 0.0015$ microfarads, $R_L = 25,000$ ohms, $R_G = 2.5 \times 10^6$ ohms. If the excitation voltage applied to the first tube is 3 volts (root-mean-square value) find from Equation (607) the excitation voltage of the second tube:

(a) At 200 cycles per sec.

(b) At 20 cycles per sec.

(c) Find what per cent of error would result in (a), also in (b) by using Equation (606) instead of Equation (607).

(d) Find the phase angle between E_3 and E_1, at 200 cycles, also at 20 cycles.

3. Circuit as in Fig. 127. $\mu = 8$, $R_p = 5,000$ ohms, $n_2/n_1 = 3$.

(a) Find what value of L_P will make the voltage gain at 20 cycles 60 per cent of its maximum value, assuming a gain curve of the general type illustrated by the solid line in Fig. 128.

(b) Find what value of transformer distributed capacitance will make the gain at 4,500 cycles 60 per cent of its maximum value.

(c) Assume that L_P varies as the square of n_1, and the transformer distributed capacitance as the square of n_2. The turn ratio is increased above the former value of 3 by increasing n_2 in the same ratio that n_1 is decreased. What value of turn ratio, so obtained, will raise the low-frequency 60-per-cent point to 40 cycles?

(d) To what frequency will the high-frequency 60 per cent point be lowered by the change described in (c)?

(e) What turn ratio will produce a peaked gain curve (similar to Curve CCC, Fig. 128) in which the 60 per cent points are at .8 and 1.25 times the resonant frequency respectively?

4. Add to the circuit of Fig. 127b a transformer leakage inductance in series with R_p; then derive, by the use of an appropriate vector diagram, an expression for voltage gain of the circuit, including the effects of the distributed capacitance. Show

that your equation describes a gain curve similar to the heavy solid line in Fig. 129 if R_p is small, and one similar to the light solid line in the same figure if R_p is large.

5. Circuit as in Fig. 130a; $n = 3$, $\mu = 8$, $R_p = 10,000$, $R_L = 200,000$.

(a) What is the voltage gain of this amplifier for frequencies high enough so that the primary reactance X_P of the transformer can be considered infinite?

(b) How large must the primary inductance of the transformer be in order to make the gain at 50 cycles per sec be 60 per cent of its value at 1,000 cycles?

6. Circuit as in Fig. 130a. $\mu = 8$, $R_p = 10,000$ ohms, $R_L = 1,000,000$ ohms.

(a) What value of n will give maximum amplification at high frequencies?

(b) What two values of n will each give half the maximum possible high-frequency amplification?

7. Circuit as in Fig. 130a. $R_p = 5,000$ ohms, $\mu = 8$, $n = 4$. The frequency is high enough so that primary reactance may be considered infinite. Determine R_L so that the gain will be 80 per cent of the value it would have if R_L were infinite.
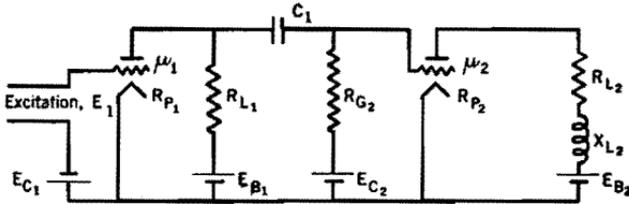
FIG. B.   Two-stage amplifier.

8. Circuit as in Fig. B.

$$R_p = 12,500 \qquad R_{p_2} = 1,500$$
$$\mu_1 = 10 \qquad \mu_2 = 3$$
$$R_{L_1} = 50,000 \qquad R_{L_2} = 500$$
$$C_1 = 0.00265 \text{ mf} \qquad X_{L_2} = 1,000$$
$$R_{G_2} = 800,000$$

The frequency of the excitation voltage E_1 is 1,000 cycles per sec, and the voltage across R_{G_2} is 6 volts, root-mean-square value.

(a) Draw the two equivalent circuits, and indicate what relationship exists between them.

(b) Find the excitation voltage for the first tube.

(c) Find the output voltage.

(d) Find, using clearly drawn vector diagrams, the phase angle between input voltage and output voltage.

(e) Find the plate circuit efficiency of the last stage.

(f) Find the decibel voltage gain of each stage, and of the whole.

9. Both tubes in the circuit of Fig. C (next page) have plate characteristics as illustrated in Fig. 104b. The choke in series with the plate battery of the first stage has a high enough inductance to prevent passage of alternating currents of the excitation frequencies to be used.

(a) For the tube in the first stage of amplification the point Q of zero excitation is to be at $E_Q = 250$ volts, $I_Q = 80$ milliamperes. Specify the proper values of plate battery voltage and grid bias voltage.

(b) Draw the equivalent circuits for this two-stage amplifier, and indicate clearly the relationship that exists between them.

(c) The excitation frequency is that which produces resonance in the series L, C plate circuit of the first stage.   On a sheet of plate characteristics similar to Fig. 104b draw load lines for each of the two stages of amplification.   Distinguish carefully between them.
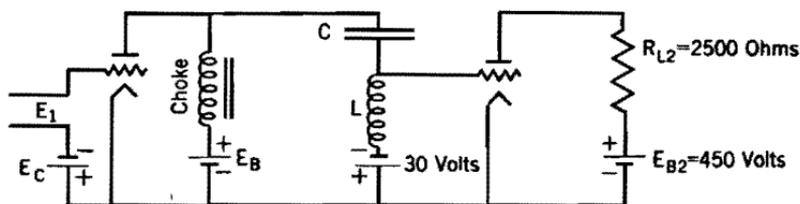
FIG. C.   Two-stage tuned amplifier employing a series resonant circuit in the first stage.

(d) Determine graphically from your set of plate characteristics the values of μ and R_p that apply for the second stage of amplification.   Indicate clearly how your values are obtained.

(e) What is the voltage gain of the second stage of amplification?

(f) At a frequency (not necessarily resonant frequency) such that the voltage gain of the first stage of amplification is 5, what is the overall gain of this two-stage amplifier?

(g) Find the phase angle between output and excitation in the first stage, at resonant frequency.

10. Circuit as in the previous problem; at resonant frequency the reactance of L is 5,000 ohms.

(a) Find the gain in the first stage of amplification, at resonant frequency.

(b) Frequency twice that of resonance; again find the gain of the first stage.

(c) Find the limit which the gain approaches as the frequency is increased indefinitely, also that approached as the frequency is decreased toward zero.

(d) Using the results of (a), (b), and (c), sketch a curve of gain vs. the common logarithm of the frequency.

11. Circuit as in Fig. 131, tube characteristics as in Fig. 132.   The frequency of oscillation is to be 50,000 cycles per sec.   Select a set of values of L_P, L_G, C, and R_L that will make R_p 20 per cent more than the least value possible for a tube having characteristics as in Fig. 132; it is furthermore required that $X_L = 20R_L$, and that the coefficient of coupling K between L_P and L_G be 0.2.   (The coefficient of coupling relates M to L_P and L_G as follows: $M = K \sqrt{L_P L_G}$.)

12. Draw the equivalent circuit and derive the equilibrium equation for R_p, for

(a) A tuned grid oscillator, Fig. 133a.

(b) A Hartley oscillator, Fig. 133b.

(c) A Colpitt oscillator, Fig. 133c.

13. Consider a tuned amplifier circuit which is similar to the circuit illustrated in Fig. 131, except that L_G is connected in series with the grid of a subsequent stage of amplification.   Tube characteristics are as in Fig. 104b, and the operating point Q is at $E_Q = 250$ volts, $I_Q = 80$ milliamperes, $E_C = -30$ volts.   Use $R_p = 700$ ohms.

(a) R_L is ten per cent of the resonant value of X_P and X_C; what values must X_P, X_C, and R_L have to make the voltage across L_P be $0.8\mu E_1$ at resonant frequency?

(b) Using the values of X_P, X_C, R_L just determined, what is the voltage gain if $M = 5L_P$?

(c) Draw a load line for this amplifier on your plate characteristic sheet, and contrast its position with that of the load line for the first stage of Problem 9.

(d) Find the voltage gain of this amplifier at half the resonant frequency, and at twice the resonant frequency.

(e) Determine the limits which the gain approaches as the frequency rises indefinitely; also that approached as the frequency falls toward zero.

(f) Using the results of parts (b), (d), and (e), sketch a curve of gain vs. the common logarithm of the frequency.

14. Circuit as in Fig. 134a, tube an RCA-24, characteristics as in Fig. 55. Grid bias is 0 volts. Specify a set of values of E_B, L, C, and R_L that will permit stable 50,000-cycle dynatron oscillations of the type corresponding to Load Line II, Fig. 134b. Make $X_L = 20R_L$.

PART II

ELECTRONS, ATOMS, AND RADIATION

CHAPTER XV

ATOMIC ENERGIES*

167. The Function of Positive Ions in Gaseous Conducting Devices.
The electronic devices dealt with in Part I contain only negatively
charged current-carrying particles, that is, electrons. In contrast to
this, gas-filled electron tubes, mercury-arc rectifiers, gas and vapor light
sources, electric arcs, glow discharges, sparks, etc., contain charged
particles of both signs, for the most part in approximately equal con-
centrations. Yet in them, as in high-vacuum electronic devices, the
important current flow is that due to electron movement. The positively
charged particles present are positive ions. Their mass per particle is
so much greater than that of the electrons that they move relatively
very slowly; in mercury vapor the velocity ratio is about 1 to 605.
Hence the rate of charge transfer due to positive ion motion is an in-
significant part of the total current.

Yet in spite of the fact that positive ions carry so little current, their
presence makes the current-voltage relation altogether different from
what it is without them. When they are not present, as in high-vacuum
electronic devices, the current density due to electron movement is
limited by space charge. When they are present, and equal in con-
centration to the electrons, there is no space charge, and the electron
current density may reach tremendously large values even though the
potential gradient is small.

Interest in regard to the ions in gaseous electrical discharges centers
around the manner and rate of their appearance and disappearance. It
is important to know where they come from and where they go, and at
what energy cost. If they are present, electron space charge is neu-
tralized and large currents can flow in response to a small potential
difference; if they are absent, current flow can take place only under the
influence of a high potential, no matter how freely electrons may be
supplied.

* The suggestions and criticisms made by Dr. S. A. Goudsmit of the Physics
Department, University of Michigan, relative to Chapters XV and XVI have been
of very great assistance to the author.

There is an interesting similarity between an ionized gas and a metallic conductor. In both, current flow consists of an electronic "wind" passing through a region populated by positively charged particles which are relatively stationary. The positively charged particles of a metal are the atoms of its structure; they may be called positively charged because each one has contributed its outer electron to the electronic wind that blows through the lattice-work of atoms. The heat motions of these atoms are vibratory in nature, so that each one retains its average position relative to its neighbors. In an ionized gas the positively charged particles are gas ions whose random heat motions carry them throughout the body of gas; however, relative to the electronic wind that constitutes the current, they are practically at rest.

168. Energy Required for Ionization; Energy-Level Diagrams. The arrangement of electrons around atomic nuclei was discussed in Section 80. A positive ion is an atom or molecule from which an outer electron has been removed. Just as an electron can be removed from a metallic surface only by the introduction of energy (heat energy in thermionic emission), so an electron can be removed from a particle of gas or vapor only by the contribution of energy. The least amount of energy necessary to remove a single electron from a metal surface is definite and measurable; it is called the work function and is usually a matter of a few electron volts. The least energy that can remove an electron from a gas particle is similarly definite and measurable in electron volts. It is called the *ionizing potential* and is between 5 and 25 volts for most gas and vapor atoms.

Energy-level diagrams similar in principle to those used in work-function representation are immensely valuable in the study of ionization and kindred processes in gases. Fig. 135 compares the energy-level diagram for an atom of sodium vapor with a metal-conductor type of diagram. To emphasize the similarity the work function of the metal on this diagram has been made the same as the ionizing potential of sodium, 5.12 volts. Below the line that is described as the "normal level" there are, in both diagrams, "occupied" or "filled" levels, representing the energy states of the electrons present in the metal or in the atom. The metal has a very large, but finite, number of filled levels, there being one for each free electron in the metal. In a sodium vapor atom there are altogether eleven filled levels, corresponding to the eleven electrons that form the "atmosphere" around the atomic nucleus. These eleven occur in groups as indicated; first two close together, then eight, then one alone.

The quantitative parallelism between Figs. 135*a* and 135*b* is limited to the spacing between the normal and zero levels, for the bottom metal

level is likely to be between two and twenty volts below the one here
called normal, as compared with a matter of thousands of volts for the
sodium-vapor atom.

169. Excited States of Atoms. Fig. 135*a* and Fig. 135*b* represent
minimum-energy conditions: in the metal, that of absolute zero tem-
perature; in the atom, the " normal state." This means the condition
in which most atoms are found in gases at ordinary room temperatures.
If an atom or a piece of metal initially in the minimum-energy or normal
state receives energy from an outside agent, one or more of the electrons
must rise out of a level that is normally filled into one or more of those
ordinarily vacant. The normally vacant levels lie above the normal
level in both metal and atom. In the metal the rising electrons or-
dinarily come from the normal level or those just below it. In a gas
or vapor atom just the normal-level electron is elevated into one of the
higher normally empty positions.

When this has happened an atom is said to be *excited*, or in an *excited
state*, or to have an electron in an *excitation level*.[X 400] Fig. 135*c* is an
energy-level diagram for an excited atom. An electron ordinarily occu-
pies an excitation level for only a very brief period of time, of the order of
a hundredth of a microsecond (10^{-8} second). As it drops to a lower
excitation level or to the normal level energy is released. This energy
usually appears as radiation. The amount of energy released in the
return to the normal state is exactly the amount originally required
to accomplish the elevation to the excited state, and is measured by the
spacing between the levels on the diagram.

There are ordinarily no levels above an atom's zero level. If an atom
receives energy in excess of that needed to lift an electron to the zero
level, the electron escapes from the atom altogether, so converting the
atom into an ion.

170. Transitions between Levels. In electrical discharges atoms
are continually being raised from the normal into excited states by
collisions with other particles. This activity is the result, directly or
indirectly, of the acceleration of charged particles by the electric field
between the electrodes. As the affected atoms return to the normal
state they radiate energy electromagnetically, sometimes in the form of
visible light. This gives rise to the luminosity of arcs, sparks, and glow
discharges.

An *ionizing collision* is one in which the outer electron of an atom is
lifted past the ionization level and escapes. Enough ionizing collisions
must occur in an electrical discharge to maintain the necessary ion con-
centration in spite of the gradual diffusion of ions out of the path of the
discharge. *Exciting collisions* lift electrons into excitation levels,

Exciting collisions occur very much more frequently than do ionizing collisions.

The vertical spacing between any two levels is proportional (*a*) to the amount of energy that must be imparted to an atom by a collision in order to produce a shift from the lower to the upper of the pair, and (*b*) to the *reciprocal of the wave length*, that is, to the *number of waves per centimeter*, of the electromagnetic radiation emitted during the down-
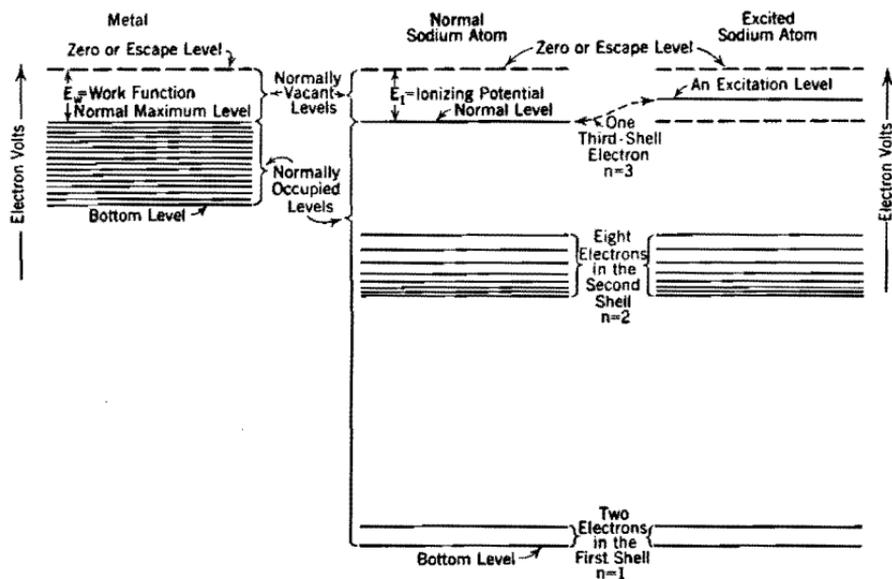
(*a*) Energy levels for the free electrons in a metal, as in Fig. 69.

(*b*) Energy levels for the 11 electrons of a sodium atom's " electronic atmosphere " in the *normal state* of the atom.

(*c*) Similar to (*b*), in an *excited state* of the atom.

FIG. 135. Comparison between an energy-level diagram for a metal and an energy-level diagram showing the normally occupied levels in a sodium atom.

ward shift. Thus the particular kinds of radiation sent out by a conducting gas are determined by the spacings between the various excitation levels.

The excitation levels lie above the normal level and below the ionization level. Hence the portions of energy-level diagrams that lie between these two extremes are especially interesting and useful. In fact, the phrase, " energy-level diagram " [X 413, DD] usually refers to just this portion alone, illustrated for sodium in Fig. 136*b*, and for neon and mercury in later figures. The proper placement of the excitation levels for the various elements is determined by analysis of spectral photographs that record, experimentally, the wave lengths of the spectral lines.

The usual cycle of energy changes from normal state to excited state, then back to normal state again, consists of abrupt elevation of an electron to some excitation level, followed by a cascading back to the normal level by a series of long or short downward *transitions* between levels. The following typical cycle for a sodium atom is illustrated in Fig. 136a:

(1) A fast-moving electron strikes a sodium atom, and causes a shift from the normal level to one that is 3.60 volts above normal. The attacking electron goes on its way with kinetic energy reduced by the amount given to the atom.

(2) The atom remains in this excited 3.60-volt state for an uncertain but extremely short period of time, of the order of 10^{-8} second.

(3) A transition (abrupt shift downward) occurs, terminating at the 2.09-volt level. Simultaneously a light impulse is radiated whose energy is correspondingly $3.61 - 2.09 = 1.52$ electron volts, at a definite wave length.

(4) Hesitation again occurs for perhaps a hundredth of a microsecond, then

(5) Another transition carries the atom to the normal state. At the same time a 2.09-volt light impulse is released. This impulse has the yellow color that is used as a chemical test for the presence of sodium. In fact, each transition corresponds to some one definite line in the spectrum of sodium.
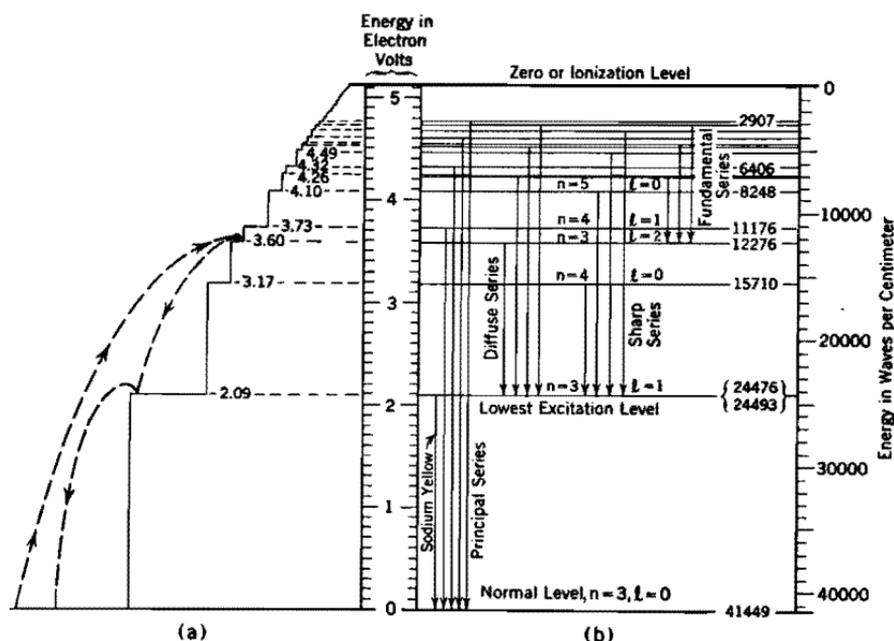
Fig. 136. Energy levels and transitions for a sodium atom. (See Table XI.)

Fig. 136a illustrates the cycle just described as the trajectory of a ball which has been thrown up a flight of stairs, only to bounce down again. The particular height to which it rises depends on the energy initially received.

Often the return to the normal state from a particular excited state may take place in any of a variety of ways.   A single transition directly to the normal state may occur, or the return to normal may take place in two, three, or several transitions.

The probability of a transition from the 3.60-volt level in sodium to either the 3.17-volt level or the normal level is practically zero.   Such transitions are never observed experimentally, and are called *forbidden* transitions.   Those that do occur are called *permitted* transitions.[DD, EE 14] On energy-level diagrams vertical lines are drawn connecting levels between which transitions are permitted; this has been done in Fig. 136*b*.

The spectrum of a sodium vapor light-source contains a spectral line corresponding to each one of the vertical lines in Fig. 136*b*, for each time a permitted transition occurs, light of its characteristic wave length is emitted.   Some of the permitted transitions are more likely to occur than others, so occur more often.   This is a partial explanation of the differences in the intensities of the various colors of light produced by a given source.

171. Electron-Volt Measure of the Color of Light.   The proportionality between the energy released in a transition and the number of waves per centimeter of the resulting radiation arises out of the fact that light emission invariably occurs in impulses of definite energy content, variously called *photons, light quanta, light particles,* and *light impulses.*   This fact has been experimentally demonstrated.

The energy of a photon is proportional to the light frequency, according to the equation[U 330]

$$W_{Ph} = hf \tag{701}$$

Here W_{Ph} is the photon's energy in ergs, f the radiation's frequency in cycles per second, and h a universal (Planck's) constant, of value 6.542×10^{-27} erg-second.   This is the same action constant that appeared in the statement of the principles governing the velocity distribution laws for gas particles.

If the light-quantum energy W_{Ph} is expressed in electron volts E_{Ph},

$$W_{Ph} = E_{Ph}e \times 10^7 = hf \tag{702 p}$$

This can be rearranged into

$$E_{Ph} = \frac{h}{e} 10^7 \cdot f \tag{703 p}$$

In all electromagnetic radiations, whether radio waves, heat, light, X-rays or cosmic rays, wave length λ times frequency must give the velocity of light c, that is

$$\lambda f = c \tag{704}$$

If λ is wave length in centimeters, and f is cycles per second, c must of course be 3×10^{10} centimeters per second.   If λ is in meters, c is 3×10^8 meters per second.

Combination of Equations (703) and (704) gives

$$E_{Ph} = \frac{hc}{e} 10^7 \frac{1}{\lambda_{cm}} \qquad (705\ p)$$

After introduction of numerical values this can be written in either of the two following forms:

$$E_{Ph} = 12335 \times 10^{-8} \frac{1}{\lambda_{cm}} \qquad (706\ p)$$

$$E_{Ph}\, \lambda_{cm} = 12335 \times 10^{-8}$$

This relation is easy to bear in mind, and serves two useful purposes: (1) it shows that the reciprocal of the wave length, also described as the
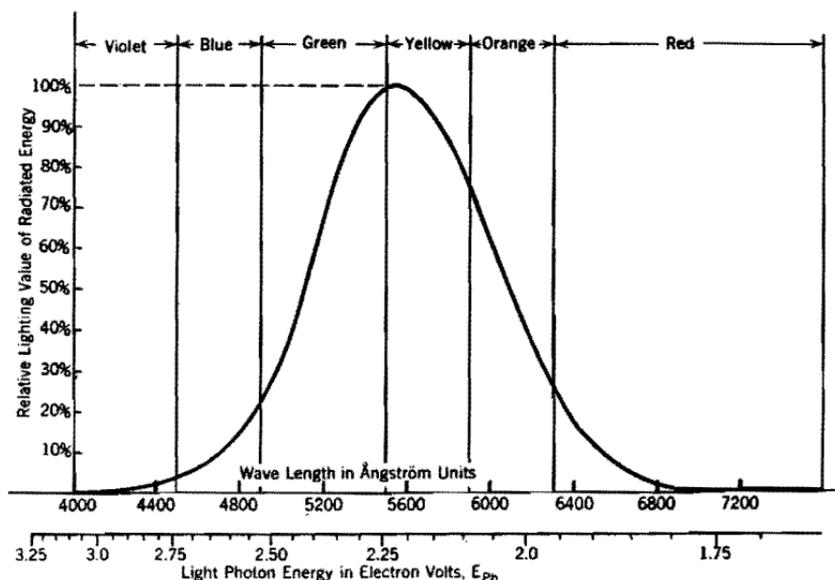
FIG. 137.   Relative sensitivity of the human eye to light of various colors. (Reproduced from *Lighting Data*, Bulletin LD-114 D of the General Electric Company.)

number of waves per centimeter of wave train, is a direct measure of the energy of a photon, and (2) it permits description of the color of light in electron volts.

The wave length of visible or near-visible radiations is often measured in angstrom units, which are units of length each 10^{-8} centimeter long. Equation (706) can be rewritten

$$E_{Ph} = 12335 \frac{1}{\lambda_{\mathring{A}}} \qquad (707\ p)$$

λ , Wave Length in Centimeters

| 10^6 | 1000 | 100 | 10 | 1.0 | 0.1 | 0.01 | 0.001 | | 10^{-6} | 10^{-9} | 10^{-12} |

Microns

| 10000 | 1000 | 100 | 10 | 1.0 | 0.1 |

Millimicrons

| 10 | 1.0 | 0.1 | 0.01 |

Meters

| 10000 | 1000 | 100 | 10 | 1.0 | 0.1 |

Ångström Units

| 100000 | 10000 | 1000 | 100 | 10 | 1.0 |

X- Units

| 100 | 10 | 1.0 | 0.1 |

Radio Communication

"Short Waves"

U.S. Commercial Broadcast, 500 to 1500 Kilocycles

Infra- Red

Heat Rays

Ultraviolet Near   Far

Visible Spectrum 4000Å to 7000Å

X - Rays

γ - Rays

Cosmic Rays

| 1000 | 10,000 | 100,000 | 10^6 | | 10^9 | | 10^{12} | | 10^{15} | | 10^{18} | | 10^{21} |

f, Radiation Frequency in Cycles per Second

| 10^{-9} | | 10^{-6} | | 0.001 | 0.01 | 0.1 | 1.0 | 10 | 100 | 1,000 | 10,000 | 10^5 | 10^6 | 10^7 |

E_{Ph} Electron Volts per Photon

Cadmium Red Line, International Spectral Wave · Length Standard, 6438.4696Å

1 Micron (μ) $= 10^{-4}$ cm.
1 Millimicron (mμ) $= 10^{-7}$ cm.
1 Ångström Unit (Å) $= 10^{-8}$ cm.
1 X Unit (X.U.) $= 10^{-11}$ cm.

FIG. 138.   Total electromagnetic radiation spectrum.   Logarithmic scales.

Here $\lambda_\text{Å}$ symbolizes wave length in angstrom units.   Fig. 137 is a graphical representation of the variation of sensitivity of the human eye to light of various colors.   Fig. 138 gives the approximate wave-length and electron-volt limits for the various kinds of electromagnetic radiations that human experience encompasses.

172. Scales on Energy-Level Diagrams.   The left-margin scale on Fig. 136*b* is divided in electron volts, and has its zero at the normal level; this is the most satisfactory scale to use when considering rates and probabilities of excitation and ionization.   The right-margin scale is divided in waves per centimeter, also called wave numbers.   Its zero is at the zero level.   It is used in calculating wave lengths corresponding to the various transitions, and in analyses leading to the placement of levels.

The wave-number difference between the 2.09-volt level and the normal level is $41{,}449 - 24{,}476 = 16{,}973$.   This describes sodium's characteristic yellow light in waves per centimeter.   Its reciprocal, 5890×10^{-8} centimeter, or 5890 angstrom units, is the corresponding wave length.   The same result can be arrived at by using the voltage scale.   The energy difference between these two levels is just 2.09 electron volts, which when used in Equation (707) gives 5890 angstrom units for $\lambda_\text{Å}$.   The placement of the zero at the top of one scale and at the bottom of the other is done as a matter of convenience in the particular services for which each scale is used.

173. Resonance Radiation; Photoelectric Action.   The light-producing transitions from high to low levels are reversible, within certain limitations.[X 412]   In particular, an atom may be excited by the absorption of a light-pulse or photon, *provided the energy of each incoming light impulse is the same or very nearly the same as the energy-difference of the transition.*   This gives rise to *resonance radiation.*

Imagine that light from a monochromatic (single-wave-length) light-source falls on a glass tube containing sodium vapor.   If the color (wave length) of the monochromatic light is varied in the direction of increasing energy, from infra-red on up, there is at first no effect on the sodium vapor.   The light passes through it with only slight diffusion.

However, when the light-source begins sending out 2.09-volt photons, the entire sodium vapor chamber becomes itself a light-source *of the same color,* sending radiations in all directions.   Various atoms are being raised to the 2.09-volt excited states by the incoming light;   on their return to normal states 2.09-volt radiation occurs.   This resonance radiation originates in the sodium vapor tube, but the energy for it comes from the monochromatic light-source.   .

If the monochromatic light-source begins sending out light carrying more than 2.09 electron volts per photon, resonance radiation stops, and

the sodium-vapor tube becomes dark again.   The reason for this is that the probability of absorption of a photon is very small unless the photon's energy is just the amount required to raise the atom to an excited state. In contrast to this, excitation by impact may result from collisions with particles having a considerable range of energies above the required value.   The difference arises from the fact that a photon must give up all or none of its energy; a moving particle may give up any fraction of its energy.

If the incoming light-impulse energy is sufficient, ionization may occur; the resulting ions are said to be produced by photoelectric action. The analogy with the energy-level diagram of a solid piece of metal (Fig. 135) can now be reversed.   *Photoelectric emission* from a metal results when incoming photons have sufficient energy to raise internal electrons from the uppermost filled (" normal ") level to the zero or escape level. Photoelectric emission is discussed in Chapters XVII and XVIII.

174. Spectral Symbolism and Electron Configuration.   The various energy levels for any particular element are distinguished from one another qualitatively by *spectral symbols*.[EE, X 469]   In addition to serving many other useful purposes, these symbols systematize the distinctions between permitted and forbidden transitions, and permit easy sorting out of the levels which contribute to any particular series of spectral lines.   The symbols tell an initiated reader certain very definite things about the positions and movements of outer electrons relative to a nucleus and the inner electrons.

A variety of systems of symbolism have been used, but all are related more or less closely to the physical theories which explain the arrangements, or " configurations " of the electron's low-lying *filled* levels, as revealed by X-ray spectra.   For this reason all discussion of the symbols of spectral levels or " terms " will be postponed until after a brief review of the principles governing the arrangements of the electrons in the atmosphere of an atom.   This review will also show why there are discrete energy states and why the various levels occur in groups, and will relate various physical and chemical properties of the elements to their atomic structure.

175. Atomic Number; Isotopes.   Each atom of an element normally consists of (1) a positively charged central nucleus which contains almost all of the atomic mass, and (2) a surrounding atmosphere of electrons whose arrangement accounts for most of the element's chemical and physical properties.

As stated in Section 80, the atomic number Z is the number of electrons in the atmosphere of an atom when the atom as a whole is electrically neutral; it is therefore also a measure of the positive charge on the

nucleus.[U 406]   The mass of an atomic nucleus is almost all due to the protons or neutrons it contains.   These are particles which have each the same mass as a hydrogen nucleus.   Neutrons are electrically neutral, but each proton has a positive charge numerically equal to the charge on an electron.[VV]

Atoms having the same atomic number but different atomic weights are called *isotopes*.[X 617, VV]   For example, the nucleus of one kind of neon has a mass about 20 times, of the other kind 22 times, that of a hydrogen nucleus.   Yet both are neon, for the atomic number is 10 for both kinds. As neon occurs in nature the two are always mixed in such proportions as to give an average atomic weight of 20.2.   The atomic number of mercury is 80; its average atomic weight, 200.61, is the composite result of a mixture of seven isotopes, of atomic weights 202, 200, 199, 198, 201, 196, in that order of prominence.   Sodium's atomic number is 11; there is only one kind of sodium, of atomic weight 23.   Deuterium, or " heavy hydrogen " is an isotope of hydrogen.   See Table V.

176. Energy Levels As Related to Electronic Motions.   An atom of sodium vapor that is struck by a fast-moving electron may have the outermost one of the eleven electrons in its atmosphere driven away by the impact.   This will happen only if the kinetic energy of the on-coming electron is at least 5.12 volts, and the geometry of the collision favorable. If the resulting sodium *ion* is again hit by a high-speed electron, it may lose its tenth electron, now the outermost one.   The energy required for this operation is 47 volts.   With sufficiently elaborate equipment it is possible to rob atoms of two, three, four, or even more electrons.   If all the electrons are removed, the atom is said to be *stripped*.   It is experimentally feasible to strip atoms of low atomic number, such as lithium and beryllium.

An atom from which one or more electrons have been removed of course exerts an attraction on any lone electron that may enter its field. The motion of a roving electron during approach toward an atomic nucleus that has been stripped of its electronic atmosphere is remotely similar to that of a comet approaching the sun.   The electron, like the comet, loses potential energy and gains kinetic energy as it comes near to the heavier body.   In both cases mutual attraction draws the smaller body inward, while centrifugal force tends to drive it farther out.

Ordinarily a comet approaching from the remote heavens possesses at every point in its path enough kinetic energy to carry it back into infinite space again.   It will begin orbital motion around the sun only if, while in the sun's gravitational field, it loses by collision, friction, or otherwise, an appreciable portion of the kinetic energy that would normally carry it away again.   An electron *invariably* loses energy while

approaching a stripped atomic nucleus, for any electrical particle that undergoes acceleration radiates energy electromagnetically. Hence any electron that approaches at all closely is presumably captured and pursues a path that carries it repeatedly around the nucleus.

Since the electron is continually losing kinetic energy by radiation, its velocity and centrifugal force should become continually less, permitting the radius of its path to become smaller and smaller. According to this reasoning the path should be a flat spiral centering at the nucleus. But motion in any path that repeatedly encircles the nucleus is essentially periodic, and therefore subject to the quantum limitation that requires the action per cycle to be an integral multiple of h; see Section 86. This requirement, combined with that of centrifugal balance, restricts the possible radii to certain discrete values, and compels successive abrupt reductions in radius of path rather than a gradual decrease.

Each possible radius corresponds to a particular level on an energy-level diagram. In pursuing the orbital path corresponding to a given level the electron radiates no electromagnetic energy. A transition between levels corresponds to an abrupt shift to a smaller orbit, along with a sudden release of a single impulse of radiant energy. Thus the impulse-like nature of light and the intermittent nature of the movement toward the nucleus are mutually interdependent, both being quantitatively related to the action-unit h.

177. Energies of the Levels; One Electron in a Nuclear Field. The necessity for balance between centrifugal force and inverse-square-law attraction of the stripped nucleus requires, for a circular path, that [X 368]

$$\frac{m_e v^2}{r} = \frac{Ze \cdot e}{r^2} \qquad \text{(708 esu)}$$

where

$$Z \quad = \text{nuclear atomic number}$$
$$e \quad = \text{electronic charge}$$
$$m_e \quad = \text{electronic mass}$$
$$v \quad = \text{linear velocity in the path}$$
$$r \quad = \text{radius of the path}$$

The limitation of action to h-units (see Section 86) requires that

$$\int_{s=0}^{s=2\pi r} m_e v \, ds = 2\pi r m v = nh \qquad (709)$$

Here ds is incremental distance around the circumference, and n is any whole number. Equations (708) and (709) permit the elimination of v, and subsequent determination of the radius of the orbit for any particular value of n.

The energies rather than the radii are of interest, however. Potential energy W_p of the electron's position is its charge $(-e)$, multiplied by the potential, Ze/r, that is,

$$W_p = -\frac{Ze^2}{r} \quad \text{(ergs)} \tag{710 esu}$$

This assumes that the potential energy is zero at an infinite radius, decreasing to negative values as the radius becomes finite. Such an assumption is permissible because the zero of potential energy, like that of electric potential in a field or in a circuit, may always be selected arbitrarily. Kinetic energy W_k is of course $\frac{1}{2}mv^2$. It may be expressed by cancelling an r from Equation (708) and dividing by 2. This leads to the expression

$$W_k = \frac{Ze^2}{2r} \tag{711 esu}$$

The net energy W is the algebraic sum of the two contributing kinds, that is

$$W = W_p + W_k \tag{712}$$

$$= -\frac{Ze^2}{r} + \frac{Ze^2}{2r} \tag{713 esu}$$

or

$$W = -\frac{Ze^2}{2r} \tag{714 esu}$$

An expression for W which does not involve either the radius or velocity is obtained by multiplying together Equations (708) and (714), then substituting into the result a value for v^2r^2 obtained from Equation (709). These operations produce the following result:

$$W = -\frac{2\pi^2 m_e Z^2 e^4}{h^2} \frac{1}{n^2} \tag{715 esu}$$

All the factors except n and Z that appear here are fundamental physical constants, and n is any whole number. It is important to bear in mind that the energy calculated by using Equation (715) or Equation (714) *is the energy necessary to remove the electron to infinity.*

The atomic number of hydrogen is 1. Therefore if $Z = 1$ in Equation (715), and n is given values from unity up, the energies for the various levels in a hydrogen atom's energy-level diagram should and do result. Similarly if $Z = 2$, those for twice-ionized helium's diagram result, and if $Z = 3$, those applicable to thrice-ionized lithium, and so on. Experimental difficulties have limited laboratory verification of the predicted values of energies for these two-particle spectra to hydrogen, helium, lithium, beryllium, and boron. However, there is every reason to

believe that this theory is valid for elements of still higher atomic numbers. For example, the energy levels for a sodium nucleus that is in the process of capturing its first electron are presumably correctly obtained by using $Z = 11$ in Equation (715).

It has just been explained how and why each level of a single-electron energy-level diagram corresponds to a particular value of n, called the quantum number for that level. The greatest value n can have is infinity. This corresponds to the ionization level, for which the energy is zero. The least value n can have is 1. The atomic number for hydrogen is also 1; the use of $n = 1$ and $Z = 1$ in Equation (715) gives an energy in ergs that corresponds to 13.53 electron volts. This is the energy of the normal level. Since that for the ionization level is zero, the difference between the energies of these two extreme levels is just that of the lower one. Therefore the ionizing potential of atomic hydrogen should be 13.53 volts. This expectation has been experimentally verified.

178. Limitations of the Orbital Physical Picture. The locations of the energy levels for the various elements have been on the whole very well explained by extensions and modifications of this theory of electronic motions. The physical picture of planetary pellet-like electrons surrounding a nucleus that grows out of such treatment is called the " Bohr atom model." [X]

Some of the modifications employed are very striking. Elliptical orbits must be used; each electron is assumed to spin on its own axis, like the earth; also, the relativity change in mass of an electron, as its velocity varies around an elliptical orbit, is taken into account. Finally there is necessary a wave-mechanical denaturing of the electrons themselves, by which each one partakes of some of the aspects of a cloud-like wave train encircling the nucleus.

Throughout all these changes in the physical picture the essential features of the mathematical expressions remain the same. Although the simple picture of planetary pellet-like electrons is inaccurate in detail, it leads to useful approximations to the major aspects of atomic behavior.

179. Three-Dimensional Quantization. The circular motion assumed in the derivation of Equation (715) is a very specialized kind of orbital motion, in that the *angular* position only of the electron varies. In a sense circular motion is one-dimensional, for it is completely described by variations in a single angular coordinate.

Since space is three-dimensional, an electron's velocity may have three mutually perpendicular components. Each component of motion is periodic, so that an electron may have three different quantum numbers.

A similar situation was encountered in Sections 86 and 87 in the statement of the principles underlying the velocity distribution laws for a gas. For that purpose gas particles were thought of as being contained within a cubical box. The restraining walls of a box are plane surfaces; similarly, the barriers that prevent the escape of electrons from metals or metallic crystals are the work functions of plane surfaces. This suggested quantization in rectangular coordinates. The total quantum number n was expressed as

$$n = \sqrt{n_u{}^2 + n_v{}^2 + n_w{}^2} \tag{716}$$

because the total kinetic energy was found to be proportional to $n_u{}^2 + n_v{}^2 + n_w{}^2$. These three contributing n's, n_u, n_v, and n_w, are the quantum numbers that correspond to the velocity components in the x, y, and z directions respectively.

The force that acts on the electrons near a nucleus is centrally directed. This makes suitable a quantization in polar coordinates,[DD 22] illustrated in Fig. 9, rather than in rectangular coordinates. The three polar quantum numbers are n_r, n_θ, and n_ϕ, corresponding to variations in r, θ, and ϕ. In this polar system, r is the radius from an origin at the nucleus, θ is the angle between the radius and the polar axis, and ϕ is a measure of revolution around the polar axis.

The total quantum number is given by the relation

$$n = n_r + n_\theta + n_\phi \tag{717}$$

This definition is used because mathematical analysis shows that for periodic motion in the force field of a central nucleus the kinetic energy is inversely proportional to $(n_r + n_\theta + n_\phi)^2$.

The change from the relation $n = \sqrt{n_u{}^2 + n_v{}^2 + n_w{}^2}$ used in Chapter VIII to $n = n_r + n_\phi + n_\theta$ used here is necessary only because the electric forces on the electrons around an atomic nucleus are centrally directed, so induce motions of polar rather than rectangular nature. *In both situations n is chosen so that the energy is dependent on n^2.* It is possible to develop a rather satisfactory theory of electron energy distributions in large atoms by methods very similar to those employed in Chapter VIII for determining electron energy distributions in metals; however, a polar quantization must be used in such an atomic analysis.

The use of polar coordinates requires the selection of some given direction in space as that of the polar axis. In atomic study the direction chosen is always that of whatever magnetic field exists. The field may be small, perhaps much weaker than the earth's magnetic field. Yet in general there is at all points in space some magnetic field, so also some one direction that has unique magnetic significance.

180. The Exclusion Principle; Grouping of the Levels. A sodium nucleus that has captured just one electron must of course be expected to gather in ten more eventually, because eleven are required to neutralize the nuclear charge. As in the case of gas particles in a box, only one electron can have the motions that correspond to a given combination of quantum numbers [DD 144] (see Section 87). The final state of each one of the eleven must therefore be different from the states of all the others. There are correspondingly eleven separate, occupied energy levels in the normal state of a sodium atom.

The various orbits in which these eleven electrons travel are in general elliptical. The major axis of each orbital path depends on the total quantum number. The various ellipses have varying degrees of flatness, and lie in different planes; that is, they have different *eccentricities* and *orientations*.

The most important differences between them have to do with the lengths of the major axes and the eccentricities. These differences can be illustrated by a two-dimensional study, because each ellipse is two-dimensional in nature, the third dimension entering only in relation to the orientation. For this purpose let $n_\theta = 0$, only n_r and n_ϕ having values. Mathematical analysis shows that the resulting quantized motion is truly elliptical, with the nucleus at one focus of the ellipse. The length of the major axis is found to be directly proportional to n^2, and the energy inversely proportional to n^2. For given length of major axis, the minor axis is proportional to n_ϕ. When $n_\phi = n$, the major and minor axes are equal, so that the motion is circular. These relations are illustrated in Fig. 139. For a demonstration of them the reader is referred to Richtmeyer.[X 386]
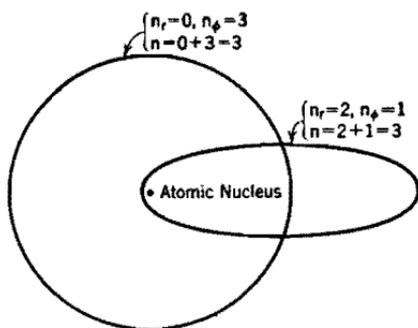
FIG. 139. Circular and elliptical orbits having the same total quantum number and therefore the same energy. The major axis of the ellipse is the same as the diameter of the circle. The minor axis of the ellipse is proportional to n_ϕ.

It has just been pointed out that the energy of an electron depends on the total quantum number, and of course a given total quantum number can result from a variety of combinations of the component quantum numbers. Therefore more than one electron can have the same energy. For example, an elliptical orbit for which $n_r = 2$, $n_\phi = 1$, calls for the same energy as the elliptical one which $n_r = 1$, $n_\phi = 2$, and the circular one for which $n_r = 0$, $n_\phi = 3$.

This explains why the energy levels occur in groups, as illustrated in Fig. 135b. The total quantum number is the same for all the levels in a group, and the energies are approximately the same within low-lying groups. The small differences in energy between the levels within a group result from minor differences in the energy contributions from the different components of motion.

181. Shells. The different quantum-number groups of electrons are often spoken of as *shells*. The reason for this is that the various groups are located at progressively greater distances from the nucleus, Equation (714) indicates that the mean orbital radius should be expected to vary approximately inversely as the energy, therefore directly as n^2. The proportionality between the length of the major axis and n^2 bears this out.

For reasons detailed later in Section 183, a shell whose total quantum number is n contains $2n^2$ possible energy states, or energy levels. This is illustrated both in Figs. 68, page 163, and in Fig. 135. Sodium has two electrons in the first, closest-in shell with lowest-lying levels, eight electrons in the second, and one lone one as a start in the third.

The path of the lone third-shell electron is a very flat, elongated ellipse. As illustrated in Fig. 139, such a path brings the electron traveling along it much closer to the nucleus than does a circular path. In fact, this outer electron's orbit is at one end *well inside the second shell of electrons*. It is said to pursue a *penetrating* orbit.

The ten electrons of the first two shells serve, for regions beyond them, as an effective screen of $\frac{10}{11}$ of the nuclear charge. Therefore for the portion of this elongated path that is *outside* the second shell, the eleventh electron lies in a central force field due to a charge of $+e$. For the portion of the path *inside* the second shell it lies in a much stronger field. The average effect of the nucleus on this penetrating electron is therefore greater than it would be on a nonpenetrating one.

Thus the *effective nuclear charge* [U 431] is greater for a penetrating than for a nonpenetrating orbit. In Equation (715), Z stands for the nuclear charge. Evidently a penetrating electron must have, according to this equation, a greater negative energy, so occupy a lower level, than a nonpenetrating one having the same total quantum number.

For example, if sodium's outermost electron normally pursued a nonpenetrating path it should be possible to calculate the ionizing potential of sodium by using $n = 3$ and $Z = 1$ in Equation (715). Such treatment predicts an ionizing potential of only 1.51 volts. The actual ionizing potential is obtainable from this equation by using $n = 3$, $Z = 1.84$. This indicates that for the actual penetrating orbit pursued in the normal state of the atom the effective nuclear charge is 1.84.

One of sodium's excitation levels has an energy of 1.51 electron volts (measured from the ionization level). This level presumably corresponds to a nonpenetrating orbit for which $n = 3$.

182. Relations between Electron Arrangement and Chemical and Physical Properties of the Elements. A static atom model, of which Fig. 68 is an example, is useful in explaining certain chemical and physical properties of the elements.[X 425] This static concept is not in conflict with the notion of orbital motion. For example, the eight electrons in sodium's second shell may be in rapid periodic motion, yet on the average react to external forces as though located at the corners of a cube.

The relation between electron arrangement and chemical properties can be illustrated by the behavior of salt in solution. The third shell of chlorine has seven electrons (see Table V). Therefore chlorine exposes to its surroundings a cube with one corner missing. When sodium chloride is dissolved in water it dissociates into sodium ions (Na^+) and chlorine ions (Cl^-). Each sodium atom gives up its outer electron in order to permit a chlorine atom to fill in the eighth corner of its cube. Both ions present completed cubes to external forces. The atoms of the inert gases, neon, argon, krypton, xenon, all noted for their chemical inactivity, present completed cubes to external surroundings.

Chemical valence is related to electron array. Thus sodium and chlorine are respectively one over and one short of being able to present completed cubes to external surroundings, and the valence for both elements is one. For magnesium the excess is two, and for sulphur the deficiency two; hence the valence for both elements is two.

183. Magnetic Quantization: $2n^2$. This question naturally arises: why can the nth shell contain $2n^2$ electrons, but no more?

The periodic motions of an electron in the central force field around an atomic nucleus are a composite of:

1. Variations in r, that is, in closeness to and remoteness from the origin at the nucleus; the corresponding quantum number is n_r.

2. Continuous rotation about the polar axis, that is, a continuous increase in ϕ. (See Fig. 9.) The corresponding quantum number is n_ϕ. Since the electron carries an electric charge, this continuous rotational motion constitutes an electric current and produces a magnetic field. The rotation may be either clockwise or counterclockwise, so that the magnetic field may have either of two opposite polarities. Some magnetic field exists at all points in space, and the direction of the polar axis is that of the magnetic field. The magnetic energy due to the combination of the external field with that produced by the electron is of course different for the two directions of electron motion. For this reason there can be *two* energy states for each numerical value of n_ϕ, except when $n_\phi = 0$. Each state can be occupied by an electron. Because n_ϕ is magnetic in nature, it is often called m_l. The significance of the subscript l will appear later.

3. Oscillations of the polar angle θ, which must always take place between 0 and π (radians). Since θ does not increase continuously, there are not two oppositely directed kinds of θ motion with differing energies. The quantum number is n_θ.

The mathematical statements of these quantizations are as follows:

$$n_r h = \oint m_e v_r \, dr \tag{718}$$

$$n_\phi h = \oint m_e r^2 \sin^2\theta \frac{d\phi}{dt} \, d\phi \tag{719}$$

$$n_\theta h = \oint m_e r^2 \frac{d\theta}{dt} \, d\theta \tag{720}$$

The results of spectrographic study indicate that each electron spins on its own axis, as does the earth. Such spinning of electric charge produces a magnetic field. The rate of spin, as measured by action in h-units, is always the same, but the magnetic field may be oriented either with or against the polar-axis field. The spin quantum number m_s is always $\frac{1}{2}$, so that the change from one orientation to the opposite is accompanied by a shift of a single h-unit. For any given array of the quantum numbers n_r, n_ϕ, and n_θ there are two half-quantum spin states, one for each spin orientation. That is, m_s may be either $\frac{1}{2}$ clockwise, or $\frac{1}{2}$ counterclockwise.

Table X lists the various combinations of n_r, n_ϕ, n_θ, that are possible for each of several total quantum numbers.[DD 145] It also indicates how the $2n^2$ levels in a shell differ from one another. The symbol l is used for the quantity $m_l + n_\theta$ (identical with $n_\phi + n_\theta$). Thus m_l can never be greater than l, and is equal to l when $n_\theta = 0$. This usage is a helpful simplification in dealing with circular or elliptical orbits which are two-dimensional.

It has been found that actual atomic behavior, as recorded by spectral lines, is best approximated by using, in numerical formulas, values that are slightly greater than the integral values 0 1, 2, 3, etc., in place of l. For some purposes $l + \frac{1}{2}$ works best, for others $\sqrt{l(l + 1)}$. However, these best approximations never require the use of numerical values greater than n. If the numerical value to be used in place of l is slightly greater than l, but does not exceed n, it is apparent that l does not exceed $n - 1$.

The restriction, in Table X, of values of l to not more than $n - 1$ is in accord with the situation outlined in the previous paragraph, as well as in accord with known placement of electrons in shells. The usage outlined in the previous paragraph also makes the use of the value 0 for l, apparently requiring the electron to collide with the nucleus, not wholly unreasonable, because the numerical value used is not quite zero.

Within each shell, the energy states available to the electrons are classified into subgroups, according to the values of l. Letters are ordinarily used to describe the l's according to the following code:

$$\text{Letter symbol:} \quad s \quad p \quad d \quad f \quad g \quad h \quad i$$
$$\text{Value of } l: \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$$

By reference to Table X it will be seen that, in any shell, a completed s subgroup contains 2 electrons, a completed p subgroup 6 electrons, a completed d subgroup 10 electrons, and so on.

In general, an atom whose outermost aspect is a completed subgroup is relatively stable chemically. Thus the stability of atoms for which the static models have cubical outer aspects (see Section 182) is due to the fact that just enough electrons are present to fill the p subgroup.

There are no elements whose outer aspects are completed third shells. Thus argon has 8 electrons in the third shell (see Table V) but the next element, potassium, places its additional electron in the fourth shell. The reason for this " jump " is explained in the following paragraph.

The first fourth-shell electron pursues a penetrating orbit, for its angular momentum is small ($l = 0$) while its radial motion is large ($n_r = 4$). On the other hand, the last 10 third-shell electrons have more nearly circular, nonpenetrating orbits ($l = 2$, $n_r = 1$). It was pointed out in Section 181 that electrons in penetrating orbits are exposed to relatively large effective nuclear charges, and have relatively large (negative) energy values. The effective nuclear charge of the first fourth-shell orbit is enough larger than that of the last subgroup of third-shell orbits to more than compensate for the increase in total quantum number from 3 to 4, in the establishment of energy values. Therefore the energy level for the first fourth-shell energy state lies below that for the last 10 third-shell states, and so is occupied before the last 10 third-shell levels.

If the spectrum of sodium is analyzed with reference to Table X, it appears that the normal level of sodium's energy-level diagram is one for which $n = 3$, $n_r = 3$, $l = 0$. The usual sodium spectrum corresponds to an energy-level diagram having four levels for which $n = 3$. Their l-values are 2, 1 (a pair), and 0, as marked in Fig. 136b.

According to Table X there should be six levels having $n = 3$, $l = 1$. Yet only two appear in the diagram. Actually all of the six states of motion can exist, but their energies all have either the 24476 or the 24493 value, so that they are shown as a pair. If a strong magnetic field is applied to a discharge while the spectral lines are being observed, the differences in magnetic energies of the clockwise and counterclockwise rotations become more pronounced, and each member of the pair splits

up into three distinguishable levels. This spectral change is called the
" Zeeman effect."[DD 69]   Many other levels are similarly affected. A
somewhat similar behavior in the presence of a strong electric field is
called the " Stark effect."[DD 78]

184. Action and Angular Momentum. One of the concepts that is
very useful in the analysis of electronic motions is that of *total angular
momentum*, symbol J.[DD 54]   The only angular momentum possessed by
sodium's eleventh electron in the normal state is that due to electron
spin, for $l = 0$.   In the $n = 3$, $l = 1$ state it has angular momentum due
both to *rotation* and *spin*.   The total angular momentum J is the vector
sum of these.   Vectorial summation of angular momenta is accomplished
by the use of vectors pointed along the axis of the rotation, in the direc-
tion of a right-handed screw's advance.   If the motions of more than
one electron are effective in causing the differences between the levels,
*the total angular momentum is the vector sum of the momenta for all the
active electrons.*

The actual vector summation is, in atomic analysis, always made
in terms of quantum units.   This is possible because action and angular
momentum have the same dimensional formula, and are subject to
corresponding action limitations.   To illustrate this the action integral
for circular motion, which is

$$\int_{s=0}^{s=2\pi r} m_e v \, ds = 2\pi r m_e v = nh \tag{721}$$

may be rewritten in polar terminology, in the form

$$\int_{\phi=0}^{\phi=2\pi} I\omega \, d\phi = 2\pi I\omega = n_\phi h \tag{722}$$

Here I, ω, ϕ, signify moment of inertia, angular velocity, and rotational
angle.   Equation (722) can be rewritten as an equation for the angular
momentum $I\omega$, as follows:

$$I\omega = \frac{n_\phi h}{2\pi} \tag{723}$$

In this form it shows that the action limitation in general requires
angular momentum to be an integral multiple of $h/2\pi$.

With this relationship in mind, it is intelligible to say that the total
angular momentum of sodium's normal level ($n = 3$, $l = 0$) is half of

an h-unit, or $\frac{1}{2}\left(\frac{h}{2\pi}\right)$, for the only angular motion is that due to spin.

On the other hand, the total angular momentum corresponding to the $n = 3$, $l = 1$ level is, in $\dfrac{h}{2\pi}$ units, either

$$J = 1 + \tfrac{1}{2} = 1\tfrac{1}{2}$$

or $\qquad\qquad J = 1 - \tfrac{1}{2} = \tfrac{1}{2} \qquad\qquad\qquad (724)$

These two forms correspond to the two opposite possibilities in the angular combination of these motions. The term values of these two levels are respectively 24492.83 and 24475.65 waves per centimeter. For $n = 3$, $l = 3$ there are again two possibilities, $J = 2 + \tfrac{1}{2} = 2\tfrac{1}{2}$ and $J = 2 - \tfrac{1}{2} = 1\tfrac{1}{2}$. The energy values for these two levels are normally indistinguishable experimentally.

Sodium is said to have a one-electron spectrum, in that the energy changes all result from variations in the motions of one electron operating outside a stable group. It is typical to have levels for such a spectrum occur in pairs, or *doublets*. By contrast mercury is said to have a two-electron spectrum, having two electrons just outside a stable group of 18. In a two-electron spectrum the total angular momentum is the vector sum of the h-unit values of the two electrons, and levels occur either in *singlets* or in *triplets*, that is, either alone or in three's.[DD][VII]   See Sections 191 and 195.

PROBLEMS

CHAPTER XV

1. Calculate from theoretical considerations, of the type described in Section 177, the energy in electron volts required to remove the second of helium's electrons, also that required to remove the third of lithium's electrons.

2. What is the effective nuclear charge for the orbit pursued by the outermost electron in a sodium atom in the lowest excited state?   In a sodium atom in the state corresponding to the 8,248 level?

CHAPTER XVI

ENERGY LEVELS FOR PARTICULAR ELEMENTS

185. Identification of Levels. The levels of an atom's energy-level diagram are identified quantitatively by their *term values*, in waves per centimeter, and qualitatively by a symbolic notation. The notation used in this text is the same as that used by Bacher and Goudsmit.[EE] In general the information conveyed by the notation includes: (1) descriptions of the *electron configurations*, that is, statements of the values of n and l for the various electrons, (2) indications as to the various ways in which angular momenta of the individual electrons combine to give different values of total angular momenta, so different energy values, (3) numerical values of the total angular momentum J, and (4) information indicating between what levels transitions may occur.

186. Energy Levels for the Arc Spectrum of Sodium. Table XI lists the term values and symbolic notations for the excited states of a sodium atom, which are sometimes described as the energy states of the NaI spectrum.[EE 307] They are also referred to as the energy states of the *arc spectrum* of sodium, because transitions between them produce the light radiated by a sodium vapor arc. The arc spectrum of sodium is said to be a one-electron spectrum, because the differences between the energies of the levels are due entirely to changes in the angular momenta and radial movements of one lone third-shell electron outside a stable 2 + 8 array.

If this outer electron is entirely removed, further contributions of energy operate on one of the eight second-shell electrons. The resulting energy states are those of the NaII spectrum, also called the *spark spectrum* of sodium. This phraseology is used because much of the light radiated by an electric spark in sodium vapor is due to transitions between the NaII energy states. The spark spectrum of sodium is similar to the arc spectrum of neon (NeI), for both are produced by 2 + 8 electron arrays. The ionization level of the NaI spectrum is the normal level of the NaII spectrum.

Table XI has four columns, with entries describing, for each level, the *term value*, the *configuration*, the *symbol*, and the *J-value*, with meanings as follows:

187. Term Values. The numbers in the term value column measure, in waves per centimeter, the distances of the levels below the ionization or zero level. The *reciprocal* of the difference between two term values

gives the wave length, in centimeters, of the radiation corresponding to a transition between those levels. (In some spectra the ionizing potential is not accurately known; then, and sometimes for other reasons, the normal level is used as the zero of reference. Waves per centimeter are then reckoned upward from the normal level rather than downward from the zero level.)

188. Configuration. The *numerals* in the configuration description give values for the total quantum numbers n for the various electrons. The *letters* describe the angular quantum numbers $l(= m_l + n_\theta)$, according to the following code:

Letter symbol:	s	p	d	f	g	h	i
Value of l:	0	1	2	3	4	5	6

The exponents to the l-value letters indicate the number of electrons having the specified values of n and l. For example, the description of sodium's normal, unexcited configuration can be explained as follows:

The whole: $1s^2\ 2s^2\ 2p^6\ 3s$

Its parts: $1s^2$: two electrons for which $n = 1, l = 0$
(these constitute the first shell)

$2s^2$: two for which $n = 2, l = 0$
(these begin the second shell)

$2p^6$: six for which $n = 2, l = 1$
(these complete the second shell)

$3s$: one for which $n = 3, l = 0$
(this one is alone in the third shell).

The configurations for the various levels in any one spectrum are ordinarily alike except for the last few electrons. It is usually necessary to give the descriptions only for those which do change. For example, in sodium's arc spectrum only the eleventh electron changes its n and l values, so that it is sufficient to abbreviate to $3s$ the description of the configuration of the normal level, and to $3p$ that for the 24493 and 24476 levels. The yellow light characteristic of sodium is due to transitions from the $3p$ to the $3s$ configuration.

189. Symbols. Within a given configuration, the various electronic angular momenta may combine vectorially in a number of different ways to give a variety of energies, so various levels. To each of these there corresponds a symbol, found in the " symbol " column. In text references the values of J are usually appended as a subscript to the symbols, but they appear in a separate column in the table. Thus the complete symbol for sodium's $3s$ level is $^2S_{\frac{1}{2}}$; those for the two $3p$ levels are $^2P_{\frac{1}{2}}^{\circ}$ and $^2P_{1\frac{1}{2}}^{\circ}$.

The levels of any spectrum can be segregated into an *odd* group and an *even* group, the *odd* terms being identified by the superscript °. Thus the $^2S_{\frac{1}{2}}$ level is an even one, the $^2P_{\frac{1}{2}}°$ and $^2P_{1\frac{1}{2}}°$ levels odd ones. Transitions must in general take place from odd levels to even levels, or vice versa; never, or very rarely indeed, do transitions take place between two even or two odd levels.

The upper prefix [2] to the symbols for the sodium terms indicate that they belong to sets of terms which for the most part are made up of close pairs, called doublets. One-electron spectra contain doublets, two-electron spectra singlets *and* triplets. The expected "multiplicities" increase systematically with the number of electrons involved.

The complete qualitative description of a level or term includes a statement of its configuration followed by the symbol for the individual level. Sodium's normal state is a $1s^2\,2s^2\,2p^6\,3s\;^2S_{\frac{1}{2}}$ level; the description may be abbreviated to $3s\;^2S_{\frac{1}{2}}$.

190. The Meanings of Symbols. The tendency of individual momenta of an outer group of one, two, three, or more electrons to combine according to a certain definite system forms the basis for the choice of symbols. When the system is followed exactly the atom is said to exhibit complete " Russell-Saunders coupling." Complicated spectra follow this system only approximately or not at all. In neon's arc and sodium's spark spectra (NeI and NaII) the deviations from Russell-Saunders coupling are so great that the usual symbolism cannot be used. An arbitrary identification of terms in order of energy values is adopted for these spectra.

The ideal Russell-Saunders combination pattern is as follows:

(1) The individual *l*-values add *vectorially* in as many different ways as will give integral results. The resultant, called L, may be

$$0, \quad 1, \quad 2, \quad 3, \quad 4, \quad \text{etc.}$$

in units $h/2\pi$. L is correspondingly described by the capital letters

$$S, \quad P, \quad D, \quad F, \quad G, \quad \text{etc.}$$

This is suggestive of the *l*-value symbolism. Small letters code the values of angular momenta for individual electrons; capital letters similarly code the vectorial summation of the *l*-values for an electron group. The individual values and their vector summation must be integral multiples of $h/2\pi$ units.

(2) The individual spin values add vectorially in as many different ways as give (*a*) integral values if there is an even number of spin values, (*b*) half-integral values if there is an odd number of spin values. The resultant is the total spin-vector, sometimes called S. Care must be taken not to confuse this S with the description of L when of value zero.

(3) The vectors L and S combine *vectorially* to give J-values. They combine in as many ways as will result in: integral values if the spin-vector is a whole number, half-integral values if the spin-vector is a half number.

191. Symbols for Sodium. Suppose this combination pattern is applied to sodium's two $3p$ levels, 24493 and 24476. Sodium has a one-electron spectrum, so that just one l-value, that of the eleventh electron, contributes to L. Therefore when the value of l is $p(= 1)$, that of L is $P(= 1)$. The spin of only the eleventh electron contributes to S; the value of S is therefore $\frac{1}{2}$. J may then be either:

$$J = 1 \oplus \tfrac{1}{2} = \tfrac{1}{2}; \qquad \begin{array}{c} L = 1 \\ \hline \longrightarrow \\ \hline \longrightarrow \longleftarrow \\ J = \tfrac{1}{2} \quad S = \tfrac{1}{2} \end{array} \quad {}^{2}P_{\frac{1}{2}}{}^{\circ} \quad 24493$$

or $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (725)

$$J = 1 \oplus \tfrac{1}{2} = 1\tfrac{1}{2}; \begin{array}{c} L = 1 \quad S = \tfrac{1}{2} \\ \hline \longrightarrow \longrightarrow \\ \hline \longrightarrow \\ J = 1\tfrac{1}{2} \end{array} \quad {}^{2}P_{1\frac{1}{2}}{}^{\circ} \quad 24476$$

These two terms constitute a *doublet*. The levels of one-electron spectra generally occur in pairs (doublets), one for each manner of addition of the spin-vector. Sometimes the separation between the levels of a pair is so small as to escape observation.

The symbol for sodium's normal level is ${}^{2}S_{\frac{1}{2}}$. This suggests that it should be one of a pair of doublets, and indicates that its J is $\frac{1}{2}$. As for all one-electron levels, the value of L is the same as that of l; here it is $S(= 0)$. The value of J is as usual $L + S$, which, however, gives in this case *only one* half-integral result, that is,

$$J = 0 \oplus \tfrac{1}{2} = \tfrac{1}{2} \qquad\qquad {}^{2}S_{\frac{1}{2}} \ 41449.0 \qquad\qquad (726)$$

$\frac{1}{2}$ can add vectorially to 1 to give two half-integral results, but $\frac{1}{2}$ added vectorially to zero gives only $\frac{1}{2}$. One member of the expected doublet does not exist. Yet the symbol is given the doublet marking because its spin-vector is such as to lead normally to doublet expectation; it belongs to a doublet system, and is so marked. All the levels of sodium's arc spectrum belong to doublet systems.

192. J-values. For convenience, and because of their importance, values of total angular momentum of the active electron or electrons are entered in separate columns in Tables, XI, XII, XIII, and XIV, rather than being appended as subscripts to the energy-level symbols. The significance of these J-values was briefly discussed in the last section of the previous chapter, and their utility is indicated in the following section.

193. Selection Principles. Some transitions between levels are *permitted*, others *forbidden*. Spectral theory explains why this is true,

but the explanation is too intricate to present here. The *selection rules* are in general as follows: [EE 16]

(1) Transitions occur only from odd to even or even to odd levels. For even levels, the arithmetical sum of all the *l*-values is even; for odd levels, it is odd.

(2) In any permitted transition *J* changes by +1, 0, or −1, except that transitions requiring changes from *J* = 0 to *J* = 0 are forbidden.

(3) In any permitted transition the *l*-value for the shifting electron changes by either +1 or −1.

(4) The *n*-value can change by any amount.

(5) Transitions which require shifts in the values of *n* or *l* for more than one electron are very infrequent and in most cases forbidden.

There are additional selection rules which still further contrast permitted and forbidden transitions. Some, like the fifth rule above, are only approximate and serve chiefly to differentiate between expected intensities of the corresponding spectral lines.

Some spectra follow the Russell-Saunders combination-pattern very closely. In such spectra the resultant spin-vector *S* remains unchanged, and the *L*-vector changes by +1, 0, or −1, in the transitions which produce the most intense lines.

194. Series of Levels in Sodium. The transitions marked on Fig. 136*b* have been arranged so as to group together those that correspond to various important *spectral series*. The transitions that give rise to a series of spectral lines terminate at a common level. They originate at members of a *series of levels* which have successively different values of *n*, but identical values of *l* and of *J*, and therefore identical symbols.

The names given to the NaI series of spectral lines are partly descriptive, partly misleading. The " principal series " contains the most pronounced lines, the first one of them being the yellow light characteristic of sodium. The names " diffuse " and " sharp " indicate the character of the spectral lines of the respective series. The transitions of the " diffuse series " originate at levels that are very close doublets (12276, $^2D_{1\frac{1}{2}}$, $^2D_{\frac{1}{2}}$, for example). The transitions of the " sharp series " originate at levels that are not doublets, even though they bear doublet symbols (8248, $^2S_{1\frac{1}{2}}$, for example), for reasons explained in the discussion following Equations (726). There is nothing especially fundamental about the series so named.

195. Mercury. Table XII contains the various descriptive names, numbers, and symbols, and Fig. 140 the energy-level diagram, for the arc spectrum of mercury.[EE 227] This is a two-electron spectrum, for there are two active ($6s^2$) electrons outside a stable 2 + 8 + 18 + 32 + 18 array.

The occurrence of singlets and triplets, typical of two-electron spectra, is well illustrated by the four $6s6p$ levels. The two active electrons have

the l-values $s(=0)$ and $p(=1)$.   Since 1 and 0 can add vectorially to give only a single positive result, $L = 1(=P)$.   The total spin-vector S may be either $\frac{1}{2} \oplus \frac{1}{2} = 1$ or $\frac{1}{2} \oplus \frac{1}{2} = 0$.   Vector addition of the value 1 for S to the value 1 for L gives three possible values for J, corresponding to the following triplets:

$J = 1 \oplus 1 = 0;$

$^3P_0^{\,\circ}$ 46536

$J = 1 \oplus 1 = 1;$

$^3P_1^{\,\circ}$ 44769

(727)

$J = 1 \oplus 1 = 2;$

$^3P_2^{\,\circ}$ 40138

If the value 0 for S is used, there can be just one J-value, corresponding to the following *singlet:*

$J = 1 \oplus 0 = 1;$ $J = L = 1$ $^1P_1^{\,\circ}$ 30113   (728)

Sometimes a " triplet " level occurs alone.   For example, the $6s7s$ configuration has for L the value zero, and has only two terms, one for each value of S, as follows:

$J = 0 \oplus 0 = 0$ $^1S_0$ singlet system

$J = 0 \oplus 1 = 1$ $^3S_1$ triplet system

(729)

This latter term is symbolized as a triplet because unit spin-vector can ordinarily add to the L-vector in three different ways.   Thus in mercury most of the other levels for which the spin-vector is 1 occur in three's.   This one occurs alone because zero added to unity can give only one vectorial result, by contrast with three vectorial integral results when added to any integral value of L other than zero.

196. Mercury Metastable States.   The levels 46536 $6s6p$ $^3P_0^{\,\circ}$ and 40138 $6s6p$ $^3P_2^{\,\circ}$ are of particular interest because no transitions can take place downward from them.   Neither of these permits a shift to the normal level, for in one J would have to change from 0 to 0, which is forbidden, in the other from 2 to 0, likewise ruled out.   There cannot be a shift from the $6s6p$ $^3P_2^{\,\circ}$ level to either of the two lower $6s6p$ levels

(46536 and 44769) because there would be no change in l in such transitions.

Therefore a mercury atom that through accident of collision, or by transitions from higher levels, reaches either the 46536 or the 40138 state remains there until impact with another particle occurs.   There
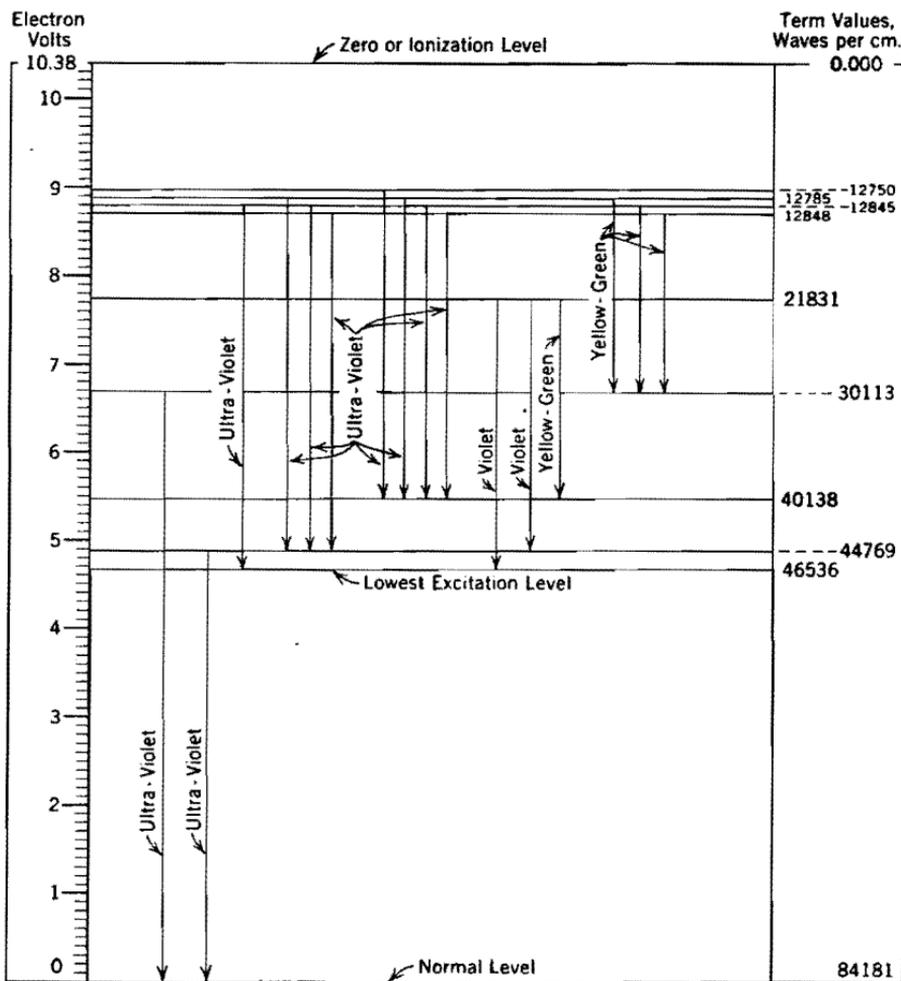
Fig. 140.   Energy-level diagram for a mercury atom.   (See Table XII.)

can be no escape from these levels to the normal level by the release of radiant energy.   Since the likely duration of these states is many thousands of times that of any of the other levels, they are called *metastable states*.[86, DD94]

Atoms in metastable states have important effects in electric arcs and glow discharges.   For example, mercury may be ionized in two

steps, the first step creating a metastable state, the next causing a further shift up to the ionized level.   An atom so ionized has received its 10.38 electron volts of energy in two successive steps, neither step requiring more than 6 volts.

An atom in a metastable state carries a substantial burden of energy that can be released by contact.   For example, if a metastable atom whose energy is 4.65 volts comes into contact with a metal surface that has a 3-volt work function, there is a substantial likelihood that an electron will leave the surface.   The electron may in fact be ejected with a kinetic energy equal to $4.65 - 3.0 = 1.65$ electron volts.   The mercury atom returns to its normal state simultaneously with the electron ejection.   Similarly a 5.43-volt metastable atom can ionize a sodium atom by accidental contact, for sodium's ionizing potential is only 5.12 volts.   If two 5.43-volt metastable atoms collide, one may be ionized while the other returns to the normal state.

197. Negative Term Values.   Most of mercury's energy levels represent situations in which only one electron has been shifted from the normal $6s$ condition.   However, two $6p^2$ levels have been observed, although they occur only infrequently.   It takes *more* energy, by about one electron volt, to produce either of these $6p^2$ levels than to produce ionization.   They therefore lie above the zero level, and are said to have *negative* term values, -7860 and -9798.

198. Light from Mercury Vapor and from Sodium Vapor; Fluorescence.   As indicated in Fig. 140, the arc spectrum of mercury includes both violet and yellow-green transitions.   However, the violet ones occur much oftener than the yellow-green ones.   The violet light produced by a mercury-vapor arc is therefore usually so intense as to mask completely the yellow-green light, in spite of the latter's much greater relative lighting value per unit of energy (see Fig. 137).   However, by using an enclosing tube made of amber glass that restricts the emergence of the violet light but not of the yellow-green light, a mercury discharge may be made to give green illumination.

The ultraviolet radiations from a mercury arc are effectively stopped by the walls of the enclosing tube, if the tube is made of glass.   Quartz is transparent to ultraviolet radiation.   It is the near ultraviolet radiation from a mercury arc that has therapeutic value, so that mercury-arc tubes employed for medical purposes are made of quartz.   The 1850 A radiation from mercury arcs is rather strong, and harmful physiologically.   For this reason quartz mercury-vapor lamps should be used only as directed by a competent physician.

The transitions in the mercury spectrum that produce visible light originate rather high up in the mercury energy-level diagram, Fig. 140.

For example, that from 21831 to 44769, (violet, 4370 Å, 2.82 volts) can occur only after at least a 7.86-volt excitation. Suppose that an atom in a mercury arc receives by electron impact 7.86 electron volts of energy, lifting it into the 21831 energy state. It may subsequently radiate visibly 2.82 electron volts of this during transition to the 44769 energy state. The energy remaining, 4.86 electron volts, is then radiated in the ultraviolet (2530 Å), giving " light " which is not useful from the standpoint of illumination. It may be said that the *luminous efficiency* of this cycle of operations from normal state to normal state again is 2.82/7.68 = 0.367, or 36.7 per cent.

Mercury's metastable states add to the overall efficiency of a mercury arc as an illuminant, because many light-producing cycles may begin and end at metastable states. For example, suppose a 2.93-volt electron impact lifts an atom from hesitation in the 46536 metastable state to the 21831 condition. The atom may then immediately return to the 46536 state for further hesitation, then repeat the experience. Each such cycle produces 2.93-volt (violet) light, and has 100 per cent luminous efficiency. Of course there is no assurance that the descent from the 21831 level will take place along the transition leading back to the level from which the cycle started. If the descent happens to follow the transition leading to the 44769 energy state, further nonluminous descent to the normal state immediately occurs. There can be no appreciable hesitation in the 44769 state, so that there is no chance for electron impact while there.

The more opportunities there are for high-efficiency luminous cycles in a discharge, the greater overall efficiency is possible. For this reason sodium vapor gives excellent promise as an illuminant.[86] The light-producing cycle consisting of electrical elevation from sodium's normal level to the 24493 level or its twin, with immediate return, is 100 per cent efficient and terminates at the normal state. Thus sodium operates very efficiently to and from its normal state, but the most efficient processes in mercury (also in neon) require sufficient energy input to the discharge to favor high concentrations of metastable particles. Some practical difficulties have arisen in using sodium-vapor light sources, because rather high operating temperatures are necessary to vaporize sodium and keep it vaporized. These difficulties have been overcome to a considerable extent.

The yellow light from sodium does not provide pleasing general illumination, but has been found very useful for street lighting. More or less successful attempts have been made to convert sodium yellow, and other kinds of monochromatic light from gaseous discharges, into

colors suitable for general illumination by having them pass through layers of fluorescent material coated on discharge-tube walls.

In conversion of light by fluorescence individual photons are absorbed by molecules of the fluorescent substance, and part of the energy so received is immediately reradiated. There is always some energy loss in the conversion, so that the energy per photon of the reradiated light is always less than that of the incident light. This explains why it is reasonable to expect to be able to scatter sodium yellow light into the orange and red parts of the visible spectrum, but not into the blue and green parts. Ultraviolet radiation can be converted into visible light by fluorescence, but infrared radiation cannot.

199. Neon. Table XIII contains the term values, and Fig. 141 an energy-level diagram, for the arc spectrum of neon.[86, EE 312] In order to avoid confusion, only the transitions that produce visible light are shown, except that two ultraviolet transitions terminating at the normal level are included. The features of special interest are as follows:

(a) There is so little adherence to Russell-Saunders coupling that the levels within each configuration are numbered consecutively, instead of being given coded letter symbols.

(b) Not more than one electron at a time ever moves out into the third shell, so that the $2p^5$ group is a part of all the configurations. The electron array $1s^2\ 2s^2\ 2p^5$, sometimes called the ion on which the levels are built, itself has a variety of energy states *similar to those in a one-electron spectrum*. This behavior illustrates the general principle that a subgroup lacking just one electron of completion behaves in general like a single electron. The angular momenta (magnetic quantum numbers m_l, Table X) of the electrons within any completed subgroup combine to give zero total angular momentum. The omission of the final electron from‾the group, therefore, leaves the resultant of the m_l's at just the value which that one would produce if it were alone in its subgroup. The energy levels of the $2p^5$ group in neon may thus be said to be those due to the ghost of the missing electron. Such a one-electron set of levels consists of doublets (see Section 191); the important doublet of the $2p^5$ neon ion consists of the following:

$$1s^2\quad 2s^2\quad 2p^5\ (^2P_{1\frac{1}{2}})$$
$$1s^2\quad 2s^2\quad 2p^5\ (^2P_{\frac{1}{2}})$$

Thus in the 25671.65 $2p^5$ $(^2P_{1\frac{1}{2}})$ $3p$ 1_1 energy state of neon there is a $3p$ electron outside of a $^2P_{1\frac{1}{2}}$ ion, while in the 23157.34 $2p^5$ $(^2P_{\frac{1}{2}})$ $3p$ 7_1 state there is a $3p$ electron outside a $^2P_{\frac{1}{2}}$ ion.

(c) There is an older notation for the neon spectrum; the older notation is included in Table XIII because many references to the neon spectrum will be found in which the older notation is used.

(d) The transitions between the 3p and the 3s levels produce most of the red light that is characteristic of neon.

(e) There are the following two important metastable states:

$$39887.61 \quad 2p^5 \; (^2P_{1\frac{1}{2}}) \; 3s \; 1_2{}^\circ$$

$$39110.81 \quad 2p^5 \; (^2P_{\frac{1}{2}}) \; 3s \; 3_0{}^\circ$$
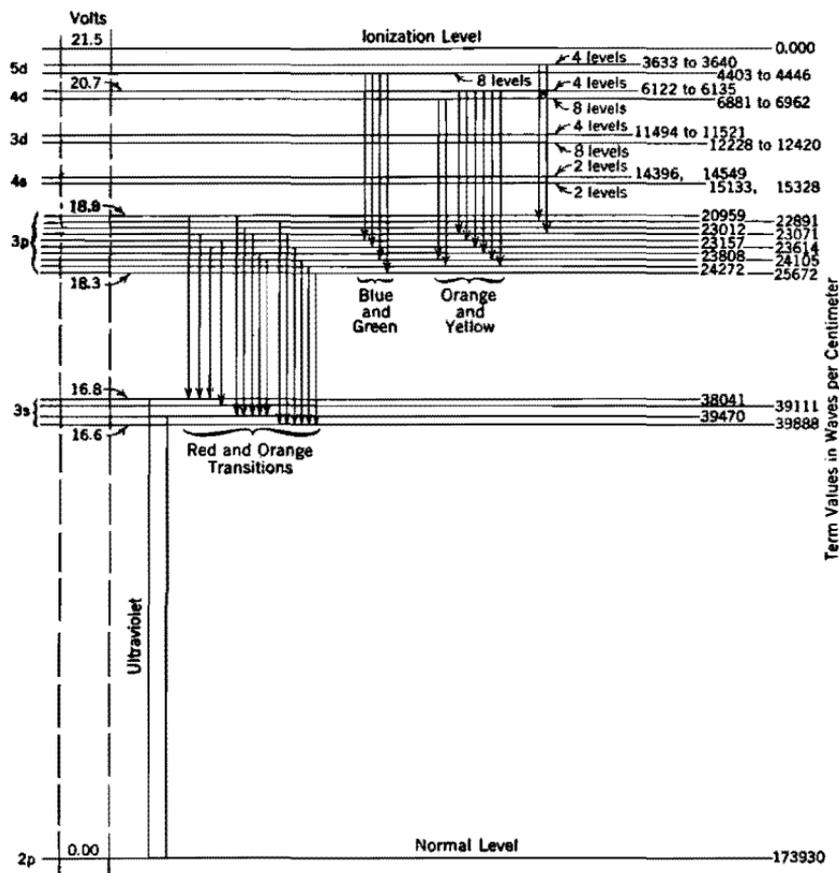
Fig. 141.   Energy-level diagram for a neon atom.   Only transitions giving rise to visible light are shown, except for two terminating at the normal level.   Not to scale.   (See Table XIII.)

A shift to the normal level from the first of these would require J to change by 2, hence is forbidden.   For the second, $J = 0$, as for the normal level, and transition between two $J = 0$ levels is forbidden.

Therefore transitions cannot occur from either of these levels to the normal. Many of the red transitions terminate on one or another of these two metastable states.

(f) The lowest excited state (39987.61, metastable) is 16.6 volts above the normal. Most of the levels from which visible light originates are a few volts higher than this. It is apparent that individual luminous cycles from normal state to normal state are very inefficient. As in mercury, cycles beginning and ending at the metastable states are efficient light-producers.

(g) Metastable neon atoms have such high energy content that they are able to ionize, by contact, atoms of many other gases which can be mixed with neon in a discharge tube. Examples of such gases are mercury, sodium, and argon.

(h) In a mixture that includes neon and some gas having low exciting and ionizing potentials, a low voltage gradient may introduce only enough electronic energy to excite and ionize the gas of lower ionizing energy. In that case the neon takes no part in the electrical discharge. At high voltage gradients in the same mixture the neon may become active. Luminous tubes that for the most part are blue, but have constricted portions that are red, make use of this behavior. The current must be the same at all points along the tube. Therefore the current density, and along with it the voltage gradient, must be greater in the constricted than in the other portions. This increased gradient results in excitation and ionization of neon, with resulting production of red color in the constricted parts.

200. Copper. The arc spectrum of copper is of interest partly because copper arcs occur frequently in engineering work, and partly because copper's quartet terms illustrate more complicated vector combinations of angular momenta than occur in either sodium or mercury. Table XIV contains some of the more important term values for CuI.[EE 175]

In the normal state of copper there is a lone fourth-shell electron outside of a completed third shell; that is, there is a $2 + 8 + 18 + 1$ electron array. A one-electron spectrum might be expected. Actually the arc spectrum of copper results from two rather distinct sets of energy states. One set is built on a $3d^{10}$ ion, corresponding to the one-electron expectation. Examples of this are $3d^{10}\, 4s\; {}^2S_{\frac{1}{2}}$ at 62308.00 (the normal level); $3d^{10}\, 4p\; {}^2P_{\frac{1}{2}}{}^\circ$ and ${}^2P_{1\frac{1}{2}}{}^\circ$ at 31772.698 and 31524.314; $3d^{10}\, 5s\; {}^2S_{\frac{1}{2}}$ at 19170.791.

The other set results from the activities of a $2 + 8 + 17 + 2$ array. It therefore consists of levels built on a $3d^9\, 4s$ ion. This set of levels produces what is essentially a three-electron spectrum, the three active

electrons being the two in the fourth shell and the ghost of the one missing from the third-shell d group (see item (b), Section 199).

The terms built on the $3d^9\,4s$ base provide an interesting illustration of the Russell-Saunders vector combination pattern for various angular momenta.[DD 108] For example, there are many energy states having the configuration $3d^9\,4s\,4p$. The three electrons that are active in this configuration have l-values of $d(=2)$, $s(=0)$, and $p(=1)$. The vector sum can be

$$L = 2 \oplus 0 \oplus 1 = 1 \qquad\qquad\qquad (730)$$

or

$$L = 2 \oplus 0 \oplus 1 = 2 \qquad\qquad\qquad (730.1)$$

or

$$L = 2 \oplus 0 \oplus 1 = 3 \qquad\qquad\qquad (730.2)$$

The spin-vector S is the half-integral vector sum of the three individual spin values, and has values as follows:

(1) If the first two of the three spin values add up to 1, S may be either $\frac{1}{2}$ or $1\frac{1}{2}$, thus

$$S = (\tfrac{1}{2} \oplus \tfrac{1}{2}) \oplus \tfrac{1}{2} = 1 \oplus \tfrac{1}{2} = \tfrac{1}{2} \ \text{ or } \ 1\tfrac{1}{2} \qquad (731)$$

(2) If the first two of the three add up to 0, S is $\frac{1}{2}$, thus:

$$S = (\tfrac{1}{2} \oplus \tfrac{1}{2}) \oplus \tfrac{1}{2} = 0 \oplus \tfrac{1}{2} = \tfrac{1}{2} \qquad\qquad (732)$$

This shows that S may have the value $\frac{1}{2}$ in either of two different ways; the two different ways give rise to different term values.

The L and S vectors add in all possible ways, vectorially, subject to the limitation that the resultant J must be a half-number, since the spin-vectors are both half-numbers.   The various combinations are as follows:

When $L = 1$ and $S = \frac{1}{2}$:   (733)

	Four $^2P°$ doublet terms:	
$J = 1 \oplus \tfrac{1}{2} = 1\tfrac{1}{2}$	if $S = 1 \oplus \tfrac{1}{2}$, at	if $S = 0 \oplus \tfrac{1}{2}$, at
or $J = 1 \oplus \tfrac{1}{2} = \ \tfrac{1}{2}$	16,428.659	5,964.26
	16,487.0	3,943.27

When $L = 1$ and $S = 1\frac{1}{2}$: (733.1)

Three $^4P°$ quartet terms:

$J = 1 \oplus 1\frac{1}{2} = 2\frac{1}{2}$ 23,289.348

or $J = 1 \oplus 1\frac{1}{2} = 1\frac{1}{2}$ 22,194.01

or $J = 1 \oplus 1\frac{1}{2} = \frac{1}{2}$ 21,364.27

When $L = 2$ and $S = \frac{1}{2}$ (734)

Four $^2D°$ doublet terms:

$J = 2 \oplus \frac{1}{2} = 2\frac{1}{2}$ if $S = 1 \oplus \frac{1}{2}$, at   if $S = 0 \oplus \frac{1}{2}$, at

or $J = 2 \oplus \frac{1}{2} = 1\frac{1}{2}$ 16,135.158   5,656.52

                        15,709.66   3,617.14

When $L = 2$ and $S = 1\frac{1}{2}$ (734.1)

Four $^4D°$ quartet terms:

$J = 2 \oplus 1\frac{1}{2} = 3\frac{1}{2}$ 18,794.05

or $J = 2 \oplus 1\frac{1}{2} = 2\frac{1}{2}$ 17,901.732

or $J = 2 \oplus 1\frac{1}{2} = 1\frac{1}{2}$ 17,763.847

or $J = 2 \oplus 1\frac{1}{2} = \frac{1}{2}$ 17,392.39

and so on. There are altogether, according to this combination system, the following terms for the $3d^9\ 4s\ 4p$ configuration:

Two pairs of P doublet terms; $\begin{matrix} 16,487 \\ 16,429 \end{matrix}$ and $\begin{matrix} 5,964 \\ 3,943 \end{matrix}$

Two pairs of D doublet terms; $\begin{matrix} 16,135 \\ 15,710 \end{matrix}$ $\begin{matrix} 5,656 \\ 3,617 \end{matrix}$

Two pairs of F doublet terms; $\begin{matrix} 18,582 \\ 17,345 \end{matrix}$ $\begin{matrix} 6,278 \\ 4,189 \end{matrix}$

Three P quartet terms, 23,289, 22,194, 21,364.

Four D quartet terms, 18,794, 17,901, 17,764, 17,392.

Four F quartet terms, 21,398, 21,155, 20,745, 20,005.

The ionizing potential of copper, 7.68 volts, corresponds to the highest level of the *one-electron* set of levels. The zero level is therefore 7.68 volts above the normal level. However, it requires more than 7.68 electron volts of energy to excite a copper atom to many of the levels in the three-electron set, so that many of copper's term values are negative. None of the negative ones are listed in Table XIV. If a copper atom receives between 7.68 and about 9 electron volts of energy it may be ionized, or it may be lifted to one of the negative levels without ionization.

PROBLEMS

CHAPTER XVI

1. The term values for three of the energy levels for an argon atom are as follows:

> 23,009
> 33,968 (a metastable level)
> 127,104 (the normal level)

(a) Calculate the ionizing potential of argon.

(b) What is the wave length, in angstrom units, of radiation resulting from a transition from the 23,009 level to the 33,968 level? In what part of the spectrum does this kind of radiation lie?

(c) How much energy, in electron volts, may a metastable argon atom in the 33,968 level release in case of a collision?

2. What must be the temperature, in degrees Centigrade, of sodium vapor in order just to permit a collision between a stationary atom and one whose energy is 15 times the average energy result in the production of yellow light due to transitions from the lowest excited state to the normal state? How large a percentage of the atoms have 15 times the average energy?

3. Suppose that an atom of neon in the higher of neon's two metastable states collides with a normal-state atom of argon, and that as a result of the collision the neon atom returns to the normal state and the argon atom is ionized. If the surplus energy all appears as kinetic energy of the electron ejected from the argon atom, what is its velocity in centimeters per second?

4. An electron possessing 25 electron volts of kinetic energy strikes and ionizes a normal-state mercury atom. If the velocity of the atom is not changed by the collision, with what velocity will the electron proceed after the collision?

5. In a region containing a mixture of neon and mercury, a neon ion collides with a mercury atom. As a result of the collision the mercury atom is ionized, and the neon atom returns to the normal state. How much kinetic energy is released in the process?

6. The wave length of the characteristic red line used as a chemical test for the presence of cadmium is 6,438 angstrom units. If a particular source emits this kind of red light to the extent of 6×10^{-8} watt, how many photons are being radiated per microsecond?

7. A mercury atom in the lower metastable state strikes a surface whose work function is 2.0 volts, and as a result an electron is ejected from the surface. If after striking the surface the mercury atom possess a kinetic energy of one electron volt, find

(a) The maximum velocity that the ejected electron can have.

(b) The velocity of the mercury atom after striking the surface.

8. (a) What are the wave lengths in angstrom units of the radiations produced by the two mercury transitions (Fig. 140) that terminate at the normal level?

(b) State the wave length, in angstrom units, of a prominent violet line in the arc spectrum of mercury.

(c) Explain by reference to Table XII how the selection rules indicate that this transition, giving rise to violet light, is a permitted transition.

(d) How much energy, in electron volts, must be given to a mercury atom in the normal state to put it into condition to radiate this kind of light?

9. (a) Using color boundaries as in Fig. 137, and neon term values as stated in Table XIII and Fig. 141, select a pair of neon levels between which a transition pro-

ducing red light is permitted, the lower level of the pair being a metastable one. Identify the levels you have selected by stating their term values in waves per centimeter.

(b) State the wave length, in angstrom units, of the transition between the selected levels.

(c) How much energy, in electron volts, must be given to a neon atom in the normal state to put it into condition to make this transition?

10. The term values of the important energy levels for copper are given in Table XIV.

(a) Draw to scale an energy-level diagram for copper (use single lines to represent groups of close levels). Label clearly the zero level, the normal level, and the level that corresponds to the lowest excited state. Show permitted transitions in the usual way (see Figs. 136, 140, and 141); in order to determine which transitions are permitted, use the selection rules stated in Section 193.

(b) Is the lowest excited state metastable?

(c) Select a pair of levels between which a transition producing green light is permitted. State the wave length, in angstrom units, of the transition you have selected.

11. Mercury vapor at a temperature of $1160°$ K, concentration 4×10^{14} atoms per cc, with ion and electron concentrations both 5 per cent of the atom concentration.

(a) Find the random ion current density in amperes per square centimeter (see Section 127).

(b) If the electron temperature is $15,000°$ K, how many atoms per cubic centimeter have energy enough to ionize a mercury atom that is in the upper of the two metastable states? (See Chapter X and Table V.)

CHAPTER XVII

PHOTOELECTRIC EMISSION AND ELECTROMAGNETIC WAVES

201. Photoelectric Emission. As stated in Section 171, radiant energy is transferred to electrons or atoms in discrete energy units of magnitude hf ergs, where f is the radiation's frequency in cycles per second, and h the ultimate unit of action. The light-quantum (photon) energy in electron volts, symbol E_{Ph}, is defined by the relation

$$E_{Ph}e10^7 = hf \qquad (735 \text{ p})$$

Here e is as usual the electronic charge. If radiation of E_{Ph} electron volts per photon strikes a surface whose (net) work function E_W is less than E_{Ph}, some of the electrons within the metal will be ejected out of it. Each one that emerges does so because it acquires from a light quantum the kinetic energy necessary to carry it past the work-function barrier at the surface (Fig. 69). This behavior is called *photoelectric emission* of electrons. In laboratory and commercial photoelectric devices the electrons are emitted into regions that are either entirely free from gas or contain an inert gas at a low pressure.

The following important generalizations can be made regarding the photoelectric emission of electrons from metal surfaces:[X 160, FF, GG, HH]

(a) Emission cannot occur unless

$$E_{Ph} \geqq E_W \qquad (736)$$

This limitation exists because each electron's ejection results from the action of a single light quantum; there is no cumulative effect. Equation (736) gives rise to the concepts " threshold frequency " and " threshold wave length " for a surface. These quantities are the values of frequency and wave length that correspond to the condition $E_{Ph} = E_W$. Thus *at threshold frequency*

$$hf = E_W e10^7 \qquad (737 \text{ p})$$

Emission occurs only for values of E_{Ph} and of frequency at or above the threshold, therefore only for wave lengths at or below threshold values. Since E_{Ph} is between 1.75 and 3.0 volts for visible light (see Fig. 137), a photoelectric surface must have a work function within or below this range to be sensitive to visible light.

(b) When the light-quantum energy of the incident light is somewhat greater than the work function, i.e., greater than the threshold value,

electrons can escape at a substantial rate. A very few of the electrons within the metal (1) have properly directed normal maximum energy before receiving the light-quantum energy, and in addition (2) are so located as to lose only the gross-work-function energy during emergence. For these few electrons the kinetic (velocity) energy E_v after escape is just $E_{Ph} - E_W$. This is the maximum energy of emergence. Most of the escaping electrons have less than this energy, for they (1) have initially less than normal maximum perpendicular-to-the-surface energy, and (2) lose more than the gross-work-function energy in the process of escape. In general then

$$E_v \leqq E_{Ph} - E_W \tag{738}$$

Fig. 142 illustrates this relation. A more familiar form of it is

$$\tfrac{1}{2}mv^2 \leqq hf - E_We \tag{739 esu}$$

(c) The energy distribution (not velocity distribution) of the emitted electrons is approximately symmetrical about the median between zero and the maximum possible value, as indicated at the upper right of Fig. 142.
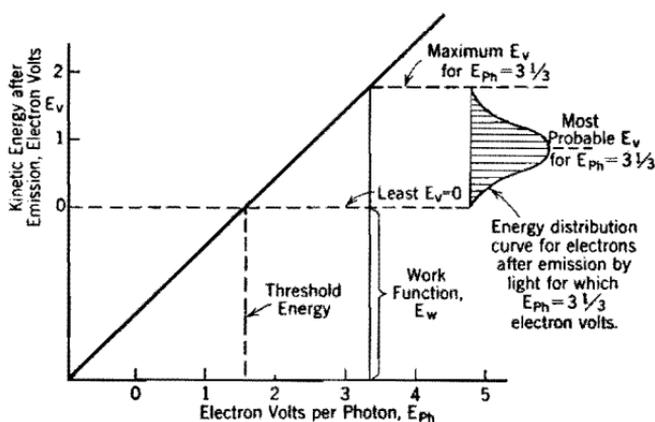
FIG. 142. Kinetic energies of photoelectrically emitted electrons.

(d) The intensity of the incoming light (ergs per square centimeter per second, or, watts per square centimeter) has no effect whatever on the maximum electron energy after escape or on the energy distribution among the escaping electrons; these quantities are controlled entirely by the light frequency, as described under (b) and (c).

(e) The number of electrons ejected per second is directly proportional to the intensity of the incident light, provided the wave length, angle of incidence, and polarization of the light remain unchanged.

(f) The response or sensitivity of a photoelectric tube as a whole, or of unit area of a sensitive surface, may be expressed variously by stating:

(1) The *quantum yield*, also called the *quantum efficiency*, as a fraction, or in per cent. When expressed in per cent, it is the number of electrons emitted per hundred incident photons. The quantum efficiency is usually not more than a few per cent.

(2) The *photoelectric yield*, in amperes per watt of light energy input; this is exactly the same thing as coulombs emitted per watt second of light energy. Photoelectric yield and quantum yield are related as follows:

$$\text{amperes per watt} = \frac{\text{quantum yield}}{EP_h} \qquad (740\ \text{p})$$

(3) The amperes (or microamperes) per lumen; the bases of reference is then the visible properties of the light rather than its intensity in watts per square centimeter.

(*g*) The magnitude of the response to light, whatever the manner of its description, is extremely sensitive to details of surface preparation. Composite surfaces consisting of a few successive thin layers, sometimes monomolecular, of selected substances, can be made to have very high photoelectric sensitivities.

(*h*) As the light frequency rises above the threshold value the sensitivity invariably increases, rather rapidly at first. In general, for clean metal surfaces, this increase is believed to continue indefinitely, but at a decreasing rate, in the manner suggested by the dotted curve in Fig. 148, page 394. However, in the corresponding curves for the specially prepared surfaces ordinarily employed in phototubes, pronounced humps appear just above the threshold, as in the solid-line curve of Fig. 148. The crest of such a hump may be expected to occur at a light-quantum energy that is between 25 per cent and 75 per cent above the threshold energy.[FF 28]

(*i*) No one has been able to measure any time lag between initial light incidence and electron ejection. If there is any time lag, it is less than a hundredth of a microsecond (10^{-8} second).

(*j*) Light can cause the transfer of electrons from one material directly into another against a retarding potential-energy barrier such as the contact difference of potential between metals. Light affects the various " rectifier " photocells in this way; see Section 223.

(*k*) Photoelectric emission or transfer of electrons must not be confused with the *photoconductive* behavior of substances, like selenium, whose electrical resistance decreases when they are illuminated.

(*l*) Photoelectric emission is not materially changed by variations in temperature, unless such variations are so extreme as to modify the structure of the emitting surface. The minor effects that temperature changes do have are quantitatively in accord with the properties of the degenerate electron gases within metals.

Many problems in the application and development of photoelectric devices are essentially optical in nature. For this reason it is necessary

to review some of the properties of light in order to establish a satisfactory foundation for study of the details of photoelectric behavior. The next several sections have to do with various aspects of electromagnetic propagation of light energy.

202. Propagation of Radiant Energy. Although light energy is delivered and received in E_{Ph} electron volt energy-units, *electromagnetic waves* transport the energy from point of origin to point of delivery.[A 292, B XXIII, G XI, U, JJ, II] The radiation of light from a point source takes place in just the same way that radio waves of alternating electric and magnetic fluxes travel radially outward in all directions from a vertical broadcasting antenna in free space. The important features of the electric flux pattern for radiation from a source (antenna) whose dimensions are small relative to the wave length are illustrated in Fig. 144a.

In light the central source of energy is essentially an electric dipole, of atomic dimensions, whose electric moment varies sinusoidally with time. The dimensions of such a source are small relative to the wave length, so that Fig. 144a is a true flux pattern for the radiation from it.

203. A Plane-Polarized Electromagnetic Wave; the Radiation Vector. At a sufficiently remote distance from a source like that of Fig. 144 the radiation approximates the *plane-polarized* wave of Fig. 143. In a plane-polarized wave the magnetic flux consists of parallel straight lines at right angles to the similarly parallel and straight electric flux lines. Both kinds of flux densities vary sinusoidally with time at a given location, and with position at a given instant.

In discussions of electromagnetic wave problems the terms "magnetic vector," "electric vector," and "radiation vector," the latter also known as "Poynting's vector," are commonly used. The *magnetic vector* for any point in the wave represents the direction and magnitude (varying sinusoidally) of the magnetic field strength H, also of the magnetic flux density B in the usual cases where the permeability is unity. The *electric vector* represents similarly the direction and magnitude of the electric field intensity F, also of the electric flux density D if the dielectric constant is unity. In the plane-polarized wave, Fig. 143, the electric and magnetic vectors are at every instant of equal magnitude. They are therefore *in phase with one another* as to space and time variations. They may be described mathematically by the following equations:

$$H = H_{\max} \cos 2\pi \left(\frac{x}{\lambda} - ft \right) \tag{741}$$

$$F = F_{\max} \cos 2\pi \left(\frac{x}{\lambda} - ft \right) \tag{742}$$

Here x measures distance in the direction of propagation; it is replaced

by r when Fig. 143 is thought of as a detail of a small portion of Fig. 144a. Also

f = the frequency of the source, hence of the variations at any point in the wave, in cycles per second.

λ = the wave length of the light in centimeters; of course $f\lambda = c$, the velocity of light.   (743)

H_{max}, F_{max} are the respective maximum values of H and F; since the maximum values are always numerically equal to one another, the instantaneous values H and F are also always equal to one another. The practical unit of H is the same as the emu unit.
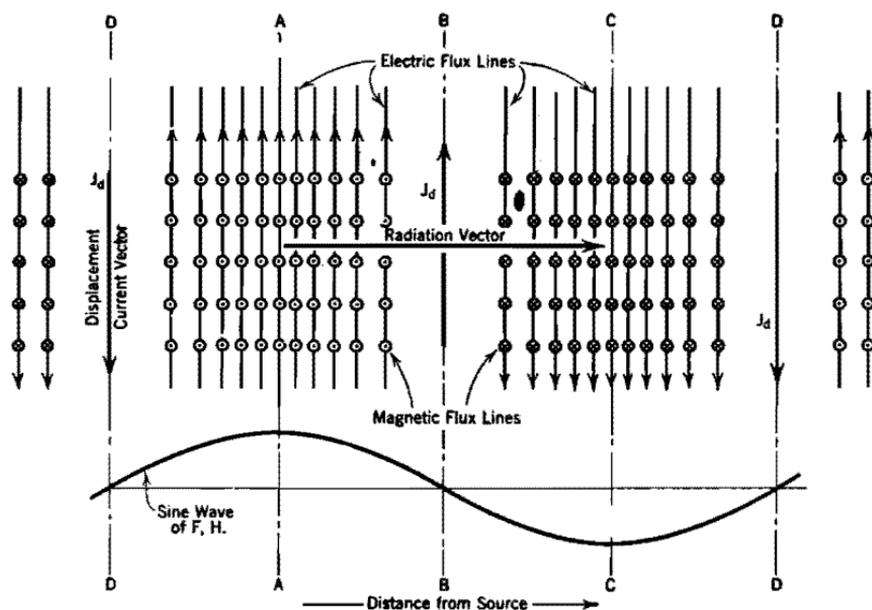
FIG. 143.   A plane-polarized electromagnetic wave; also, a detail of a small portion of Fig. 144a.

Poynting's radiation vector measures the instantaneous intensity of *radiant energy transmission* in ergs per square centimeter per second.   It is perpendicular to the other two vectors and proportional to their product according to the equation

$$P = \frac{c}{4\pi}HF \qquad \left(744 \begin{array}{l} F \text{ in esu} \\ H \text{ in emu} \end{array}\right)$$
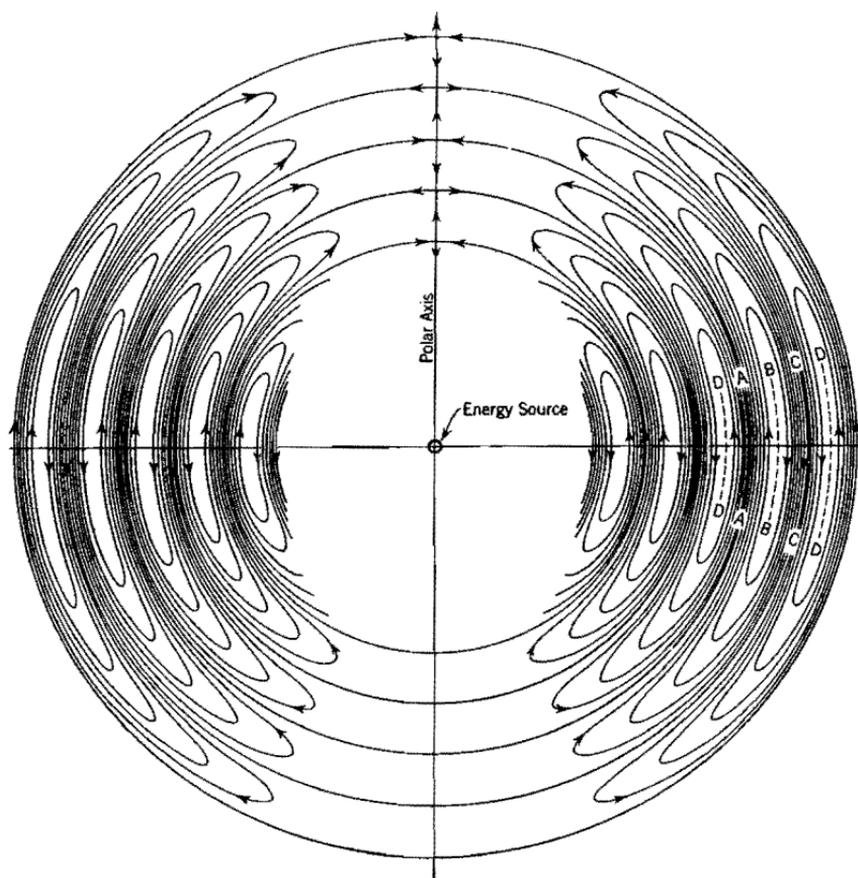
(a) Electric flux lines traveling outward from a radiating energy source in free space. Dimensions of the radiating energy source small relative to the wave-length.
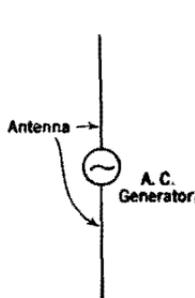
(b) Detail of the energy source.
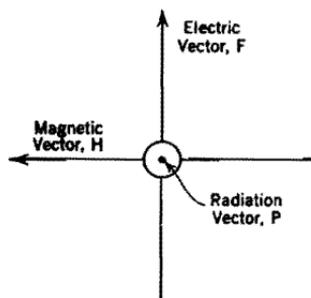
(c) Electric, magnetic, and radiation vectors in the plane AA.

FIG. 144.   Electromagnetic radiation from a point source.

In terms of cyclic variations in the plane-polarized wave

$$P = \frac{c}{4\pi} H_{max} \ F_{max} \ \cos^2 2\pi \left(\frac{x}{\lambda} - ft\right) \quad \left(745 \begin{array}{l} F \text{ in esu} \\ H \text{ in emu} \end{array}\right)$$

or

$$P = P_{max} \cos^2 2\pi \left(\frac{x}{\lambda} - ft\right) \tag{746}$$

The average value of P throughout a cycle is obtained from the fact that the cyclic average of the cosine squared is $\frac{1}{2}$, therefore

$$P_{av} = \frac{c}{8\pi} H_{max} \ F_{max} = \frac{c}{8\pi} F_{max}^{2} = \frac{c}{8\pi} H_{max}^{2} \quad \left(747 \begin{array}{l} F \text{ in esu} \\ H \text{ in emu} \end{array}\right)$$

In addition to being a measure of the rate of flow of energy, the instantaneous value of P determines the rate of absorption of radiant energy. For light passing through an absorbing medium, the probability that a light quantum of energy will be delivered in unit time within unit volume is proportional to P.[U 332] The proportionality constant is a property of the absorbing medium, and is of course zero in a vacuum.

In ordinary problems involving liquids and solids this probability is very large. Millions of light quanta are delivered within each cubic millimeter per microsecond, and the aggregate rate of energy input is measured in watts per unit volume.

204. The Mechanism of Propagation. Every portion of both kinds of flux lines illustrated in Figs. 144a and 143 should be thought of as traveling outward radially from the source with the velocity of light. The magnetic flux lines are rapidly enlarging circles centered on the axis of the source. The electric flux lines are rapidly elongating *closed loops*.

Appreciation of the significance of these loops is made easier by a little study of Fig. 145, which represents the electric flux pattern in the radiation from a vertical antenna projecting from the ground.* Fig. 145 is in fact Fig. 144a with the bottom half removed. The conducting ground surface is a plane of symmetry in which the flux pattern may be imaged.

In spite of the fact that the ground plane in Fig. 145 is a good conductor, it is not an equipotential, for each electric flux line emerges from it, loops around, and returns to it. There is (a) an alternating, outwardly moving charge density in the ground plane, and (b) a potential difference

---

* Actual broadcast radiation differs from that illustrated in Fig. 145, because the lengths of broadcast antennas are not small relative to the radiated wave lengths. The radiation field in a true broadcast wave is a systematic composite of fields radiated by elementary lengths of the antenna. Each elementary length radiates a field like that illustrated in Fig. 144a. Field strengths in the composite are obtained by integration over the entire antenna length.[F 424, G XI, 132]

between points in the ground plane, corresponding to the line integral $-\int F\,ds$ around the flux-line loops.   Thus there are traveling waves of charge and potential on the ground that resemble the traveling waves of charge and potential that are produced in transmission lines by lightning.[A 298, KK]   These traveling waves exist because it takes time for a voltage impulse to travel from one point to another even in a perfect conductor.   The velocity of travel is the velocity of light.
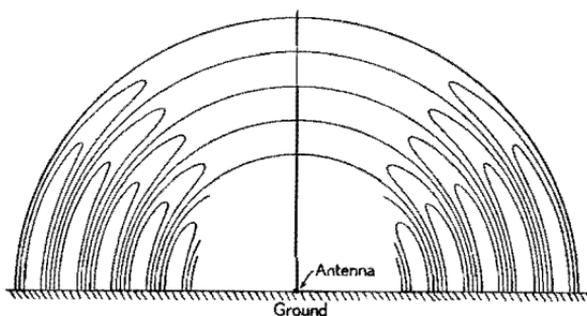
FIG. 145.   Electric flux lines traveling outward from a vertical radio antenna, the antenna being short relative to the wave length.

The interrelations between the electric and magnetic fluxes of Figs. 144a and 143 can be described by reference to familiar concepts.   The electric field intensity represented by the electric vector is the result of " cutting " by magnetic flux.   The magnetic flux is as usual associated with an electric current.   Since in free space there can be no conduction current, the magnetic flux must be associated with current of the displacement type, like that between the plates of an ordinary condenser passing an alternating current.   The displacement current density J_d is proportional to the rate of change of electric flux density,[A 290, B XVII] thus, if the dielectric constant is unity,

$$J_d = \frac{1}{4\pi}\frac{dF}{dt}\qquad\text{(748 esu)}$$

It is apparent from the detail picture, Fig. 143, that the surface AA is being rapidly " cut " by magnetic flux, because the magnetic flux is moving outward, toward the right.   In fact, since the magnetic flux density has its greatest value along AA, that plane is for the moment being cut at a more rapid rate than any other near-by plane.   The velocity of travel of the magnetic flux is everywhere the same; therefore the rate of cutting and resultant electric field strength are greatest wherever the magnetic flux is most dense.   This explains why the electric

and magnetic vectors are always in phase with one another in free space.

Along DD and BB the electric flux density is zero but rapidly changing, because of the outward flux movement. The electric flux density at BB is increasing positively, and at DD increasing negatively. Hence there is at both planes an electric current density of the displacement type; see Equation (748). Along DD the displacement current density is a maximum and is directed downward. It is zero at AA and has a maximum value in the upward direction along BB. Thus an electric current flows downward along DD, turns around near the lower branch of the polar axis (Fig. 144a), and flows upward along BB. It completes the electric circuit by another turn downward near the upper branch of the polar axis, so returning along the original DD path. The current in this closed circuit produces a magnetic flux, which is of course the flux of the advancing electromagnetic wave, and has a maximum value along AA.

Thus the advancing magnetic flux produces an electric field whose time rate of change constitutes a current that corresponds exactly to the advancing magnetic flux. This explains how the wave continually reproduces itself by advancing. *Electromagnetic wave equations* describe this behavior in the form of differential equations which can be solved for various geometrical patterns. Fig. 144a is a graphical representation of the solution for the region around an oscillating electric dipole.

205. Details of the Field around a Radiating Energy Source. The solution of the electromagnetic wave equations for the region around an oscillating electric dipole is expressed mathematically by the following equations, which were used in preparing Fig. 144a:[U 290, JJ VIII, G 368]

$$H_\phi = \frac{M\omega^2}{c^2} \frac{\sin\theta}{r} \cos 2\pi\left(\frac{r}{\lambda} - ft\right) \qquad \left(749 \begin{array}{l} H \text{ in emu} \\ M \text{ in esu} \end{array}\right)$$

$$H_\theta = 0 \tag{750}$$

$$H_r = 0 \tag{751}$$

$$F_\phi = 0 \tag{752}$$

$$F_\theta = -\frac{M\omega^2}{c^2} \frac{\sin\theta}{r} \left\{ \left[1 - \left(\frac{\lambda}{2\pi r}\right)^2\right] \cos 2\pi\left(\frac{r}{\lambda} - ft\right) - \frac{\lambda}{2\pi r} \sin 2\pi\left(\frac{r}{\lambda} - ft\right) \right\} \tag{753 esu}$$

$$F_r = +\frac{2M\omega^2}{c^2} \frac{\cos\theta}{r} \left\{ \frac{\lambda}{2\pi r} \sin 2\pi\left(\frac{r}{\lambda} - ft\right) + \left(\frac{\lambda}{2\pi r}\right)^2 \cos 2\pi\left(\frac{r}{\lambda} - ft\right) \right\} \tag{754 esu}$$

In these equations M is the maximum electric moment of the oscillating dipole. The subscripts r, θ, and ϕ identify the polar coordinate

components of H and F according to the polar coordinate system of Fig. 9. Of course $\omega = 2\pi f$.

At distances remote enough from the oscillating source particle to make the radius very large relative to the wave length, the quantity $\lambda/2\pi r$ is very small relative to unity, so that factors containing it can be dropped out of the equations. Such treatment converts them into the following expressions, which are very similar to Equations (741) and (742):

$$H_\phi = H_{\max} \cos 2\pi \left(\frac{r}{\lambda} - ft\right) \tag{755}$$

$$F_\theta = F_{\max} \cos 2\pi \left(\frac{r}{\lambda} - ft\right) \tag{756}$$

The quantities $H_{\max}$ and $F_{\max}$ have the following significance:

$$H_{\max} = \frac{M\omega^2}{c^2} \frac{\sin \theta}{r} \qquad \cdot \left(757 \begin{array}{l} H \text{ in emu} \\ M \text{ in esu} \end{array}\right)$$

$$F_{\max} = -\frac{M\omega^2}{c^2} \frac{\sin \theta}{r} \qquad (758 \text{ esu})$$

The factor $\cos 2\pi \left(\frac{r}{\lambda} - ft\right)$ in Equations (755) and (756) expresses the space-time periodicity of both fluxes. At a given radius r/λ is a constant, and H_ϕ, F_θ vary sinusoidally with time. At a given moment ft is a constant, and there is sinusoidal variation of field strengths with distance from the source. As always in free space, the two fluxes are in phase with one another. Equations (757) and (758) show that at any given radius both H and F decrease from a greatest value at the equator of Fig. 144a to zero at its poles.

It is often stated, and is for all practical purposes true, that the electric and magnetic vectors in the radiation from a point source are everywhere equal to one another and perpendicular to the radius. Equations (757) and (758) provide the mathematical justification for this statement.

At points for which the radius is *small* relative to the wave length, r/λ is negligible relative to ft, and only the terms containing $(\lambda/2\pi r)^2$ are of consequence in Equations (749–754). The flux pattern then described by these equations is that in the near neighborhood of an electric dipole of sinusoidally varying moment. The traveling aspect of the fluxes is submerged in their stationary but alternating aspect.

At points that lie between the extremes of remoteness from and nearness to the source particle, all three types of terms, those containing $(\lambda/2\pi r)^2$, $(\lambda/2\pi r)^1$, and $(\lambda/2\pi r)^0$ contribute to the determination of the field pattern.

206. Polarized Light.[II] Fig. 143 illustrates flux patterns for radiation that is *monochromatic*, in that it has a single frequency. It is also *plane-polarized*, the *plane of polarization* being that including the magnetic vector and the radiation vector. The plane of polarization is therefore perpendicular to the electric vector.

Circularly polarized light consists of two synchronized plane-polarized waves for which the electric vectors are (a) mutually perpendicular, as are the axes of the source particles also, (b) 90° out of phase with one another,[r] (c) of equal amplitude. The resultant electric vector is always constant and equal to $F_{\max}$, but continually rotating, as illustrated by the sequences in Fig. 146. Since F is always constant, the radiation vector P is also always constant and of value

$$P = \frac{c}{4\pi} F^2_{\max}$$

Since P is constant, this equation expresses its average value. Comparison with Equation (747) shows that this average P, for the resultant, is just the sum of the average values of P for the two contributing waves.

$2\pi\left(\frac{x}{\lambda}-ft\right)$	F_1 $F_M \sin 2\pi\left(\frac{x}{\lambda}-ft\right)$	F_2 $F_M \cos 2\pi\left(\frac{x}{\lambda}-ft\right)$	$F_1 \oplus F_2$
0°	•	→	→
30°	↑	→	↗
45°	↑	→	↗
60°	↑	→	↗
90°	↑	•	↑
120°	↑	←	↖
135°	↑	←	↖
150°	↑	←	↖
180°	•	←	←

FIG. 146. Rotation of the electric vector in a circularly polarized electromagnetic wave.

Elliptically polarized light consists of two synchronous wave trains that are 90° out of phase as to time, and at right angles, but unequal in magnitude. The cyclic average value of P turns out again to be the sum of the averages of the components.

207. Composition of Light from Various Source Particles.[11] A real source of light consists of an immense number of individual radiating particles (molecules or atoms). The time-phase relations and the polar axes of the various particles are continually changing in a completely random manner. A point receiving light from such a source experiences an electric field strength that is the resultant at each point and instant of the electric vectors due to all the individual particles. This resultant has always a particular momentary direction and magnitude, so that all light has *at every instant a definite plane of polarization*. Light is not said to be plane-polarized unless the plane of polarization remains constant.

Thus monochromatic light that is not polarized consists of successive wave trains having a common frequency but with randomly directed axes of polarization. Over a period of time an infinite variety of planes of polarization are observed. However, no one plane has a lasting predominance.

Of course the magnitudes as well as the directions of the electric and magnetic vectors at any point in the field of a composite nonpolarized source are continually varying between widely separated extreme values. The average value of both vectors is zero, because oppositely directed fields occur with equal frequency and strength.

The radiation intensity from such a source of course also varies widely. The radiation intensity is always positive, because it is proportional to the square of either contributing vector. It can be proved mathematically that, because of the randomness of the variations in time-phase relations, the time average of the intensity at any point in the field is the sum of the individual time averages of the radiation intensities due to the contributing source particles.

208. Interference.[11] Interference between two light waves occurs when their electric and magnetic vectors are equal and 180° out of phase with one another, so that $H = F = 0$. In order to be observed, this condition must persist for an appreciable period of time in spite of the random changes in the phase positions and polarizations of the source particles. Interference is experimentally realizable by splitting the light from a single source into two parts and making one part travel a half-wave-length-longer path than the other before they are rejoined. Whatever random variations occur in one part are exactly followed in the other because of their common origin. The 180° phase difference that arises from the difference in distance traveled makes the electric vectors of the two parts combine to give zero field strength after rejoining. The same is true of the magnetic vectors.

209. Reflection;   Standing Waves and Nodal Layers. Investigators of photoelectric phenomena have attached particular importance to the magnitudes and directions of the electric vectors at photoelectrically active surfaces.   For this reason it is worth while to review briefly the facts regarding the behavior of the electric vector in a wave
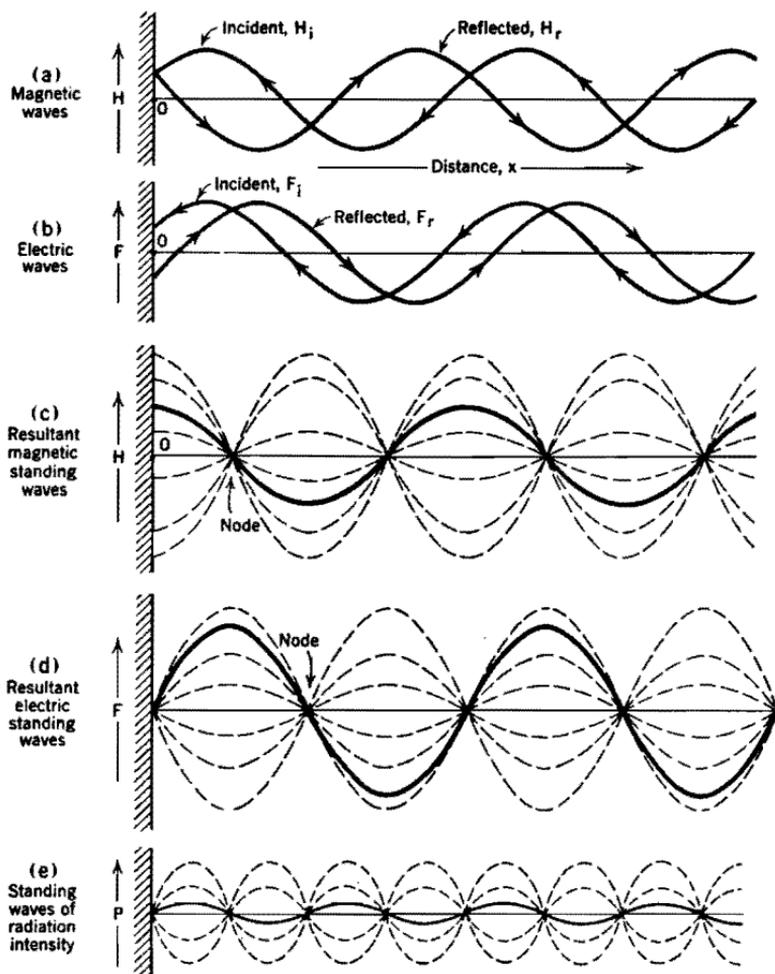
FIG. 147.   Standing electromagnetic waves resulting from 100 per cent reflection.

incident on such a surface.   It is useful to bear in mind during this discussion that distances between atoms or molecules in solids and liquids are usually only a few angstrom units, while the wave lengths of ordinary light are measured in thousands of angstrom units.   It is not until the X-ray region of the spectrum is approached that wave lengths of radiation waves are comparable with atomic spacings.

When an electromagnetic wave strikes a material surface part of the radiation is reflected. The remainder passes into or through the material substance behind the surface. If the angle of incidence is less than 90°, the electric and magnetic vectors of the incident wave each have components parallel to and perpendicular to the surface. The discussion in the following paragraphs applies only to the components parallel to the surface; of course these are the only components in the case of perpendicular incidence.

In the process of reflection the direction of the magnetic vector is unchanged, but that of the electric vector is reversed.[II, G XII]   The superposition of the reflected beam on the incident beam produces in the region outside of the reflecting surface a set of *standing waves* of electric and magnetic flux densities. These standing waves resemble the vibrations of a violin string.

The variations in the magnetic vectors of the incident and reflected beams (100 per cent reflection, normal incidence) for a particular instant are illustrated in Fig. 147a. The magnetic vector H_i for the incident wave is expressed by the equation

$$H_i = H_m \cos 2\pi \left(\frac{x}{\lambda} - ft \right) \tag{759}$$

Here H_m is the maximum value attained, x any distance from the surface, f the frequency, and λ the wave length. The term ft determines the phase position, at this particular instant, of the wave relative to the surface. The magnetic vector H_r of the reflected wave is

$$H_r = +H_m \cos 2\pi \left(\frac{x}{\lambda} + ft \right) \tag{760}$$

The positive sign in front of H_m emphasizes the fact that the magnetic vector is in the same direction in the reflected and incident beams; only the direction of travel is reversed.

Fig. 147b illustrates the similar electric vector relationship. The electric vector, however, reverses in the reflection process, so that the sign reverses in the F_r equation, thus

$$F_i = F_m \cos 2\pi \left(\frac{x}{\lambda} - ft \right) \tag{761}$$

$$F_r = -F_m \cos 2\pi \left(\frac{x}{\lambda} + ft \right) \tag{762}$$

The resultant instantaneous electric vector F is the sum of F_i and F_r. Mathematically:

$$F = F_m \left[\cos 2\pi \left(\frac{x}{\lambda} - ft \right) - \cos 2\pi \left(\frac{x}{\lambda} + ft \right) \right] \tag{763}$$

By the use of trigonometric relations this expression can be converted into

$$F = 2F_m \sin 2\pi f t \sin 2\pi \frac{x}{\lambda} \qquad (764)$$

The solid line in Fig. 147d describes the point-to-point variations in F for the particular instant chosen for Fig. 147b.

The factor $\sin 2\pi x/\lambda$ in Equation (764) describes a sinusoidal variation with x *at a phase position that is independent of time.* Only the *amplitude* of the sine wave pattern varies with time, as demanded by the factor $\sin 2\pi f t$. The dotted lines in Fig. 147d indicate the space variations in F for various moments of time. Evidently the electric field is made up of waves that rise and fall as time passes, but " stand still " in that they occupy fixed locations in space. They are therefore called *standing waves* in contrast with the *traveling waves* of Figs. 144a and 143. The zero points of a standing wave are called *nodes.*[G][XII]

The magnetic field outside the reflecting surface consists of a set of standing waves with nodes midway between those for the electric field, as illustrated by the contrast between Figs. 147c and 147d. The variations in the radiation vector are represented in Fig. 147e as comprising a standing wave whose frequency is twice, wave length half, that of the contributing vectors. The nodes in all of these waves are caused by local interference between the incident and reflected beams.

Fig. 147 illustrates the standing waves that are produced when the incident wave is entirely reflected. When it is only partially reflected the amplitudes of the incident and reflected vectors are not equal, and the electric and radiation vectors have at least small values at the surface. The smaller the per cent reflection, the greater the radiation intensity at and near the surface.

Actual surfaces are not optically perfect. In composite photoelectric surfaces, reflection may take place throughout a layer rather than at a surface. The outer parts of the layer may be subject to standing waves caused by reflections from deeper layers. In thin layers of a photoelectrically active substance such as sodium, placed over a silver base, most of the reflection is likely to take place at the sodium-silver boundary. There is then a standing wave in the sodium. If the thickness of the sodium layer is substantially less than one-fourth of the incident wave length (such layers are easily produced) the radiation intensity within the layer may be small even for intense illumination, because it is occupied by the node of a standing wave of radiation intensity.[87]

It has been stated earlier that the rate of absorption of light energy is proportional to the radiation intensity P. Photoelectric emission is

one manifestation of the absorption of light energy.   It is therefore reasonable to expect that no photoelectric emission can take place from a layer of material in which the presence of a node, like that at the surface in Fig. 147e, makes $P = 0$.

This same idea is inherent in the statement that the probability of acceptance of a light quantum of energy by an electron is proportional to the value of P which the electron experiences.   Therefore no electron within a nodal layer can accept radiant energy, for in such a layer $P = 0$.

Experiments have shown that photoelectric emission is in fact strongly affected by the details of the radiation-intensity pattern within surface layers whose thickness is a small fraction of the incident wave length. In general the reasonable supposition that most photoelectrons originate in layers of metal that are close to the surface is confirmed by these experiments.

Care must be taken not to generalize too freely from the preceding discussion.   The production of nodes and standing waves is only one of many aspects of surface behavior that can markedly affect the emission from composite surfaces.

210. Light Penetration and Absorption.[II, U 256, G XII]   Suppose a clean surface in a vacuum is exposed to light of frequency f, energy per light quantum E_{Ph}.   Part of the light is reflected away from the surface, the remainder continuing inward at an angle depending on the angle of incidence and the index of refraction of the material behind the surface. The fraction of the light energy that is reflected depends on the properties of the surface layers, and not on the rate at which absorption of light energy takes place after the beam has entered.

The unreflected portion passes into the interior, decreasing in intensity due to progressive loss of energy.   An *absorption coefficient* describes the relation between radiation intensity and the rate of absorption of energy.   If the absorption coefficient is large the unreflected portion of the wave penetrates only a very short distance.   Therefore to be transparent a material must (a) reflect only a small part of the light that strikes it, and (b) absorb light energy slowly.   The energy that is absorbed from the radiation is received within the interior volume, though sometimes chiefly in layers near the surface.   It is accepted in packets having an energy of E_{Ph} electron volts each.   The energy is delivered to the electrons or atoms of the material, which are thereby agitated, so that there arises a tendency toward an increase in temperature.

211. Mechanism of the Photoelectric Ejection of Electrons.   The mechanism of electron ejection is not sufficiently well understood to permit a complete description of an electron's adventures on the way out of a metal, and there is no point in repeating the speculations and

hypotheses that have appeared in various books and technical papers. It is, however, helpful to review some of the facts with which the prevailing incomplete and somewhat contradictory theories have to deal.[FF, HH]

Presumably the electrons that emerge acquire energy from the light in regions that are (a) inside the work-function boundary, yet (b) near enough to the surface so that the average value of the radiation intensity has an appreciable value, also (c) not so deep in the metal that collisions with atoms on the way out absorb all the excess energy before the surface is reached.

Throughout the metal, therefore including the regions from which photoelectrons start, the velocity distribution of the electrons is that of a degenerate gas, as illustrated in Figs. 91 and 92, pages 225 and 228. The corresponding type of energy distribution is illustrated in Fig. 71, page 175. Since photoelectric devices ordinarily operate at or near room temperature, the completely degenerate distribution, for $T = 0$, Fig. 71a, is a very close approximation to the true one.

Even the highest-energy (normal maximum level) electrons lack E_W electron volts of having energy enough to escape beyond the metal boundary. Since an electron can receive only E_{Ph} electron volts of energy from the entering light, it is evident that, as stated in Section 201, no electrons can be ejected by light for which $E_{Ph} < E_W$.

The total kinetic energy of an electron in the metal is the sum (numerical, not vector) of the three component parts that correspond to the velocities in the three coordinate directions. Assuming the x-coordinate axis chosen perpendicular to the surface through which escape is to take place, only the energy E_u, corresponding to the x-directed velocity u, is effective in taking an electron across the boundary. The total velocity c (this must not be confused with the velocity of light) corresponds to total energy E_c. Numerically:

$$u = 5.93 \times 10^7 \sqrt{E_u} \qquad (765 \text{ p})$$

$$c = 5.93 \times 10^7 \sqrt{E_c} \qquad (766 \text{ p})$$

At threshold frequency the only electrons emitted are those which:

(a) Are in the normal maximum level before accepting a light quantum of energy, also,

(b) Lose no energy by collisions with ions or electrons of the metal before reaching the boundary, also,

(c) Arrive at the work-function boundary traveling at exactly right angles to it, for only the normal-to-the-surface energy contributes to escape.

As the energy per light quantum or photon of the incoming light

arises above the threshold value (E_{Ph} becomes greater than E_W) the photoelectric yield (electrons per light quantum) increases, because:

(a) It is possible for electrons initially below the normal maximum level to be raised to or above the emission level, also,

(b) Affected normal maximum level electrons can lose some energy between point of energy acceptance and the barrier, yet arrive with a value of E_u greater than the gross work function, also,

(c) Electrons, initially in or near the normal maximum level, that have lost little or no energy prior to reaching the barrier, can escape even though they reach the surface with a velocity not exactly at right angles to it.

Since the temperature is never quite zero, there are always some electrons above the normal maximum level. For this reason there is not a perfectly sharp cut-off at threshold frequency, and the cut-off becomes less sharp as the temperature of the emitting surface is increased.

212. Color Sensitivity; Selective Photoelectric Emission.[FF,HH] For the reasons just stated, the quantum yield from a given surface should be expected to increase as the energy per photon of light increases above the threshold value. Experiments have shown that the photoelectric current from actual surfaces increases in general accord with this expectation. The dotted curve in Fig. 148 is typical of quantum yield curves for clean metal surfaces.

However, a quantum yield curve for a typical specially prepared emitting surface of the type used in industrial phototubes will ordinarily have a maximum point at a value of E_{Ph} that is between 25 and 75 per cent higher than the threshold value, as illustrated by the solid-line curve in Fig. 148. Beyond the maximum the curve decreases sharply, but not to zero, then rises again in a manner similar to that observed for clean metal surfaces. Surfaces whose quantum yield curves have one or more of these crests are said to exhibit a "selective photoelectric effect."[FF 28, HH V]

Color sensitivity or quantum yield curves for a complete phototube may be very different from those for the photoelectric surface within the tube, because the glass walls through which the light enters are more transparent to some wave lengths than to others. Glass is opaque to ultraviolet light, so that the color sensitivity curves of glass envelope tubes decline as E_{Ph} rises above the visible range, instead of rising continually, as do the curves for the surfaces themselves, as shown in Fig. 148.

Many different manufacturing formulas are used for preparing special surfaces for commercial phototubes. Devices employing the " Cs-O-Ag "

surface, prepared by causing deposition of a one-atom-thick layer of caesium on a mixed caesium-oxide and silver-oxide layer, over a silver base, have been rather widely used.[FF 41]   The photoelectric yield curve for a typical Cs-O-Ag tube has two maxima, one between 3000 A° and 4000 A°, the other between 7000 A° and 8000 A°, with satisfactory emission at all intermediate wave lengths.

In selective photoelectric emission, the location of the threshold and the location and prominence of the crest, or crests, are markedly affected by the details of surface preparation.   It is in fact difficult to keep manufacturing processes sufficiently uniform to produce tubes with
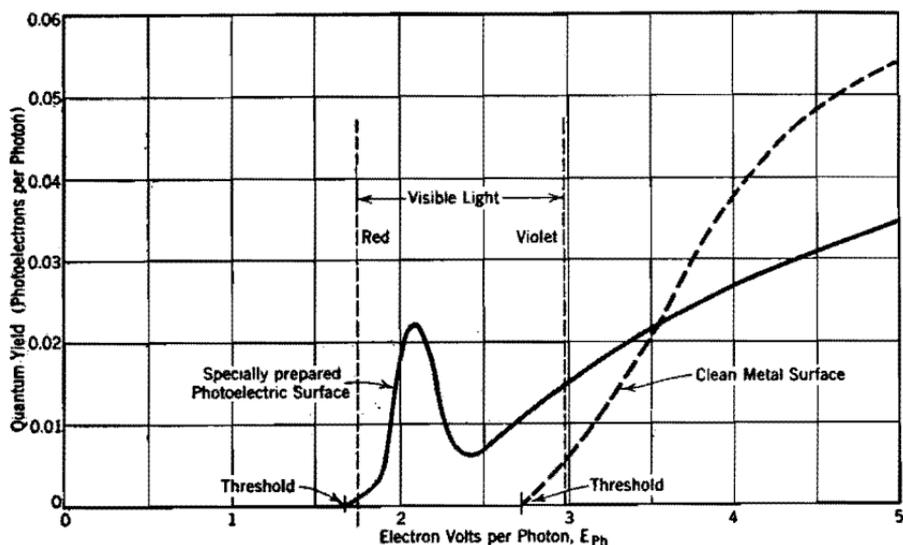
FIG. 148.   Typical color-sensitivity curves for photoelectric surfaces. The crest just above threshold energy in the solid-line curve illustrates the "selective photoelectric effect."

like characteristics.   The locations and prominence of the crests are also considerably affected by changes in light polarization, if the light strikes the surface at an angle that is less than 90° (see also the last three paragraphs in Section 213).

Theoretical developments based on the degenerate gas electron velocity distribution within a metal, Figs. 91 and 92, have led to very satisfactory explanations of the shapes of clean-metal quantum yield curves just above their thresholds, even in regard to the rather minor effects of temperature changes.[88]   However, no very convincing or complete theoretical explanation has yet been advanced for the shapes of the various experimentally obtained quantum yield curves for composite surfaces.   Complete explanations must take into account op-

tical effects, such as those due to variations of reflection and absorption coefficients of the emitting surface with wave length, the effects of nodes in standing waves (Fig. 147e) and of polarization. Some experimental evidence indicates that selective photoelectric emission is in some cases almost entirely a result of variations in optical constants, which markedly influence the pattern of nodes and anti-nodes in the standing waves produced by reflection, and so the emission.[87]

Considerable attention has been given to the possibility that the effects of potential hills and valleys may account for selective emission through complicated work-function barriers. Electrons can shuttle back and forth (in the $\pm x$ direction), with quantized periodicities, within the valley of a work-function barrier like that illustrated in Fig. 149. The valley may be so thin that the energy levels within it are
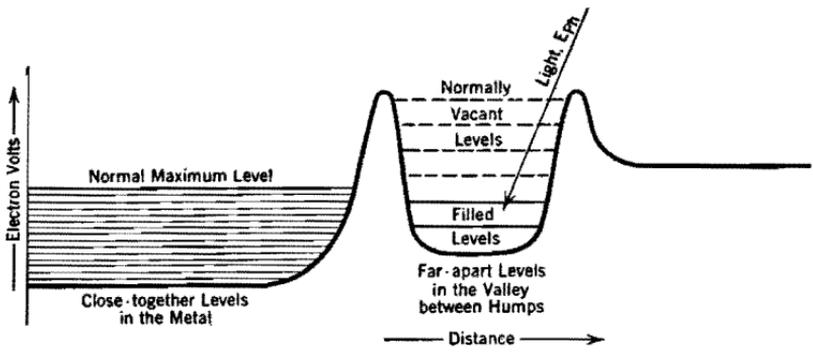
FIG. 149. Selective photoelectric effect. If E_{Ph} just equals spacing between valley levels, the electron concentration in the valley's normally vacant levels may conceivably be made large enough to permit appreciable emission through the outer hump.

spaced as much as a volt or two apart; see Section 89 for a statement of the dependence of level spacing on other quantities.

It is at least conceivable that the color sensitivity of photoelectric emission through such a barrier may be markedly affected by the wide spacing between the levels in the valley. For example, exact " resonance " (i.e., equality) between E_{Ph} of the incident light and the spacing between the valley levels may conceivably lead to a considerable electron concentration in high-up normally vacant valley levels. This may in turn cause appreciable electron passage through the outer hump, if thin enough. Such behavior would result in a peak in the quantum yield curve at the wave length corresponding to the E_{Ph} that equals the spacing between the valley's levels.[89, FF 32]

Section 106 contains a discussion of the passage of electron waves through humps. The presence of electrons in the valley in the first

place probably has to be explained by such passage through the inner hump.

213. Energies of Escaping Electrons. The electrons that are ejected from a surface photoelectrically all acquire E_{Ph} electron volts of energy from the incoming light. However, they start from various energy levels, lose various amounts of energy on the way out, and approach the surface internally at a variety of angles. Some barely have normally directed energy enough to escape; the velocities of such electrons after escape are extremely small. At the other extreme are those few that start from the normal maximum level, lose no energy
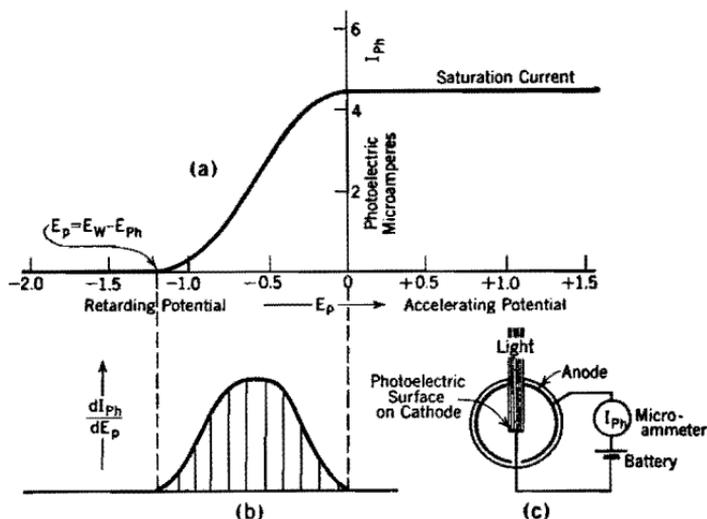
(a) Directly observed integrated distribution curve. Light is of constant intensity, energy per photon E_{Ph} electron volts.

(b) Distribution curve.

(c) Phototube and circuit.

Fig. 150. Energy distribution among photoelectrons, central cathode vacuum phototube. Potentials electrostatically measured.

prior to reaching the surface, and happen to arrive at the surface moving just at right angles to it. The kinetic energy $E_{v\,max}$ of each such electron after escape is just

$$E_{v\,max} = E_{Ph} - E_W \qquad (767)$$

The energies E_v of the other escaping electrons are distributed between zero and the maximum value in the manner illustrated at the upper right of Fig. 142.[HH 114]

The energy distribution curve is usually fairly symmetrical about the median between zero and the maximum. Its shape is found experimentally by measuring the photoelectric current that flows to an adja-

cent electrode in spite of a retarding potential. The curve like that in Fig. 150a thus obtained is in reality an integrated distribution curve for the energies of the escaping electrons. Fig. 150b is the distribution curve that corresponds to Fig. 150a. As explained later in Section 215, contact difference of potential must usually be taken into account in preparing curves like that in Fig. 150a from experimental data.

The magnitudes of the *directional* components of the electron's velocities after emission are sometimes of interest. Available evidence indicates that the directions of the velocities on emergence are symmetrically distributed about a normal to the surface, regardless of the angle of incidence of the light, or of its polarization.[HH 129]

The effects of polarization of light on photoelectric emission[HH] are usually discussed in terms of the electric vector's components parallel to and perpendicular to the surface. As long as light strikes a photoelectric surface at right angles ("normal incidence") the electric vector is parallel to the surface regardless of the plane of polarization (see Fig. 143) and polarization has no effect whatever on the photoelectric emission.

Light making an angle of less than 90° with the surface may be plane-polarized in any one of various directions. At one extreme condition the electric vector is parallel to the surface, so that there is no perpendicular-to-the-surface component. At the other extreme condition the electric vector lies in a plane perpendicular to the surface, and has the greatest perpendicular-to-the-surface component that is possible for any given angle of incidence. The greater the deviation from 90° incidence, the greater is the maximum possible perpendicular component of the electric vector.

Available evidence indicates that the introduction of a component of the electric vector perpendicular to the surface favors an increase of the x-directed velocity components (those normal to the surface) among the escaping electrons, at the expense of velocity components parallel to the surface.

PROBLEMS

Chapter XVII

1. A certain photoelectric surface has a work function of 2.5 volts.

(a) Will this surface emit photoelectrons when exposed to red light? (See Fig. 137.)

(b) When exposed to violet light?

(c) What is the color of light that has the threshold wave length for this surface?

(d) Ultraviolet light whose wave length is 1,850 angstrom units strikes this surface. What is the energy in electron volts of the fastest emitted photoelectrons?

2. A certain point source radiates 20 watts of monochromatic light. The radiation is distributed equally in all directions.

(a) Find the numerical value of Poynting's vector at a distance of 50 cm from the source.

(b) Find the value of the cyclic maximum value of the electric vector at that distance, in volts per centimeter.

(c) Find the value of the cyclic maximum value of the magnetic field strength at that distance, in flux lines per square centimeter.

(d) A photoelectric surface is located 50 cm from this light-source. Find the photoelectric current density if the quantum yield is 2 per cent, and the wave length 5,890 angstrom units (sodium yellow).

(e) What is the corresponding " photoelectric yield "?

3. Sodium yellow light, wave length 5,890 angstrom units, is incident on a surface consisting of a layer of sensitive photoelectric material 500 angstrom units thick over a base metal (often silver is used as a base metal). Assume that the cyclic maximum value of the electric vector in the incident wave is 2 volts per cm, and that a negligible absorption of light energy takes place in the thin layer.

(a) If at the boundary surface between thin layer and base metal 90 per cent of the light is reflected, find the value of Poynting's vector at that boundary; also at the outer face of the thin layer.

(b) Find Poynting's vector at these two locations if only 10 per cent of the light is reflected from the inner boundary.

4. An electron in the path of electromagnetic radiation experiences simultaneous moving electric and magnetic fields. Analyze the force action on such an electron, with particular reference to the discussions given in Sections 31 and 204. In calculating the voltage induced in an antenna wire by a radio wave, should the effect of the moving magnetic field alone, of the electric field alone, or of both, be taken into account?

5. The direction of propagation of a certain beam of plane-polarized light makes an angle of 30° with a surface that the beam strikes. The maximum strength of the electric vector in the incident beam is F_m.

(a) What is the maximum value the perpendicular-to-the-surface component of the electric vector can have? What angle does the corresponding plane of polarization make with the surface?

(b) What is the least value the perpendicular-to-the-surface component can have, and what angle does the corresponding plane of polarization make with the surface?

(c) If the plane of polarization makes an angle of 60° with the surface, what is the strength of the perpendicular-to-the-surface component of the electric vector?

6. A certain surface has a gross work function of 7 volts, and the kinetic energy of electrons in the normal maximum energy level (see Chapter VIII) is 4.5 electron volts. An electron, within the metal, whose kinetic energy is 3 electron volts has its energy increased to 9 electron volts by acceptance of a photon's energy.

(a) What was the wave length of the photon, in angstrom units?

(b) If after receiving the photon's energy the electron moves in general toward the surface, but arrives moving at an angle of 30° with the surface, will it escape?

(c) If on arrival at the surface its direction of motion is perpendicular to the surface, but it is barely able to escape, how much energy has it lost by impact after accepting the photon's energy?

(d) What is the least angle at which it can arrive at the surface, yet be able to escape, if it loses no energy by impact after accepting the photon's energy?

CHAPTER XVIII

PHOTOSENSITIVE DEVICES

214. Volt-Ampere Response of a Vacuum Phototube in Which the Electron Receiver Surrounds the Emitter.[FF, GG, HH] Fig. 150c in the preceding section illustrates a geometrical arrangement of emitting and receiving electrodes in which nearly all ejected electrons must travel toward the opposite electrode, for the reason that the electron-receiving anode almost completely surrounds the centrally located electron-emitting cathode. The anode surface is silvered to make it a good electrical conductor and light reflector. Such a device is sometimes called a *central cathode* phototube.

Fig. 150a illustrates the kind of volt-ampere curve to be expected with this geometry when the space between the electrodes is completely evacuated and monochromatic light used. As mentioned in Section 213 this curve is essentially an integrated energy distribution curve for the emergent electrons.

If the receiving electrode is negative by an amount greater than $E_{Ph} - E_W$ volts no electrons can reach it. The fastest ejected electrons can barely reach it if the potential difference is just $E_{Ph} - E_W$. If there is no potential difference between the electrodes, all emitted electrons reach the receiving surface; hence no increase in current results from making the receiving surface positive. A tube of this type is said to "saturate" at zero voltage.

215. Effects of Contact Difference of Potential in a Phototube. Contact difference of potential between electrode materials must be taken into account in interpreting measurements of potential differences between phototube electrodes, if such measurements are made by ordinary electromagnetic (D'Arsonval) meters. The order of magnitude of contact difference of potential is the same as that of the energies of the emitted photoelectrons, and the field produced by it just as real as that due to a battery, yet electromagnetic meters do not register its effects. For this reason the potential scale shown in Fig. 150a corresponds to electrostatic potential measurements, which do register the effects of contact difference of potential.

Suppose that in the circuit illustrated in Fig. 150c the work functions of the emitting and receiving surfaces are 1.8 and 4.0 volts respectively. The anode's receiving surface is then $4.0 - 1.8 = 2.2$ volts below the

emitter's potential when the battery voltage is zero.   If violet (3.0-volt) light strikes the emitting surface, the fastest ejected electrons have 3.0 − 1.8 = 1.2 electron volts of kinetic energy.   This is not enough to carry them to the opposite electrode against the 2.2-volt field established by the contact difference of potential, so that the current is zero when the battery voltage is zero.

Current first appears when an increasing battery potential reaches 1.0 volts.   An electrostatic voltmeter across the tube then registers a negative 1.2-volt potential (2.2 − 1.0 = 1.2) which the fastest ejected electrons can just overcome.   Saturation occurs when the battery potential is 2.2 volts, just enough to compensate for the 2.2-volt contact difference of potential between cathode and anode.   Thus cut-off and saturation occur at electrostatically measured potential differences of −1.2 and 0.0 volt respectively, as illustrated in Fig. 150a.

216. Volt-Ampere Response of a Vacuum-Type Tube in Which the Emitter Surrounds the Receiver.[FF, GG, HH]   Phototubes are usually arranged somewhat as shown in Fig. 151b, with a small electron-receiv-

(a) Volt-ampere curve at constant light intensity, energy per photon E_{Ph} electron volts.

(b) Phototube and circuit.
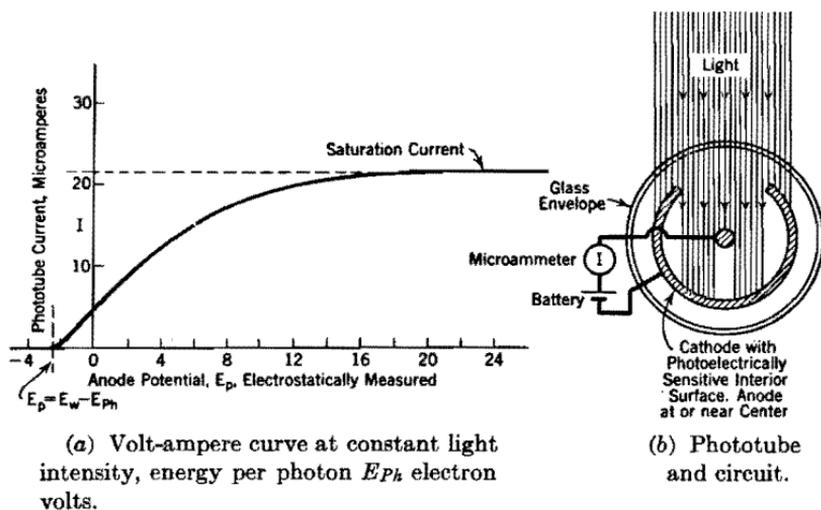
Fig. 151.   Central anode vacuum-type phototube with cylindrical cathode.

ing electrode located at a position that is more or less completely surrounded by a photoelectrically emitting electrode.   This *central anode* arrangement permits the effective use of a large photoelectrically active surface.

In a central anode vacuum-type tube saturation occurs as before, but at a much higher voltage, as illustrated in Fig. 151a.   At zero potential difference (electrostatically measured) only a very few electrons reach the receiver.   Although at this potential condition many

electrons are emitted, and there is no energy barrier to obstruct their approach to the central receiver, very few of them happen to be aimed directly at it. Most of them strike and reenter the emitter. With sufficiently sensitive meters it is possible to measure currents down to a cut-off at $E_W - E_{Ph}$ volts, as with a central cathode tube.

When the anode is made positive, in a central anode tube, there is a definite field drawing electrons toward the center. If this field is strong enough, all electrons reach the anode, producing saturation. At voltages above the zero-field value but below saturation many electrons are deflected toward the central anode, but acquire sufficient kinetic energy to carry them on around it and into the emitter again. In a similar way nonrecurring comets are deflected toward the sun, yet pass partially around it and on out of range of the solar system.

217. Use of Gas to Amplify Photoelectric Currents.[GG VII, FF III] The electron emission from photoelectric surfaces is small at best; the current from most vacuum phototubes, even those exposing considerable sensitive area to incoming light, must be measured in microamperes. By introducing an inert gas at a low pressure it is possible to amplify the photoelectric current without seriously modifying the response to light. At the same time the volt-ampere characteristic is changed in such a way as to increase the convenience of the device as a circuit element. The gas must be inert, that is, chemically inactive, in order to avoid deleterious chemical effects on the photoelectric surface within the tube. It must also be one that is not readily absorbed by the metallic parts or glass of the tube. Argon, neon, and helium have been used.

The amplification is a result of the production of an ionization current that is proportional to the photoelectric emission current. The amount of amplification (rarely more than ten to one) is dependent on the gas pressure in the tube, the requirements for excitation and ionization of the gas, the applied potential, and the interelectrode geometry.

Fig. 152a represents a phototube in which the cathode (left) and anode (right) are parallel planes. The light that strikes the cathode ejects electrons from it photoelectrically to the extent of J_{Ph} amperes per square centimeter. J_{Ph} is directly proportional to the light intensity, dependent on the wave length of the incoming light according to the color sensitivity curve for the particular surface used, and is entirely independent of the anode potential. For the sake of simplicity, edge effects, necessarily present if the spacing between the electrodes is appreciable relative to their length or width, are neglected in the discussion.

The type of volt-ampere curve to be expected with this geometry is

illustrated in Fig. 152*b*.   Whether the tube is evacuated or gas-filled, the current reaches the value I_{Ph}, corresponding to current density J_{Ph}, at about zero volts, just as in Fig. 150.   All emitted electrons then reach the anode.   If there is no gas in the tube, the current remains constant at this value for all positive anode voltages, as indicated by the dotted horizontal line in Fig. 152*b*.

If inert gas is present at a pressure of a few millimeters of mercury, the volt-ampere curve is just the same as without gas until the anode

<table>
<tr><td>(a) Geometrical arrange-
ment.</td><td>(b) Volt-ampere curves, monochromatic
light of constant intensity.</td></tr>
</table>
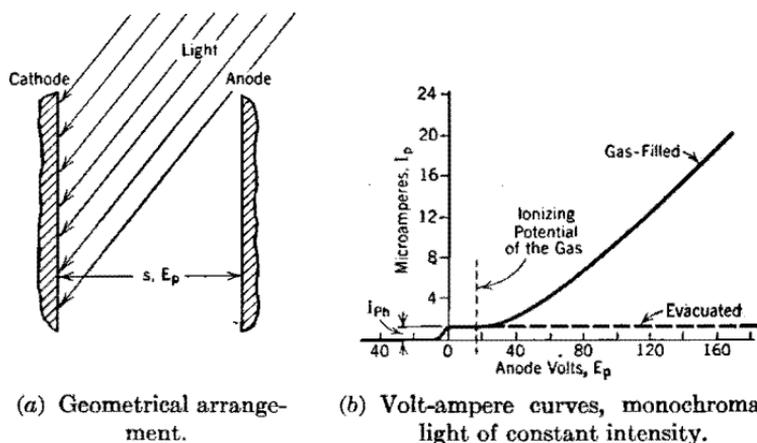
FIG. 152.   Effect of inert gas in a phototube with parallel plane electrodes.

voltage rises above the ionizing potential of the gas.   The increase in current above I_{Ph} that occurs after the ionizing potential is passed is due to an ion and electron current in the gas that is proportional to I_{Ph}. Because the additional ionization current is proportional to the photoelectric emission current, all ordinates of the volt-ampere curve increase in proportion to any increase in light intensity.   This direct proportionality exists only as long as the illumination is not excessive.

Thus the gas *amplifies* the photoelectric current, the amplification factor being I_p/I_{Ph}.   The amplification increases with rising anode voltage, as is indicated by the shape of the solid volt-ampere curve in Fig. 152*b*.

218. Mechanism of Gas Amplification; Elastic and Inelastic Collisions.   If in Fig. 152*a* the anode-to-cathode voltage is E_p, and the anode distance s, the electric field F in the intermediate region is

$$F = -E_p/s \qquad \text{(volts or statvolts per centimeter)} \qquad (768)$$

An electron photoelectrically ejected from the cathode experiences acceleration due to this field.   Acceleration continues unchecked until

collision with a gas particle takes place.   Many collisions occur before the anode is reached.

A collision may be *elastic*, like that between billiard balls; then the total kinetic energy and the x, y, z, momentum components are differently divided between the particles after collision, but have the same combined value as before.   Or a collision may be *inelastic;* then it results in excitation or ionization of the atom, with consequent absorption of energy, so that the total kinetic energy and combined momentum components are not the same after as before the collision.[W III]

Suppose the gas used is argon, ionizing potential 15.7 volts, lowest excitation potential 11.5 volts, atomic weight 39.9, mass ratio $m_g/m_e =$ $1824 \times 39.9 = 72,800$.   Photoelectric devices ordinarily operate at room temperature, that is, in the neighborhood of 300° K.   This is low enough so that for most purposes *the heat energies of the gas atoms may be neglected, that is, the atoms may be thought of as stationary.*

Consider a collision between one of the argon atoms and an electron that has acquired 10 electron volts of energy from the field.   Since this is not enough energy either to excite or ionize the argon atom, the collision must be an *elastic* one.   Then, although the electron may rebound with greatly changed *direction* of motion, it loses very little energy.   As Loeb points out, the result is like that of throwing a tennis ball at a boat; not very much energy is imparted to the boat, because its mass is so large relative to that of the ball.

A derivation, not given in this book, based on the conservation of energy and momentum, shows that on the average the fraction f of an electron's energy that is transferred to the atom in elastic collisions is[V 602, NN, 96]

$$f = \frac{8}{3} \frac{m_e m_g}{(m_e + m_g)^2} \left(1 - \frac{E_g}{E_e} \right) \qquad (769)$$

m_e, m_g are the masses, E_e, E_g the energies, of the electron and atom.   In the present case this reduces to $f = 2.67\ m_e/m_g$ because of the relatively small atomic heat energies.   The numerical value of f is therefore 2.67/72,800, a very small number.   Because an electron on the average transfers only this very small fraction of its energy to an atom during an elastic collision, a ten-volt electron is still a ten-volt electron after making an elastic collision.   As long as the collisions are elastic, and the kinetic energy given to the atoms negligible, *an electron's energy change depends only on its progress in the direction of the field.*

Suppose, for example, that the anode potential and spacing are respectively 100 volts and 5 centimeters, making the potential gradient 20 volts per centimeter or 2 volts per millimeter, as in Fig. 153.   If as in trajectory $A A A A$ in that figure the encounter just after a 10-volt

collision happens to be on the rebound, 2 millimeters behind the 10-volt

FIG. 153. Typical electron trajectories *AAAA* and *BBBB*. The circles represent gas particles.

collision, the electron strikes with only 6 volts of energy, and must make another inelastic collision. The *BBBB* trajectory illustrates a rebound that advances the electron in the direction of the field. The electron enters the second *BBBB* collision with 12 volts of energy, for it has advanced one millimeter in the direction of the field since the first collision.

This second *BBBB* collision may or may not be an elastic one. The electron enters it with enough energy to cause excitation, for the least excitation energy of argon is 11.5 electron volts. If excitation does occur, the electron emerges from the collision with only 0.5 volt of energy, and begins a new sequence of elastic collisions, between which it again builds up to the excitation or higher energy.

(a) Energy variations.
(b) Path of the electron's motion.
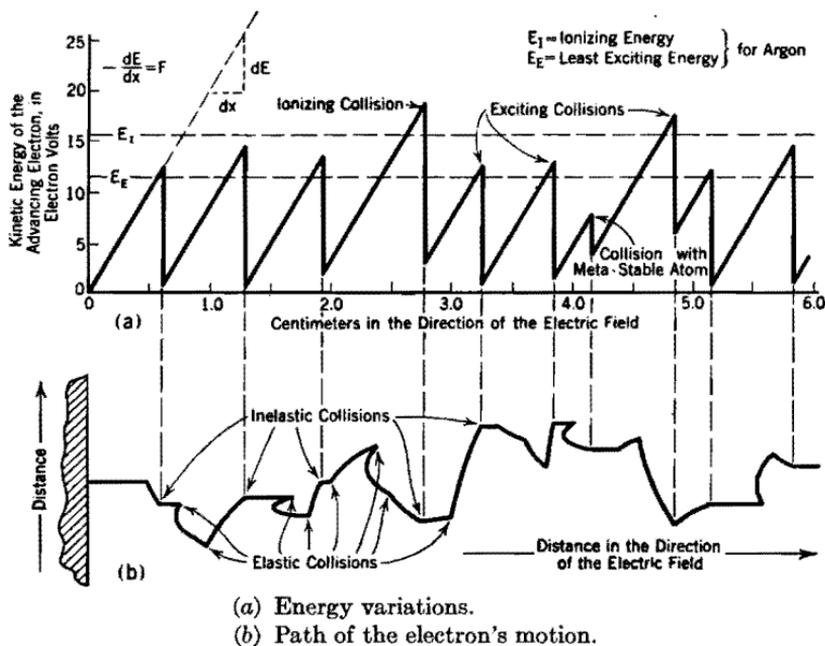
FIG. 154. Advance of an electron driven by an electric field through argon at low pressure.

But excitation may not occur at the second *BBBB* collision.   When the electron's energy exceeds the excitation energy, there is a definite *probability of excitation;* perhaps between three per cent and thirty per cent of the 12-volt collisions result in excitation.   If excitation does not occur, the collision is an elastic one; the electron travels on, still with its 12 volts of energy.   It advances, acquiring more energy, until an inelastic collision does occur.

If the first inelastic collision occurs after the electron's energy has exceeded the ionizing potential, there is a definite (though perhaps unknown) *probability* of ionization's occurring.

Thus a great variety of things happen to an electron as it travels toward the anode.   Its energy builds up, drops off, builds up, drops off, in the manner illustrated by Fig. 154*a*.   Each slant line represents the building-up process, and may include several free paths, as shown in Fig. 154*b*.   The inelastic collisions result in vertical drops in the energy line; the elastic collisions have no effect on the shape of this line, for they do not affect the electron's energy. Ordinarily the peaks in the saw-tooth line must lie above the dotted excitation-energy line at 11.5 volts.   Very rarely a peak like that at about 4.2 centimeters may occur, as a result of a collision with a metastable atom which can accept less than 11.5 electron volts of

Fig. 155.   One electron emitted from the cathode; 14 ionizing collisions; 15 electrons enter the anode.

energy.   Ionization is represented on the saw-tooth diagram by a vertical drop of 15.6 volts.

Amplification of the initial photoelectric current results from the fact that each ionizing collision results in the production of a new electron as well as an ion.   The new electron pursues its own journey to the anode, exciting and ionizing atoms as it goes.   Each electron so produced makes more electrons.   The more electrons there are, the faster the electron stream grows.   Fig. 155 illustrates how the ionizing collisions make each original photoelectron become the parent of an extensive genealogical tree.   In that figure 14 ionizing collisions are represented, so that 15 electrons enter the anode as a result of one's leaving the cathode.   If this represented the average behavior in a particular de-

vice, the actual tube current would be 15 times the photoelectric emission current.

The large mass and size of the ions produced by electron impact make their drift toward the cathode very sluggish. They are very unlikely to make ionizing collisions with atoms.

219. Dependence of Amplification on Electrode Spacing and on Ionization Rate. The amount of amplification obtained by the use of gas in a phototube depends chiefly on: (1) the spacing between the electrodes, and (2) the number of new electrons produced per centimeter of each electron's advance in the direction of the field. This number of new electrons per centimeter of advance will be called α'. The prime is used to avoid confusion with the symbol α used in Chapter X to describe the velocity that is characteristic of a given temperature. α' can of course be described as the number of ionizing collisions per centimeter of an electron's advance, because each such collision produces a new electron. α' depends on the gas concentration, on the exciting and ionizing potentials and probabilities for the gas used, and on the ratio of field strength to gas concentration.[LL 79, MM, OO]

Suppose L electrons per square centimeter per second cross a plane at x centimeters from the cathode of Fig. 152a, moving toward the anode. In going a differential distance dx each one of them produces $\alpha' dx$ ions, and the same number of electrons. This results in a change dL in the electron flow, thus

$$dL = L\alpha' dx \qquad (770)$$

In terms of electron current density J_e, obtained by multiplying L by the electronic charge, this is

$$dJ_e = J_e \alpha' dx \qquad (771)$$

Integrating this between $x = 0$ and $x = s$

$$J_e = J_{Ph} \epsilon^{\alpha's} \qquad (772)$$

where J_{Ph} is the current density at $x = 0$, and is of course photoelectric in origin. Equation (772) expresses the fact that the current in the tube is *proportional to the photoelectric emission current*. The proportionality factor is an exponential function of $\alpha's$. Evidently the amplification increases rapidly with an increase in the spacing, and with any growth in α'.

220. Dependence of Ionization Rate on Gas Concentration and on Field Strength. The nature of the dependence of α' on the concentration of gas particles and on electric field strength will be brought out by a comparison of Figs. 156a, b, and c. All three of these diagrams are similar in nature to Fig. 154a.

The top one of the three, Fig. 156a, represents a typical electron's energy variation along the cathode-to-anode journey when the gas concentration is so low that there are very few collisions. The electron mean free path is very long. The collisions that do occur involve considerable energy; however, experiments have shown that the mean

(a) Low gas concentration.
(b) High gas concentration.
(c) Intermediate gas concentration.
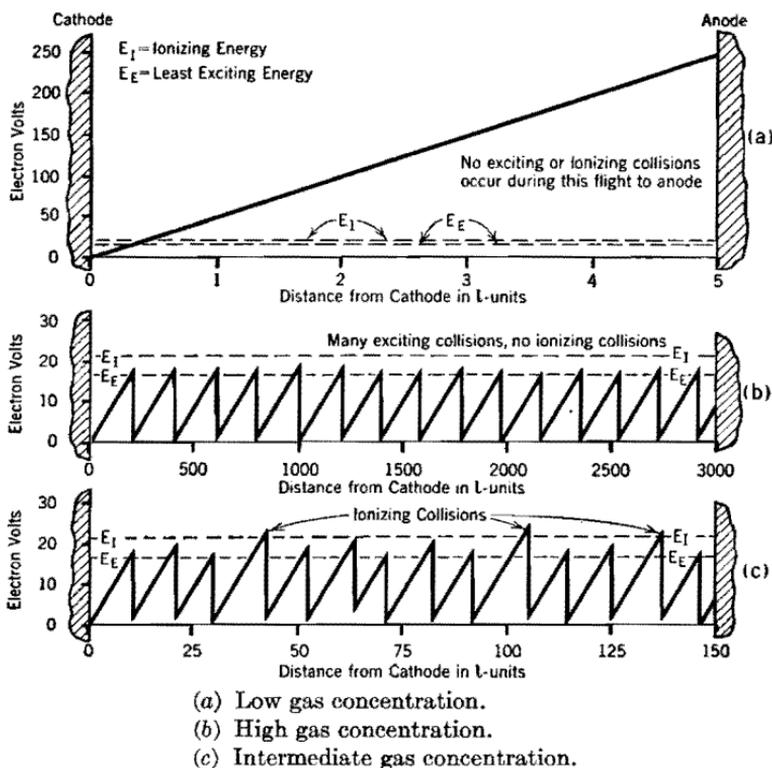
Fig. 156. Dependence of ion production on gas concentration. An l-unit of distance is one electron mean free path.

free paths of high-velocity electrons are relatively long (see Table XV) and that only a moderate fraction of the collisions with energies well above ionizing energy result in ionization.[98]

Because collisions of all kinds occur infrequently, and not all of those that do occur are ionizing collisions, very few new electrons are produced by electron impact when the gas concentration is low. Most of the electrons, like the one whose uneventful journey is diagrammed in Fig. 156a, pass directly to the anode without experiencing either exciting or ionizing collisions. α' is vanishingly small, and the phototube current is just the photoelectric emission current. There is little or no amplification.

In Fig. 156b (note the change in the vertical scale) the behavior if the gas concentration is very great is indicated. There are many collisions per centimeter of path, so many, in fact, that long before an electron acquires ionizing energy it experiences an exciting collision that drops its energy back near to zero. Thus again ionization is very unlikely, α' is very small, and there is little or no amplification.

Fig. 156c illustrates the behavior with an intermediate gas concentration. There are not as many collisions per volt along the rising front of a saw-tooth as in Fig. 156b, because the gas particles are farther apart. After the E_E level is passed along any one rising front, there is an appreciable probability [98] that each collision will be inelastic, so terminating the saw-tooth. However, there are not nearly as many such opportunities for termination, per volt of rise, in Fig. 156c as in Fig. 156b, so that occasional saw-teeth in Fig. 156c do rise above the ionization level. The collisions that terminate these tall saw-teeth result in ionization of gas particles and consequent introduction of new electrons. α' has an appreciable value, and there may be considerable amplification. At ordinary room temperatures, this situation exists when the gas pressure is a few millimeters of mercury.

The distance scales in Fig. 156 are marked in units whose length is one electron mean free path, l centimeters. These units will be called l-units. For convenience the symbol α'' will be used to describe the number of electrons produced per 100 l-units of electron advance. From the definition of α''

$$\alpha'' = \alpha' \cdot 100l \tag{773}$$

The use of Equation (487) converts Equation (773) into

$$\alpha' = \frac{\alpha''}{100l} = N_g \alpha'' \frac{\pi b^2}{100} \tag{774}$$

N_g stands for gas concentration, b for the gas particle radius. These equations show that as long as α'' remains constant, α' is proportional to the gas concentration.

The next few paragraphs are devoted to a demonstration that α'' is a function of the slope of the fronts of the saw-teeth in Fig. 156c, the slope being measured in volts per l-unit. α'' is the product of (1) the number of saw-teeth per 100 l-units by (2) the fraction of the total number that rise above the ionization level. It is necessary to show that both factors of this product depend on the slope just mentioned.

As an electron's energy increases along the rising front of a saw-tooth, the electron experiences collisions. The slope of the rising front determines the number of collisions that occur between the E_E and E_I levels, for two reasons, as follows:

(*a*) The slope of the rising front determines, inversely, *the number of l-units traversed* in rising from the E_E to the E_I level.

(*b*) Total motion, not just forward motion, produces collisions. For this reason the number of collisions *per l-unit* of forward motion depends in some direct way on the ratio of total motion to forward motion; it is in fact proportional to the ratio

$$\frac{\text{Total energy}}{\text{Forward energy}}$$

Rebounds at elastic collisions take place in random directions, so that, on the average, forward motion does not persist after the free path of its origin. Therefore whatever forward energy an electron has is acquired from the field between *successive* collisions, and is therefore proportional to the slope of the rising front of the saw-teeth. Hence the number of collisions per *l*-unit of forward motion is also proportional to the ratio:

$$\frac{\text{Total energy}}{\text{Slope}}$$

Thus the slope determines, again inversely, *the number of collisions per l-unit of forward motion.* The total energy to be used in these ratios is some average value intermediate between E_E and E_I, because only collisions in that range are of interest.

The result of this double dependence on slope is to make the number of collisions along the rising front of a saw-tooth between the E_E and E_I levels *inversely* proportional to the *square* of the slope in volts per *l*-unit. If there are only a few such collisions, because the slope is steep, there is an appreciable chance that any saw-tooth will rise above E_I, so that the fraction

$$\frac{\text{Number of saw-teeth that terminate above the } E_I \text{ level}}{\text{Total number of saw-teeth}}$$

has an appreciable value. The steeper the slope becomes, the larger this fraction becomes. This fraction also depends to some extent on two properties of the gas particles, namely, the spacing between the E_E and E_I levels, and the probability of excitation in case of collision between these levels.[98] However, as long as the slope remains constant, and the same gas is used, this fraction remains the same. This establishes the dependence on the saw-tooth front slope, in volts per *l*-unit, of one of the two factors that constitute α''.

The average saw-tooth width, in *l*-units, varies:

(*a*) Approximately inversely as the slope, for the saw-tooth heights are approximately fixed, in that nearly all of them terminate in the relatively narrow excitation band between the E_E and E_I levels. This important dependence of saw-tooth width on slope is of an inverse first-degree type.

(*b*) According to minor variations in the average height of termination of the saw-teeth, within this narrow band. Any effects that tend to make many saw-teeth rise

above E_I also tend to increase the average height of the saw-teeth. For this reason the average saw-tooth width, like the per cent of tall saw-teeth, is dependent in some direct way on the square of the slope. Since this direct dependence affects only the increment in addition to a minimum saw-tooth width, corresponding to the minimum height E_E, it is relatively unimportant.

Only as long as the slope and gas are unchanged do the average saw-tooth *width*, so the number of teeth per 100 l-units, remain constant, according to both (a) and (b). This establishes the dependence on slope of the second of the two α'' factors. Thus both factors of α'', therefore α'' itself, are constant only as long as the saw-tooth slope, in volts per l-unit, is unchanged, and the same gas used.

Now suppose the gas concentration is increased. According to Equation (487) the electron mean free path is thereby shortened. If it is desired to keep the saw-tooth slope in volts per electron mean free path constant, the electric field strength must be increased, and in the same ratio as the gas concentration. Thus the slope and therefore α'' remain constant as long as the ratio F/N_g of electric field strength to gas concentration is constant. Conversely, α'' changes whenever that ratio changes. In other words, α'' is some *function* of that ratio.[127]

The fact that α'' is a function of F/N_g can be used to convert Equation (774) into the following:

$$\alpha' = N_g\phi\left(\frac{F}{N_g}\right), \quad \text{or} \quad \frac{\alpha'}{N_g} = \phi\left(\frac{F}{N_g}\right) \tag{775}$$

The function $\phi(F/N_g)$ vanishes when F/N_g is very small (Fig. 156b) and N_g vanishes when the gas becomes sufficiently rarified (Fig. 156a). There is a definite value of gas concentration at which α' and the amplification are a maximum, for given field strength.

Equation (775) appears frequently in the technical literature relative to the type of gaseous-conduction current here discussed, which is called a *Townsend current* to distinguish it from the very different type of conduction current that flows in electric arcs and glow discharges. The validity of the general relation expressed by Equation (775) has been well established by experiment.[LL 82, MM]

The above discussion illustrates the convenience of using the electron mean free path as a yardstick in the study of conducting gases. Values of electron mean free paths in various gases, for various electron energies, can be determined from Table XV.

221. Gas Amplification Limited by Space Charge.[GG VII, FF III] In a phototube employing gas amplification many more electrons arrive at the anode per second than are photoelectrically emitted from the cathode per second. However, as in any element in a closed circuit, the current

entering the tube at one terminal must be the same as the current leaving at the other terminal.   Therefore as many electrons must leave the cathode per second, in one way or another, as enter the anode per second.

Of course each time a new electron is produced by collision in the gas, a new ion is also formed.   The ions drift to the cathode under the influence of the electric field, their movement being very slow because their large mass makes the ion *mobility* g_i very small.   Mobility, discussed later in Sections 233–237, is the ratio of average drift velocity to the electric field strength that produces it, and is measured in centimeters per second per volt per centimeter.   All the ions do ultimately arrive at the cathode.   Upon arrival each one steals an electron from the cathode surface and becomes a neutral gas particle again.   This explains how the extra electrons leave the cathode.

In the apparatus represented by Fig. 152a, let J_{Ph}, J_e, and J_i symbolize respectively the photoelectric emission current density from the cathode, the current density at any point in the gas due to electron flow, and the current density at any point in the gas due to positive ion flow.   Let v_e and v_i represent the velocities of electron and ion drift respectively, ρ_e and ρ_i the respective space-charge densities, and g_e and g_i the respective mobilities.

Now suppose the gas amplification to be 10 to 1.   This requires that at all points in the tube the total current density must be $10J_{Ph}$.   Just outside the anode the entire current is due to electron flow, because the anode is not a source of ions, so no ions can be moving away from it. Therefore, just outside the anode,

$$J_e = 10J_{Ph} \tag{776}$$

Just outside the cathode the electron current density is J_{Ph}.   Therefore, in order to make up the requisite total current, the positive ion current density, represented by the ions moving toward the cathode, must be, just outside the cathode,

$$J_i = 9J_{Ph} \tag{777}$$

At all points within the tube

$$J_e = \rho_e v_e, \text{ or } J_e = \rho_e F \cdot g_e \tag{778}$$

$$J_i = \rho_i v_i, \text{ or } J_i = \rho_i F \cdot g_i \tag{779}$$

g_i is only a fraction of a per cent of g_e, because of the contrast between the masses of the two kinds of particles.   For this reason the product $\rho_i F$ near the cathode must be very much larger than the product $\rho_e F$ near the anode, even though J_i near the cathode is about the same as J_e

near the anode.   A very small electron space charge driven by a gentle gradient is to be expected near the anode, but a considerable positive ion space charge driven by a steep gradient must exist near the cathode. The solid line in Fig. 157 illustrates the type of potential distribution that corresponds to these requirements in the geometry illustrated by Fig. 152a, page 402.

Suppose the illumination is increased. J_{Ph} grows in proportion to the illumination, and to a first approximation so also do J_e and J_i. However, if the illumination becomes intense enough the space charge near the cathode becomes so large as to flatten out materially the potential distribution in the middle region and near the anode.   This condition is illustrated by the dotted line in Fig. 157.   According to Fig. 155 and Equation (772) a very important part of the amplification is that due to the ions produced in and beyond the middle region.   If the curve in the middle region flattens out, α' there is reduced, so that the amplification becomes less.   This explains why the amplification may be less for very intense than for moderate illumination.   Under some conditions an increase in the illumination may reduce the total tube current.
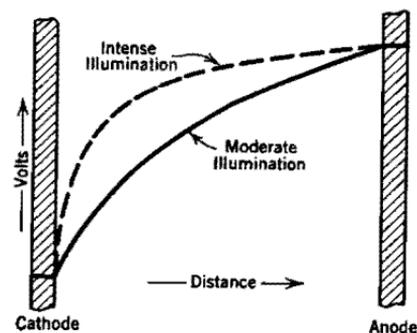
Fig. 157.   Potential distribution in a gas-filled phototube with parallel plane electrodes.

The effect of positive ion space charge on tube operation is much less pronounced in a tube with a cylindrical cathode surrounding a small anode cylinder than in the parallel-plane geometry of Fig. 152a.   In the cylindrical device the area of cathode surface is much greater than the area of the anode surface.   Therefore J_i near the cathode is only a small fraction of J_e near the anode.   This contrast tends to compensate for the similar contrast between the mobilities, so that only a moderate space charge is necessary near the cathode.   Furthermore, because of the nature of the geometry, the electric field near the anode will remain relatively steep even after the appearance of a considerable positive space charge near the cathode.

With any geometry, a definite equilibrium condition corresponds to any given value of J_{Ph}.   In the equilibrium condition the electric field in the middle region is just steep enough to produce enough ions to set up a space charge near the cathode that will maintain the existing intermediate field.

When the phototube current is large because of intense illumination,

there is likely to be a steep gradient just outside the cathode, as illustrated by the dotted line in Fig. 157. This steep gradient causes tendencies toward (a) release of electrons from the cathode by positive ion bombardment, and (b) ionization by positive ions in the gas near the cathode. Probably the (a) effect is very much more important than the (b) effect in all ordinary apparatus.

If either or both of these effects are moderate, they merely increase the amplification somewhat, so more or less completely offset the space charge limitation to amplification. If either effect is pronounced, a glow discharge forms between the electrodes. The severe positive ion bombardment of the cathode that accompanies a glow discharge will rapidly destroy the specially prepared photoelectric surfaces of most phototubes.

A glow discharge is self-sustaining, and the current through it in no way dependent on the illumination of the tube. The type of current that flows during normal use is not self-sustaining, and stops entirely when the light is cut off.

Gas amplification can be used in connection with thermionic currents that are subject to grid control. The effect of gas on thermionic currents, especially space-charge-limited ones, is complicated by the accumulation of positive ions near the cathode that always accompanies gas amplification.

222. Volt-Ampere Properties of Gas-Filled Phototubes; Phototube Circuit Analysis.[GG VII, FF III] Fig. 158a illustrates the nature of the volt-ampere characteristics to be expected of a gas-filled phototube. A glow discharge is initiated, in the manner described in the previous section, if operation is carried across the dotted curve ABC.

No matter how weak the illumination may be, a glow discharge forms if the anode voltage reaches that at A, here about 109 volts. On the other hand, no matter how strong the illumination may be, a glow discharge cannot form if the anode potential is less than the BC value, about 75 volts. Furthermore, reduction of the anode voltage to less than about the BC value discontinues a glow discharge that has previously been formed. The widening of the dotted line AB to a dotted region BC indicates that above B the transition from the ordinary light-controlled phototube current to a self-sustaining glow discharge takes place gradually rather than abruptly.

The effect on the current in a simple direct-current phototube circuit, Fig. 158b, of changes in light intensity, is predictable by the use of a load line like those used in the analysis of amplifier circuits. Such a load line has been drawn in Fig. 158a. Variations in light intensity cause shifts from point to point along the load line.

The approximate proportionality between phototube voltage and

current indicated by Fig. 158a is of particular interest in that it shows that for the current-carrying polarity a gas-filled phototube acts much like a resistance whose magnitude is inversely proportional to the light intensity. A phototube is of course also a rectifier.

An alternating component of current can exist in a phototube circuit either because there is in the circuit an alternating voltage source, or because the light intensity varies periodically.

Suppose that steady light, of such intensity as to produce the solid volt-ampere line in Fig. 159a, strikes the tube in the circuit diagrammed
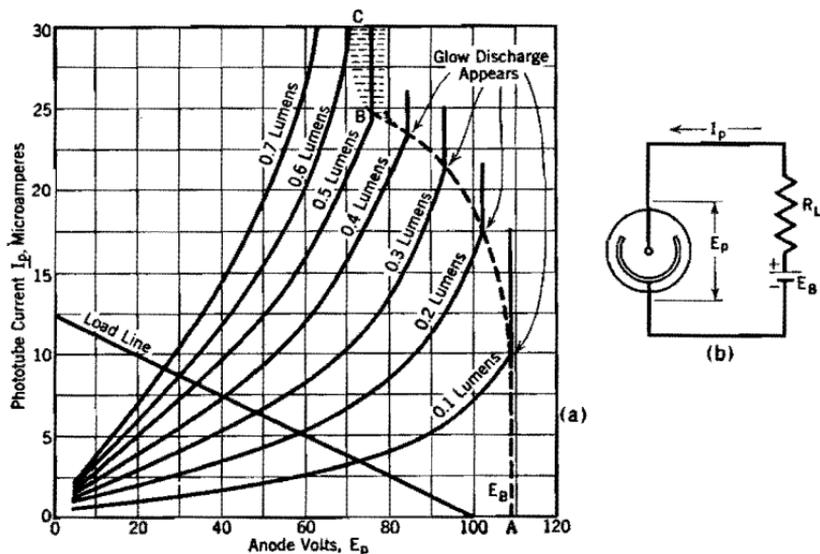
FIG. 158.   Volt-ampere characteristic curves for a gas-filled phototube, for various intensities of illumination; also, load line corresponding to a simple direct-current circuit.

in Fig. 159b. E_1 represents an alternating voltage introduced into the circuit by a transformer whose primary is connected to some alternating-current power source.   The magnitude of the alternating-current voltage E_2 that appears across the load resistance R_L can then be predicted by means of the equivalent circuit shown in Fig. 159c.   The quantity R_p should be defined, as in amplifying tubes, as $\partial E_p/\partial I_p$; it is approximately an inverse function of the light intensity.

A method of graphically representing the effects of periodic variations in light intensity, with fixed plate battery voltage, is illustrated in Fig. 160.   Owing to the fact that R_p varies approximately *inversely* as the light intensity, a single-frequency sinusoidally varying light beam produces harmonics in the alternating voltage across the load resistance.
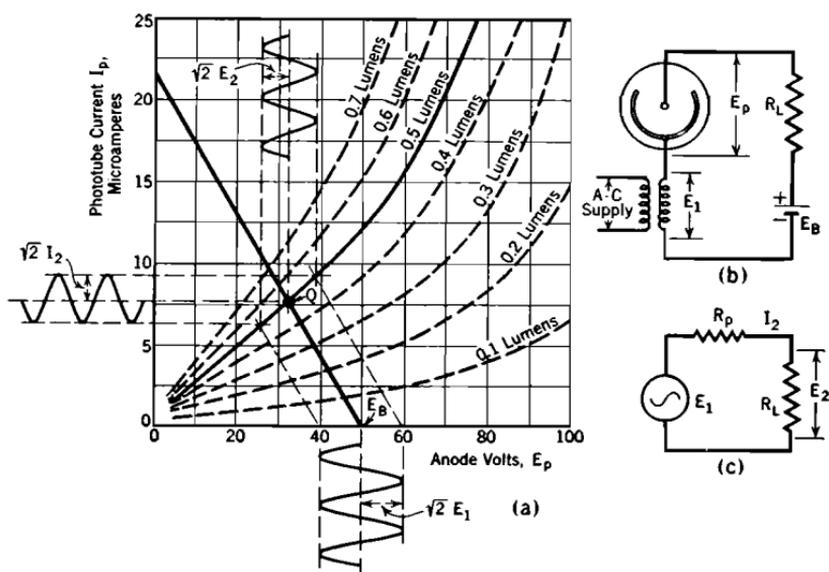
(a) Tube characteristics and load line.

(b) Actual circuit.

(c) Equivalent alternating-current circuit. $R_p = \left(\dfrac{\partial E_p}{\partial I_p}\right)$

FIG. 159. Gas-filled phototube in a circuit containing an alternating-current as well as a direct-current voltage.
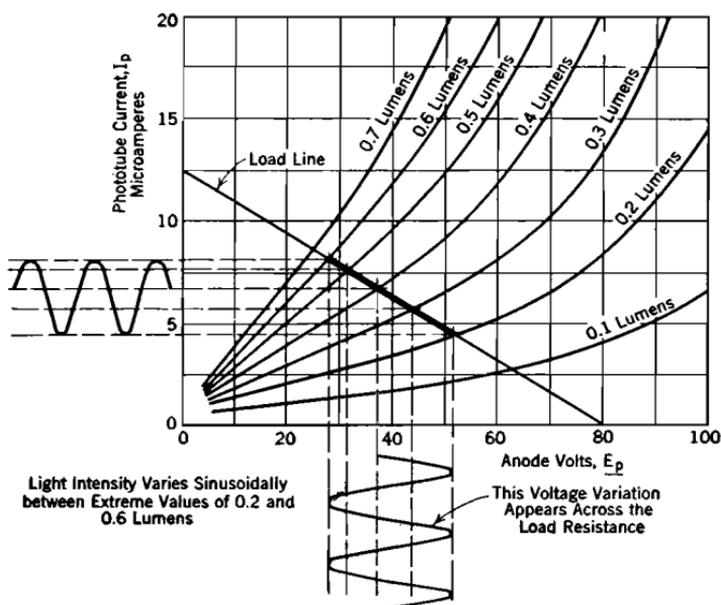
FIG. 160. Variations in phototube current and voltage resulting from sinusoidal variations in light intensity. The heavy portion of the load line is the current-voltage locus. Circuit as in Fig. 158b.

If the ratio of the change in light to total light is small, the harmonics are not pronounced.

Various other phototube circuit arrangements can be analyzed either graphically, or mathematically by the use of empirical equations for the useful portions of the characteristic curves.

223. Rectifier-Type or "Sandwich" Photocells; Semiconductors.[68, GG IX, FF IV, HH IX]   In the types of phototubes so far described the energy associated with current flow in the external circuit comes chiefly from the battery, the magnitude of the current being controlled by the light. A rectifier-type or " sandwich " photocell, illustrated by Fig. 161a, itself provides whatever energy appears in the external circuit, the incoming light being the primary energy source. There is no auxiliary battery. The light generates an electric current rather
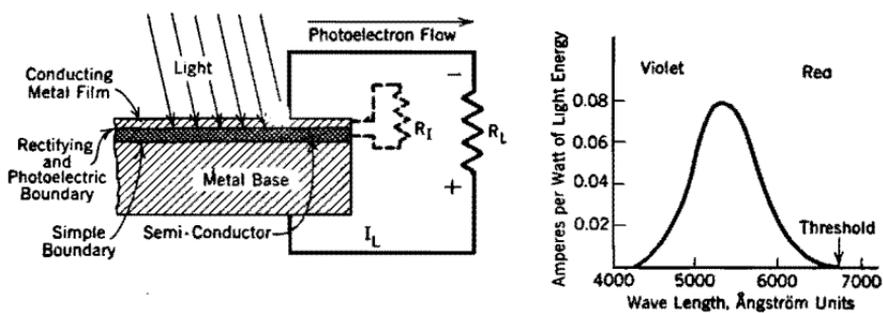
(a) Cell and circuit.   (b) Typical color sensitivity curve.

FIG. 161.   Rectifier-type photocell.

than an electric potential, but the voltages against which such a cell can drive current are small. There is a general similarity between the operation of a rectifier-type photocell and that of a central cathode vacuum-type cell (Section 214) connected into a circuit that contains a resistance but no battery.

The geometry of a rectifier-type photocell, Fig. 161a, suggests the descriptive term " sandwich." The light penetrates through a thin metal film to the boundary between the film and a *semiconductor*. Usually the semiconductor is either cuprous oxide (Cu_2O) or selenium. When such a " front-wall " cell is illuminated the direction of electron flow within the cell is opposite to the direction of light entrance. In the external circuit the photoelectrons flow in the direction indicated in the figure. In " back-wall " cells, not now widely used, light is made to enter through the semiconductor rather than through the metal. In all cases the photoelectrons travel from semiconductor to metal across the rectifying and photoelectric boundary.

A cell assembly like that illustrated in Fig. 161a serves as a rectifier

when used in connection with an external alternating-current power source; that is the reason for the use of the term " rectifier-type photo-cell." When driven by an external voltage source, electrons flow freely from copper to oxide; but the resistance to flow from oxide to copper is very high. An individual sandwich serves satisfactorily for rectification if the alternating-current voltage does not exceed a small value. Stacks of sandwiches can serve effectively as rectifiers for alternating-current potentials of several hundred volts.

A *rectifying and photoelectric boundary* between metal and semi-conductor may be formed by sputtering (see Section 247) a thin film of metal (gold and platinum have been used) over the surface of the semi-conductor. An alternative method that has been used in preparing sandwich rectifiers and back-wall photocells is to form an extremely thin layer of cuprous oxide over copper by heating the copper in air at about 1040° C, the processing being completed by annealing at about 600° C. Still other methods can be used for forming this type of boundary over selenium.

A *simple boundary* is formed by pressing the base onto the semi-conductor. A simple boundary has neither rectifying nor photo-electric properties.

In Chapters VIII and IX the behavior of electrons in metals was described in terms of energy levels, as illustrated in Fig. 69, page 165. At absolute zero all the levels below the normal maximum are com-pletely filled, those above the normal maximum completely empty. At room temperature a very few of the electrons occupy normally vacant levels that are just above the normal maximum level. These low-lying normally vacant levels are sometimes called *running levels*. This ter-minology is used because the flow of electrons from one part of the metal to another is believed to be due chiefly to the movement of the electrons in these levels. Such electrons have energy enough to enable them to pass through or over internal potential barriers such as those between crystals and grains. Boundaries of this sort, symbolized in Fig. 162*a*, *b*, and *c* by vertical lines, are believed to be to a considerable extent responsible for the electrical resistance of metals. The application of an electric field to a metal corresponds to a tipping of the energy levels into an inclined position. The electrons then flow from each compart-ment into the next over the tops of the intervening barriers, in some-what the way water flows down a fish-ladder.[70]

In insulators the electrons of the atoms may be represented as re-siding in definite kinetic energy levels, just as in conductors and in individual atoms. However, the electrons in insulators are firmly attached to individual molecules, and cannot readily accept additional

kinetic energy or move about.   Correspondingly, there is in the energy-level diagram for an insulator a very extensive " no-level " range of energy values just above the filled levels, as illustrated in Fig. 162b. Quantum requirements forbid the existence of energy states within the no-level band.

Thus in an insulator the normal maximum level is an absolute maximum level; no electrons can rise above it.   Consequently the levels may be tipped to a steep incline without any electrons passing from one

(a) Metal conductor.   (b) Insulator.   (c) Semi-conductor.

(d) Rectifying and photoelectric boundary between a metal and a semi-conductor at equilibrium.
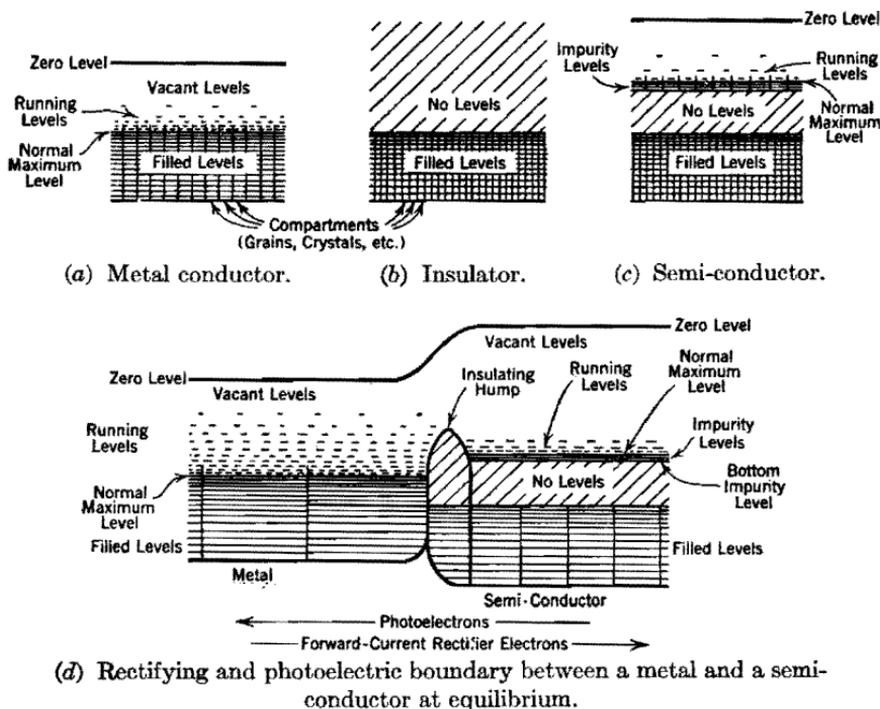
Fig. 162.   Metal, insulator, and semi-conductor energy-level diagrams.

compartment to the next; that is, a steep gradient does not produce a current.   The fish-ladder has been covered over at the tops of the steps.

Semiconductor properties represent a sort of composite of the two extremes just described.   As illustrated in Fig. 162c, a semiconductor has " running levels " above a thin layer of filled levels that are in turn above a no-level band.   The few high-up filled levels are sometimes called " impurity levels " because they are presumed to contain, in many substances, the electrons of impurities that are present.   For example, in cuprous oxide the impurity levels are believed to be those containing the electrons associated with excess oxygen, that is, oxygen present over and above that necessary to satisfy the chemical formula Cu_2O for

cuprous oxide. This supposition is based partly on the known fact that the conductivity of cuprous oxide is closely related to the excess oxygen content.

The conductivity of a semiconductor is in general limited by the fact that there can be but few electrons in the running levels, partly at least because the impurity-level reservoir that supplies them is shallow. An increase of the temperature of a semiconductor increases its conductivity, presumably by increasing the number and height of the electrons that rise out of impurity levels into running levels.

According to one fairly convincing hypothesis a reducing action that accompanies the sputtering-on of the metal film of a cuprous oxide cell results in the removal of the excess oxygen from an extremely thin layer of oxide adjacent to the metal film. Therefore this extremely thin layer becomes an insulating film. Within it there are no running levels, so that a thin insulating " hump " appears in the potential-energy curve for the photoelectric and rectifying boundary, as illustrated in Fig. 162d. Other manufacturing processes presumably produce similar insulating humps, though perhaps in very different ways.

A hump of the type illustrated in Fig. 162d may be the result of a negative space charge, as Grondahl[126] has suggested. According to his point of view the high-density electron gas in the copper tends to diffuse into the oxide, which contains a low-density electron gas. The diffusing electrons on entering the oxide set up a negative space charge which restricts further diffusion to an equilibrium rate. This space charge produces a convex-upward hump in the potential-energy curve, because it produces a downward convexity in the potential curve (see Chapter I). The effect of a potential hump would be essentially the same whether attributed to an insulating layer or to space charge.

In the absence of light or any external applied voltage a boundary like that illustrated in Fig. 162d reaches an equilibrium condition in which the transfer of electrons from semiconductor to metal occurs just as rapidly as the transfer from metal to semiconductor.[68, 69, 70] Section 106 contains a discussion of the transfer of electrons through humps.

Any electron's transfer through the hump must begin and end at the same level. Subject to this limitation, transfer can take place from either running or filled levels to either vacant or running levels. Transfer cannot terminate in filled levels because no additional electrons can be accommodated therein. Transfer cannot originate in either vacant or no-level ranges of energy values because there are no electrons there to be transferred. Transfer either way is of course impossible at levels below the bottom impurity level of the semiconductor.

At levels just above the semiconductor's normal maximum, transfer

can take place both ways. Presumably the direction and rate of net transfer at any such level depends primarily on the difference between the electron concentrations at that level on the two sides of the hump, and on the thickness of the hump at that level. Within the impurity filled level range transfer takes place only from right to left, and at a rate that is independent of the concentration in the running levels at the left of the hump, but does depend on hump thickness.

In the equilibrium condition, illustrated in Fig. 162d, the relative heights of the levels on the two sides of the hump are such that the net left-to-right transfer through the hump between running levels is presumably just equal to the right-to-left transfer from the impurity filled levels, so that the overall net transfer is zero.

If an external voltage is applied that *lowers* all the semiconductor's levels, many more electrons pass from left to right than before, because the semiconductor's lowest running levels are shifted downward into line with high-concentration metal running levels. Furthermore, there may conceivably be free flow over the top of the hump. Thus electrons flow freely from left to right in response to a slight applied voltage. This is the direction of forward-current electron flow when the boundary is used as a rectifier. Once electrons have entered the semiconductor's running levels, they can be rapidly removed to a suitable terminal by a moderately strong electric field.

The free flow of forward current through the interior of the semiconductor takes place at a relatively high velocity of travel and small electron concentration. Free entrance of electrons into the semiconductor cannot increase the total concentration of electrons within the semiconductor, for there must always be just the right number of electrons present to neutralize the positive charges on the ions of the structure. Only the *distribution*, among the various energy levels, of the total electron content can be affected by heat, light, or electric circuit action.

If an external voltage is applied that *elevates* the semiconductor's levels in Fig. 162d, the semiconductor's running levels are shifted into line with vacant levels on the other side of the hump, so that only right-to-left transfer can take place. This constitutes an inverse rectifier current. However, the maximum inverse current is small, because of the small concentration in and therefore small transfer from the semiconductor's filled and running levels. Thus the boundary offers a high resistance to current flow in the inverse direction.

The explanation that has just been given of the rectifying action of a semiconductor "sandwich" is a fairly generally accepted one, at least in regard to its major features. As a little study of Fig. 162d will show,

an almost equally satisfactory explanation can be built on the assumption that all electron transfer takes place over the top of the hump, no penetration through the hump occurring. On the other hand, rectifier action at least can be satisfactorily explained on the assumption that the hump is infinitely high and of uniform thickness.

Various explanations, most of them based on diagrams in general similar to Fig. 162d, have been advanced to account for the observed photoelectric properties of rectifier-type photocells. The one given below is presented because it is easy to follow, and illustrates nicely the kind of reasoning that can be used in connection with energy-level diagrams, rather than because of any great confidence that it can withstand critical examination.

Light that reaches the boundary illustrated in Fig. 162d affects chiefly the electrons at and below the normal maximum levels in the two substances. At ordinary room temperatures the electron concentrations in the running levels are too small to permit such electrons to receive a large fraction of the total light energy delivered, even though they may be large enough to supply forward rectifier current freely in the manner described above.

The observed photoelectric behavior becomes understandable if it is assumed that the photoelectric transfer of electrons is the result of electrons, at or below normal maximum levels, being lifted entirely over the top of the insulating hump. At and just above the threshold energy, E_{Ph} is large enough to lift electrons from the semiconductor's normal maximum level high enough to pass over the hump, but it is not large enough similarly to lift normal maximum copper electrons up and over the hump. Therefore only photoelectric electron transfer from right to left takes place. However, if E_{Ph} is made still greater, as at and beyond the violet end of the visible spectrum, electrons are transferred in both directions over the hump photoelectrically. Ultimately the two opposite rates of transfer are equal, so that the photoelectric current in the external circuit falls to zero, in accordance with Fig. 161b.

Experiments have shown that a rectifier-type photocell should be thought of as a constant-current rather than as a constant-potential electric generator. At a given light intensity and wave length, electrons in some definite number per second are transferred photoelectrically from semiconductor to metal. A satisfactory equivalent circuit is shown in Fig. 163a.

R_I is the resistance to internal return flow, that is, passage of the photoelectrons back through the rectifying and photoelectric boundary, while R_L is the external or load resistance. When R_L is a meter of very low resistance the cell is short-circuited. Practically all the photo-

electrons then pass through the meter, and but very little return current flows internally. Measurements have shown that under these conditions the current through the short-circuiting meter is directly proportional to the light intensity, as indicated by the straight line through the origin in Fig. 163b. C represents the internal capacitance of the cell.

When R_L is large, practically all the photoelectrons return internally, through R_I, so that the measured external voltage is the potential drop through the internal parallel path. Of course this internal drop depends on the internal current, and might therefore be expected to be proportional to light intensity. However, the rise in terminal voltage results in a decrease in the internal resistance, so that the open-circuit

(a) Equivalent circuit.   (b) Response to variations in light intensity.
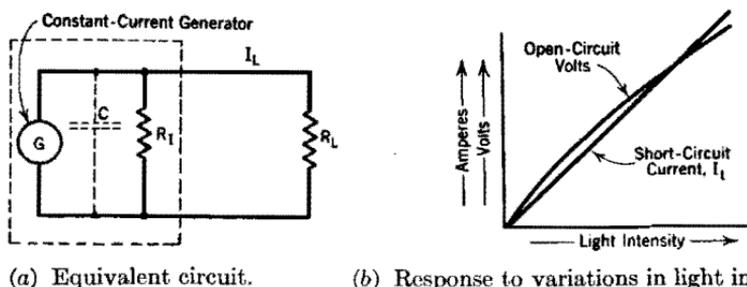
Fig. 163.   Rectifier-type photocell properties.

voltage is in fact only approximately proportional to the light intensity, as indicated by the curved line in Fig. 163b.

R_I is of course inversely proportional to the area of the cell, in square centimeters. For that reason it is undesirable to have more cell area than is exposed to light, if the internal return current is to be kept to a small fraction of the total photoelectric current. With a cell area of 5 square centimeters it is reasonable to expect a cuprous oxide cell to have an internal resistance measurable in hundreds of ohms, and a selenium rectifier-type cell to have an internal resistance measurable in thousands of ohms.

The photoelectric yield and quantum yield of rectifier-type photocells are much greater than those for photoelectrically emitting surfaces of the types used in vacuum and gas-filled phototubes. The overall sensitivity, in current units per lumen, of a rectifier-type cell is likely to be a little better than that for a gas-filled phototube. Rectifier-type cells possess the merits of simplicity and economy. Large enough areas might conceivably be prepared to generate substantial amounts of power from sunlight.

However, it is not feasible with rectifier-type cells, as it is with

emission cells, both vacuum and gas-filled, to make the currents in relatively high-voltage circuits directly dependent on illumination. Furthermore, rectifier cells cannot always be conveniently used to excite the grid circuits of electronic amplifiers, but emission phototubes are very well adapted to such use.

224. Photoconducting Cells.[FF, GG, HH]   In all of the photosensitive devices so far discussed the light produces a transfer of electrons across some kind of potential barrier. In other words, the light either produces or directly releases an electron flow.

In a photoconducting cell illumination results in a decrease of the electrical resistance of a semiconductor. Selenium, alloys of selenium and tellurium, and a thallous oxysulphide (in the "thalofide" cell) are semiconductors whose resistance is decreased by illumination, and have therefore been used in photoconducting cells. Selenium cells are more widely used than any others of this general type.

A photoconducting cell is always designed so as to require the current from an external circuit to flow along a very thin layer of the light-sensitive semiconducting material. If a thick layer were used, the portion that receives light would constitute but a small part of the total resistance, because of the relatively shallow penetration of light.

When light strikes a selenium cell, the resistance changes very rapidly at first, later slowly but for a long time. The amount of the change is relatively greater at low than at high temperatures, because the resistance of selenium decreases more rapidly with rise in temperature than with increase in illumination.

Experiments indicate that there are two distinct kinds of current flow in semiconductors whose resistance changes with illumination. One kind is said to be due to "primary" electrons, the other to "secondary" electrons. The total current is the sum of that due to the movements of these two groups of electrons. The primary current is directly proportional to the light intensity and appears as soon as light strikes the cell. It is presumed that the primary electrons are those lifted out of filled levels into running levels by direct action of light photons. In selenium the primary current is a small fraction of the total current. The secondary current, which is in selenium very much larger than the primary current, is presumed to be due to some indirect effect of the primary current, possibly the elimination of obstacles to the flow of all running-level electrons. The secondary current is not always directly proportional to light intensity, and a matter of minutes may be required for it to reach its final value after a change in light intensity.

225. Time-Lag in Photosensitive Devices.[FF, GG, HH]   There is no measurable time-lag between the application of a light beam and the

emission of photoelectrons from the cathode of a phototube.[HH 32]   There-
fore vacuum phototubes respond as faithfully to variations of light inten-
sity at frequencies of thousands and scores of thousands of kilocycles as
to variations whose periods are measured in minutes or hours.   However,
the interelectrode capacitance may have an appreciable effect on the
circuit response at very high frequencies.   Only vacuum phototubes
operate satisfactorily at the very high frequencies that must be used in
commercial television.

Gas-filled phototubes exhibit appreciable time-lags.   The establish-
ment of the equilibrium current for any given illumination requires the
accumulation of a positive space charge near the cathode.   This accumu-
lation takes some time because of the low mobility of the positive ions.
Because of this time lag, the current in a gas-filled phototube circuit
does not follow light variations faithfully at extremely high frequencies.

However, the " dynamic " properties of gas-filled phototubes, that
is, their response to rapidly varying light intensities, are usually satis-
factory at frequencies within the audio-frequency range (up to ten or
fifteen thousand cycles per second).   Such tubes are in fact very exten-
sively used for reproducing music and speech from sound tracks on
films.

In rectifier-type photocells there is no time-lag between application
of light and the photoelectric transfer of electrons from semiconductor
to conductor.[FF 89]   However, the geometry of such a cell makes its
electrostatic capacitance relatively large.   For high-frequency illumina-
tion the equivalent circuit must include a condenser, shown dotted in
Fig. 163a.   At sufficiently high frequencies this condenser constitutes an
internal short-circuit.

Photoconducting cells exhibit very pronounced time-lags, because of
the delay in the establishment of the final value of secondary current.

PROBLEMS

Chapter XVIII

1. A certain high-vacuum central anode photocell consists of two concentric
spherical electrodes, the outer one having ten times the radius of the inner one.
Suppose the directional distribution of the emitted photoelectron obeys what is some-
times called the cosine distribution law, expressible as follows:

$$\frac{dN_\phi}{d\phi} = N \cos \phi$$

Here ϕ is the angle an electron's direction of emission makes with a perpendicular to
the surface at the point of emission; N is the number of electrons emitted per
second, and dN_ϕ the number emitted per second having directional angles between
ϕ and $\phi + d\phi$.   Find the ratio of the photoelectric current when the electric field
inside the tube is zero to the saturation current.   (See Fig. 151.)

2. If, in Fig. 152, $s = 5$ cm, what is the value of the quantity symbolized as α' in Section 219,

(a) When $E_p = 60$ volts.

(b) When $E_p = 160$ volts.

(c) If the gas pressure is originally 2 mm of mercury, to what new value must it be changed in order to make α' at 160 volts the same as it was formerly at 60 volts?

3. (a) Explain how the presence of an inert gas in a phototube produces a tube current that is

(1) Proportional to the photoelectric emission current, and, at the same time,

(2) Several times as large as the photoelectric emission current?

(b) Why is the gas amplification reduced if the gas pressure becomes

(1) Too low?

(2) Too high?

4. Suppose that two gases have the same ionizing potential, the same probability of ionization, and the same atomic radius, but that the least excitation potential is much less for one than for the other. At moderate gas concentrations which gas will have the larger value of α'? Which will serve most effectively for gas amplification of phototube current?

5. (a) With the tube and circuit of Fig. 158, if $E_B = 100$ volts, what is the least value of load resistance that can be used without danger of initiating a glow discharge when the illumination becomes very intense?

(b) With the chosen value of load resistance, what will be the phototube current in microamperes when the illumination is 0.1 lumen? When it is 0.7 lumen?

6. Suppose that, in Fig. 159b, $E_B = 100$ volts, $R_L = 5,000,000$ ohms, and $E_1 = 10$ volts. Find the values of E_2 and I_2 that correspond to illumination of

(a) 0.1 lumen

(b) 0.7 lumen

(c) For which illumination are the harmonics in I_2 most pronounced?

7. Circuit conditions as in the previous problem, except that the load consists of 5,000,000 ohms of resistance in series with a reactance of 2,500,000 ohms. Illumination 0.7 lumens. Assume that $I_2 = 5$ microamperes, and has a sinusoidal wave form. Determine the instantaneous values of the excitation voltage,

(a) When the instantaneous value i_2 of the current I_2 is at its greatest positive value,

(b) When i_2 is at its greatest negative value,

(c) When i_2 is zero but increasing,

(d) When i_2 is zero but decreasing.

8. Conditions as in Fig. 160, with sinusoidally varying illumination. Plot a curve of I_p vertically, illumination horizontally. Your curve is qualitatively similar to the dynamic characteristic of a triode. By a method similar to that outlined in Section 149 determine the per cent second-harmonic distortion in I_2 and E_2.

9. Conditions as in Fig. 160, except that the load contains an inductive reactance of 3,200,000 ohms in series with R_L. The value of R_L is as indicated by the placement of the load line in Fig. 160. Assume *sinusoidal variations of I_p*, above and below a point Q at $E_p = 37$, lumens = 0.4. Root-mean-square value of I_2 is 3 microamperes. Locate points around the elliptical current-voltage locus, in the general manner suggested by Problem 3, Chapter XII. (Note that sinusoidal light variation will not produce sinusoidal current variation.)

CHAPTER XIX

ELECTRIC ARCS AND GLOW DISCHARGES

226. Appearance. Probably the most familiar form of electric arc is that drawn between two separating electrical contacts, as between the blade and jaw of a knife switch. Such an arc consists chiefly of a well-defined and extremely hot *core*, of small diameter, surrounded by a flame. The core terminates at extremely hot, well-localized, and small electrode areas called the *cathode spot* and the *anode spot*.

Arcs in gases at low pressure do not have well-localized, extremely hot anode spots, nor well-defined arc cores, but many of them do have small, well-localized, and very hot cathode spots. The cathode " hot spots " of mercury-pool-type mercury vapor rectifiers dodge about erratically on the mercury surface. Some experimenters have reported the occasional observance, under special conditions, of a mercury-pool-type arc without a cathode spot, electron emergence from the cathode taking place at a very moderate current density and over a large area.[90]

Many of the commercially used mercury vapor rectifying tubes have electrically heated oxide-coated cathodes that release electrons thermionically. Arcs in such devices have neither cathode nor anode spots.

A typical light pattern for a glow discharge is shown in Fig. 169b, in Section 245. The most familiar form of glow discharge is that in red neon commercial display lights. The most striking form is that which exhibits *striations*, that is, alternate bright and dark bands in the positive column.

The mechanism by which electrons are released from the cathode of a glow discharge demands an overall discharge potential of at least seventy-five volts, more often several hundred volts. Electric arcs can be maintained by much smaller voltages. Current densities in glow discharge cathode spots are small fractions of a per cent of those that exist in the very much smaller and hotter arc cathode spots. Glow discharge cathodes do not emit electrons thermionically.

227. Definite Values of Arc Current, Not of Arc Voltage, Required by Circuits. An arc between the blade and jaw of a knife switch is in reality part of the switch, for the circuit is not open until the arc breaks. The electric arcs in circuit breakers and in all varieties of gas-filled electronic rectifying devices are in reality parts of switching mechanisms. A six-phase power rectifier is essentially a high-speed switching device

that performs the same functions as the rotating commutator of a direct-current generator.

Now it is typical of any switch, and of any arc used as a switch, that the current is either zero, or has a magnitude that is determined entirely by the external circuit. The voltage across a current-carrying switch or arc-switch is just enough to compel passage of the required current. Thus external circuit conditions impose a definite arc *current*, not a definite arc potential, in an arc used as a switch.

A direct-current electric welding arc is not used as a switch, yet it is the welding current, not the voltage, that is selected by the operator to suit the work he is doing. Power supply equipment for electric arc welding is built to maintain the current constant at some selected value, the arc voltage varying according to momentary changes in arc length. Electric arcs used for street lighting operate best at substantial currents and very moderate voltages, hence are always used in series rather than in parallel. Power is supplied to them by constant-current rather than by constant-potential apparatus. Thus the circuits used for arc welding and for street lighting arcs impose definite arc currents, not definite arc potentials. The same is true in most other industrial applications of electric arcs.

In view of these facts, it is desirable to approach most problems of arc behavior by assuming current magnitudes dependent on circuit conditions external to the arcs, and working toward evaluation of the voltages required to maintain the arcs. This is almost as true with glow discharges as with arcs. It is of course not true of an extensive class of problems having to do with the *initiation* of arcs.

228. Plasmas and Plasma Boundaries. Electric arcs are ordinarily associated with the flow of large currents at low voltages, and glow discharges with the flow of small currents at relatively high voltages. However, there is much greater similarity between electric arcs and glow discharges than is usually realized; their external contrasts in appearance and electrical characteristics are so prominent as to mask the essential similarities in their internal makeup. All the contrasts between them are incidental to the existence of different mechanisms for procuring electron release from the cathode. Both can exist at either high or low gas pressures, and both assume definite normal cross sections unless restricted to smaller ones by physical boundaries.

A discharge of either type consists essentially of:

(1) A main current-carrying region, named by Langmuir the *plasma*.[91] A plasma is similar to the interior of a metal conductor in that it contains electrons and positive ions in equal concentrations, hence no space charge, and in that a considerable drift current of electrons can be

driven through it by a moderate potential gradient (see Section 167). In addition to the longitudinal electron drift current density, there are in a plasma random electron and ion current densities (see Section 127) and an outwardly directed radial ion current density.[93] The *random* electron current density is always very much greater than the electron *drift* current density. The random current densities may also be called *diffusion* current densities. The *radial ion current density*, although small, is of special importance because it results in the removal of ions from the plasma. The ion diffusion current density has a similar but often much less important effect.[92, 93]
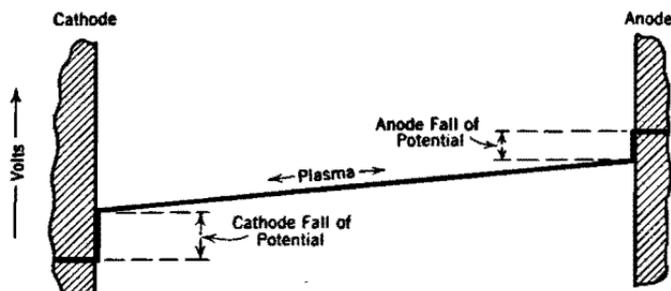
(a) Longitudinal potential distribution in an arc.
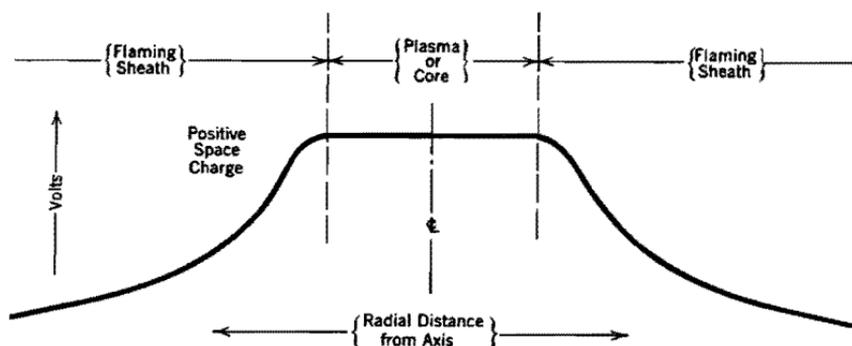
(b) Detail of transverse potential variation for an open-air arc.

Fig. 164.   Potential variations in and around an electric arc.

(2) A set of plasma boundary regions, which have properties very different from those of the plasma. For example, space charge is a prominent feature of the boundary regions. The boundary regions are classified in a succeeding paragraph, and are analyzed in Chapter XX.

The plasma of an electrical discharge usually has a different potential than the regions outside it. Figs. 164a and 164b illustrate typical longitudinal and transverse potential variations for an electric arc. The sharp curves and bends in the potential lines in these figures indicate

the existence of considerable amounts of space charge in the boundary regions. There is usually a slight radial gradient in a plasma.[93]

Plasma boundary regions may be classified; there are:

(a) The *cathode fall space*, containing the *cathode fall of potential*, located between plasma and cathode. Arcs and glow discharges differ only in respect to conditions in this one boundary region through which the main electron stream enters from the cathode. The cathode fall of potential of a glow discharge is often a matter of hundreds of volts. In contrast with this, the cathode fall of potential of an arc is approximately the ionizing potential of the active gas, so is usually less than 25 volts. Almost all problems regarding the initiation of an arc or glow, that is, arc-over, sparking, arc reignition, and electrical breakdown generally, have to do with the setting up of the cathode fall space mechanism which procures electron release from the cathode surface.

(b) The *anode fall space*, containing the *anode fall of potential*, between anode and plasma. The main electron stream leaves through this boundary region.

(c) All other boundary regions, whether between plasma and free space, or between plasma and material surfaces. In this book these regions will sometimes be called "inactive boundary regions," sometimes by the more widely used name "sheaths."[92] The current through the inactive boundary regions is either zero or small, and is incidental to the main current. The sheath around the grid of a grid-controlled gaseous rectifier (thyratron) is an example of an inactive boundary region through which there is some current flow.

229. Properties of a Plasma. The similarity between a plasma and the interior of a metal conductor has already been mentioned. A plasma contains gas particles, electrons, and positive ions, the two latter kinds of particles in equal concentrations. The ion-electron concentration is rarely more than a few per cent, and often a small fraction of a per cent, of the gas particle concentration. A drift current of electrons, driven by the electric field, accounts for the observed current flow. The ion drift current is negligible because the ions are too heavy to move rapidly.

The flow of electron drift current in response to the force due to the electric gradient results of course in the introduction of energy into a plasma. This energy appears originally as directed kinetic energy of the electrons, given to them by acceleration in the direction of the field. This originally directed kinetic energy of the electrons is immediately scattered into random directions and magnitudes, so that the energy input to the electron stream results in a high average random electron energy, which may very conveniently be interpreted as a high electron temperature. The scattering or "randomizing" of the electron energies

takes place so quickly and thoroughly that the average velocity of electron drift in the direction of the field is always very small relative to the average random velocity. [91, 92, 93, 94]

The electron temperature is ordinarily much higher than the temperature of the gas and ions, especially when the gas pressure is low. The electron temperature often exceeds 15,000 degrees Kelvin. There is nothing mysterious about the simultaneous existence of two different temperatures in the same region. It is a familiar fact that a heated substance has a higher temperature than an adjacent unheated substance, even though the two are in intimate contact. Heat energy flows from the heated to the unheated substance at a rate that is dependent on the temperature difference and the intimacy of contact. In a plasma the electrons are heated by the receipt and subsequent scattering of energy from the electric field, while the ions and gas particles are not so heated. It is true that the ions also are accelerated by the field, but power input is the product of force by drift velocity, and the ion drift velocity is negligible, so that direct input of energy from field to ions occurs at a negligible rate.

Since the electrons are heated as a result of exposure to the field, while the gas particles and ions are not, the electrons must be hotter than the gas particles and ions. Furthermore, there must be a continual transfer of energy from the electrons to the two kinds of heavier particles. In a plasma part of this energy transfer is accomplished by elastic collisions (see Section 218). Inelastic collisions, which result in radiation subsequent to excitation, and in the production of ions, account for most of the rest of the energy transfer out of the electron stream.

The energy transferred by *elastic* collisions appears as heat energy of the gas particles and ions. The electrons and gas particles are in more intimate contact, that is, have more frequent elastic collisions, at high than at low gas pressures. Therefore there is more rapid heat energy transfer from electrons to gas particles, so a smaller temperature difference between them, at high than at low gas pressures.

The electron temperature must always be high enough so that a few electrons have energies sufficient to ionize at collisions with gas particles, for the radial and random motions of the ions cause them all eventually to reach and pass through the plasma boundaries. This loss of ions must be made up by the production of fresh ions within the plasma.

Loss of ions by recombination of electrons and ions to form neutral particles is negligible in plasmas, for reasons discussed in the next section. Recombination occurs freely beyond plasma boundaries.

230. Recombination Occurs in Boundary Regions, Not in Plasmas. Practically no recombination of electrons and ions to form neutral

particles occurs in plasmas.   Recombination is discouraged by high gas-and-ion temperatures in atmospheric-pressure plasmas, and by scarcity of the right kind of encounters in low-pressure plasmas.[73, 95, V XI, LL 64, MM 20]

Analyses of the mechanisms of recombination have indicated that the probability of attachment of two approaching charged particles is very small if the relative velocity of approach is high.   Also, *large* attachment energy favors the existence of a *small* probability of attachment.[95, 73]

The average velocity is high among plasma electrons, because of their small mass and high temperature.   Therefore there are in a plasma only very few low-velocity electrons, yet only low-velocity ones are at all likely to recombine.   Furthermore, the energy of attachment (that is, the ionizing potential) for direct electron-ion joining is relatively large.   Thus the rate of direct electron-ion recombination in a plasma is negligible.

Recombination in two successive operations is in general much more likely to occur than direct electron-ion joining.   The two-stage process is as follows:

(1) An electron combines with a neutral gas particle to form a negative ion.[V 613] The energy of attachment for such an operation is small, usually only a minor fraction of an electron volt.   Therefore the likelihood of negative ion formation is much greater than the likelihood of direct electron-ion joining.   However, as in the direct process, only low-velocity electrons have an appreciable chance of forming negative ions.   Furthermore, atoms of the inert gases, neon, argon, krypton, and xenon do not form negative ions.

(2) Subsequently the negative ion " recombines " with a positive ion; that is, the negative ion and a positive ion meet and an electron is transferred from one to the other.[V 583]   Meetings two of such particles occur as a result of their random heat motions.   Once near to one another, the electrostatic attraction between the two ions makes them pursue closed orbits about a common center of gravity.   This keeps them near together until the electron transfer takes place.   Presumably the rather large energy of recombination is released, as usual, in the form of radiation.

However, this two-stage recombination process cannot be completed unless a negative ion persists long enough after the collision that produces it for a recombining encounter to occur.   Even moderately high gas temperatures interfere with two-stage recombination by destroying the negative ions almost as soon as they are formed, because the attachment energy between an electron and a gas particle is so small that they are easily parted by collisions with other moderate-energy gas particles.   There is scarcely the remotest chance of negative ions persisting long enough, in an atmospheric-pressure plasma, for the two-stage recombination process to be completed.

In the relatively cool boundary regions outside an atmospheric-pressure plasma negative ions can persist for appreciable periods.   Electrons

diffuse from the plasma to the boundary regions, and there form negative ions. Positive ions that arrive from the plasma are neutralized by contact with the negative ions.

In low-pressure plasmas the small concentrations of particles of all kinds discourage recombination. The number of attachments of any sort per cubic centimeter per second is proportional to the product of the concentrations of the two kinds of particles involved. At low gas pressures there are relatively few electrons and gas particles per cubic centimeter; negative ions form very infrequently, so that again even two-stage recombination is negligible.

In either high-pressure or low-pressure plasmas the presence of any material surface in the path of ion and electron diffusion precipitates recombination on the surface. Electrons enter or attach themselves to the surface freely, and ions easily steal them from it by contact. For this reason material-surface plasma boundaries are "ion sinks." An ion that approaches a material surface is at once removed from the plasma. This ion-removing property of material surfaces explains the effectiveness of many of the tricks used in alternating-current circuit breaker construction. It is important to make the ion concentration in circuit-breaker arcs fall off as nearly as possible in proportion to the cyclic decrease in arc current, in order to discourage arc reignition after the cyclic current zero. Removal of ions, called *de-ionization*, is favored by a large exposure of the plasma to surfaces. Forcing the arc to play in narrow slots increases the plasma surface exposure, and the introduction of oil spray exposes the plasma to oil drop surfaces. Individual oil drops do not last long, for the energy released at their surfaces by recombination heats them rapidly to vaporization.[97]

Whatever the details of the situation may be, ions and electrons that remain within a plasma rarely recombine, while those that adventure outside a plasma recombine very quickly.

231. Scattering of Electron Energies; Electron Velocity Distributions. Experiments have shown that at pressures so low that the electron mean free path is several centimeters, the electrons of gaseous discharges may be expected to have essentially Maxwellian velocity distributions.[92, 94] Collisions of electrons with gas particles are, in such discharges, so infrequent as to invalidate any attempt to account for the randomizing of electron energies by a collision mechanism. There are not enough collisions to do the job.

The originally *directed* energies of such low-pressure plasma electrons can acquire the observed Maxwellian *random* velocities only as a result of direct or indirect transfer of energy from one electron or a group of electrons to others. Those that acquire energy become the high-

velocity members of a Maxwellian distribution, those that lose it the low-velocity members.   Various mechanisms of energy transfer between electrons have been proposed.   Probably the low-pressure randomizing of electron energies can be most reasonably attributed to the effect of various types of local random variations of plasma potential.[94]   These random potential variations presumably result, in accordance with Poisson's law, from random variations in electron and ion concentration.   According to this supposition the line representing the actual potential distribution between cathode and anode has pronounced irregular variations above and below the smooth line (Fig. 164a) that represents the average condition.   The transverse potential variation is presumably similarly variable.   The plasma may be likened to a stretched tarpaulin flapping in a strong wind.   The local random variations in plasma potential are probably similar in nature and cause to the local random variations in conductor potentials which give rise to thermal agitation noises (see Section 126).

At high gas pressures, as in open-air arcs, the electron mean free path is short enough so that collisions between electrons and gas particles can conceivably account for the conversion of electron energies from directed to random form.   However, Morse, Allis, and Lamar[130] have shown that the velocity distribution of the electrons resulting from conversion purely by an elastic collision mechanism is not Maxwellian.   Furthermore, the experimental evidence for the existence of the Maxwellian distribution of velocities among the electrons of high-pressure discharges is not at all definite.   To the extent that the electrons in a discharge do not have a Maxwellian velocity distribution, the concept of electron temperature becomes difficult to work with quantitatively except in an approximate way.

The equation given by Morse, Allis, and Lamar for the distribution of random velocities in the electrons of a plasma, when scattered solely by elastic collisions, is as follows, in a notation similar to that employed in Chapter X:

$$\frac{dN_c}{dR} = \frac{4N}{\Gamma\left(\frac{3}{4}\right)} R^2 \epsilon^{-R^4} \tag{779.1}$$

The ratio R used here is defined as follows:

$$R = \frac{c}{\sqrt{\frac{2eFl_e}{\sqrt{3}\,m_e}}\sqrt{\frac{m_g}{m_e}}} \tag{779.2 esu}$$

c is total electron velocity, l_e the electron mean free path, F the electric field strength that drives the electrons, e the electronic charge, and

m_e and m_g the electron and gas particle masses respectively. In Equation (779.1), N is the electron concentration, and dN_c the number of electrons per cubic centimeter whose velocities lie between c and $c + dc$. $\Gamma\left(\frac{3}{4}\right)$ is the Gamma function of $\frac{3}{4}$, and has the numerical value 1.2254.

In a Maxwellian distribution the average energy is $\frac{3}{2} E_T$; in the distribution derived by Morse, Allis, and Lamar it is

$$\left. \begin{array}{l} \text{Average total electron} \\ \text{energy for the distri-} \\ \text{bution described by} \\ \text{Equation (779.1), in} \\ \text{electron volts} \end{array} \right\} = \frac{\Gamma\left(\frac{5}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} \sqrt{\frac{m_g}{3 m_e}} Fl_e \qquad (779.3\ \text{p})$$

$$= 0.427 \sqrt{\frac{m_g}{m_e}} Fl_e \qquad (779.4\ \text{p})$$

Note that this average is very much larger than Fl_e.

The corresponding expression for the average drift velocity of the electrons is as follows:

$$\text{Drift velocity} = \frac{1}{\Gamma\left(\frac{3}{4}\right)} \sqrt{\frac{2\pi e Fl_e}{\sqrt{3} m_e}} \sqrt{\frac{m_e}{m_g}} \qquad (779.5\ \text{esu})$$

$$= 3.78 \times 10^7 \sqrt{\frac{Fl_e}{\sqrt{m_g/m_e}}} \qquad (779.6\ \text{p})$$

For the same average energy, Morse, Allis, and Lamar's distribution gives a much smaller proportion of high-energy electrons than does the Maxwellian distribution, therefore less excitation and ionization.

Even Morse, Allis, and Lamar's distribution equation is only approximate, in that it takes no account of the loss of energy from the electron stream by excitation and ionization, also in that it assumes the electron mean free path to be independent of electron velocity. Actually the electron mean free path varies considerably with electron velocity (see Table XV).

232. Plasma Cross-Section; Equilibrium, Pinch Effect, and a Least-Energy Requirement. The plasmas of atmospheric-pressure arcs and glow discharges, also of low-pressure ones in chambers large enough to permit unlimited expansion, are observed to have more or less definite cross-sections. The cross-sectional area of a plasma grows as the current increases, the current density changing but little. The restriction to a definite cross-section can be explained reasonably well on the basis of equilibrium conditions and a least energy requirement.[93] The magnetic "pinch effect" force, that acts radially inward on any current carrier, may be partly responsible for the restriction of plasmas to limited cross-sections.

A plasma must satisfy two equilibrium requirements, as follows: (1) the rate of ion production within the plasma volume must equal the rate of ion loss through the boundary; and (2) the rate of energy input (current times voltage) must equal the rate of energy loss through the boundary surfaces. This energy loss includes that due to radiation, heat conduction, and the removal of energy of ionization in the escaping ions.

It has been pointed out that arc circuits demand definite arc *currents*, and permit whatever arc voltages the specified currents require. At any given gas concentration the two equilibrium requirements just stated can be satisfied, for any given arc current, by a plasma of almost any cross-section.

However, if the cross-section were to be very large, there would be a high rate of energy loss because of the large exposed boundary surface, and a considerable gradient would be required to introduce the energy necessary to maintain equilibrium. A very small cross-section would also result in high total energy loss, because the plasma temperatures and charged-particle concentrations would be so great as to make the energy loss *per unit boundary-area* excessive. At some intermediate cross-section the product of exposed boundary area by the energy loss per unit area has a minimum value; an arc of this cross-section requires the least energy for its maintenance, and is the one automatically assumed by the arc.

The fact that the cross-section of a high-pressure discharge is much smaller than that of a low-pressure one is an illustration of the *principle of similitude*, discussed subsequently in connection with the geometry of the cathode fall space of a glow discharge, Section 246. There it is pointed out that the mean free path in a gas is the yardstick by which a discharge lays out its own pattern; if the yardstick is small, as at high pressures, all overall dimensions are small.

Since energy and ions pass out through the boundary surface, and are produced within the plasma volume, the least-energy *shape* of plasma cross-section is that having a minimum ratio of surface to volume, that is, of circumference to area. The cross-section is, therefore, normally circular. If physical boundaries restrict the plasma to a smaller cross-section than that naturally assumed, or to a noncircular section, the gradient and overall voltage must be greater than in the absence of the restriction.[97]

233. Ion and Electron Mobilities; Drift Currents in a Plasma. The electron drift current density J_e in a plasma is related as follows to the electron concentration N_e, average electron drift velocity v_e, and electronic charge e:

$$J_e = N_e e v_e \qquad (780)$$

The following equation defines the electron *mobility* g_e in terms of average electron drift velocity and field strength:

$$v_e = g_e F \qquad (781)$$

Mobility is measured in centimeters per second per volt (or per statvolt) per centimeter. Similar equations relating to the motions of the ions are obtained by using subscripts i instead of e.

During each free path the forward component of motion of an ion or electron, of mass m, is increased by means of an acceleration Fe/m (esu) caused by the electric field of strength F. Each free path terminates in a collision. In a high-pressure plasma most of the collisions are elastic, not inelastic (see Section 218).

In some elastic collisions the forward components of the charged particle's motions are reversed, as is that of a tennis ball which makes a direct hit on a post and bounces directly backward. In some of them the forward components are unchanged, as when a tennis ball barely grazes the side of a post. In the course of many collisions, these two extreme kinds occur with equal frequency, so that the average persistence of forward motion as between them is zero. Any particular angle of partially reversing bounce can be similarly paired with the same angle of forward glancing bounce to give an average of zero forward persistence. Thus on the average the forward energy acquired by a charged particle between collisions is immediately converted into randomly directed energy, provided the particles hit remain stationary. There is, on the average, no persistence of forward motion into free paths subsequent to that of its origin.

For this reason a fairly good approximation to the average drift velocity v of a charged particle is the average of the forward velocities that would be acquired during individual free paths if there were zero forward motion at the beginning of each free path. The average velocity v during any uniform acceleration a enduring for time t after a start from rest is $\frac{1}{2}at$. Therefore if t represents the time between collisions, and if the forward velocity is zero at the beginning of each free path,

$$v = \frac{1}{2} \frac{eF}{m} t \qquad (782 \text{ esu})$$

The average time between collisions is $l/\bar{c}$, the mean free path of the charged particle divided by the charged particle's average velocity. (A car going 60 miles an hour passes a town every quarter of an hour if the towns are 15 miles apart; $15/60 = \frac{1}{4}$.) $\bar{c}$ is the average velocity regardless of direction, and is usually very much larger than the drift

velocity.   Using $l/\bar{c}$ for t

$$v = \frac{1}{2} \frac{eF}{m} \frac{l}{\bar{c}} \qquad \text{(783 esu)}$$

The corresponding expression for mobility is[V 47]

$$g = \frac{1}{2} \frac{e}{m} \frac{l}{\bar{c}} \qquad \text{(784 esu)}$$

The derivation just given has assumed all free paths to be equally long.   Of course they are not equally long; see Section 132.   A derivation that is similar to the above but takes into account the free path distribution leads to the following mobility equation:[V 47]

$$g = \frac{e}{m} \frac{l}{\bar{c}} \qquad \text{(785 esu)}$$

Equation (785) still neglects the actual forward and backward motions with which the successive free paths begin.   Furthermore, if an accelerated particle's mass is comparable with that of its target, the target is moved by the collision, and there *is* persistence of forward motion beyond the free path in which it is acquired.   If a tennis ball hits a small tree, the tree bends.   This bending reduces the velocity of backward rebound in case of a direct hit, but does not affect the continued forward motion after a grazing collision.   The two extreme types of collision result in net average forward motion after collisions.

Compton's mobility equation [73] takes into account this persistence of motion, the distribution of free paths, and the distribution of charged-particle velocities in all directions.   Compton's mobility expression is

$$g = \frac{8}{3\pi} \frac{e}{m} \frac{l}{\bar{c}} \sqrt{\frac{m + M}{M}} \qquad \text{(786 esu)}$$

Here l, $\bar{c}$, and m are the mean free path, average thermal velocity, and mass, for the charged particles, and M is the mass per particle of the gas within which the motion takes place.

For the *ions* in a plasma, $m = M = m_g$, so that

$$g_i = \frac{8\sqrt{2}}{3\pi} \frac{e}{m_g} \frac{l_g}{\bar{c}_g}, \text{ and } v_i = \frac{8\sqrt{2}}{3\pi} \frac{e}{m_g} \frac{l_g}{\bar{c}_g} F \qquad \text{(787 esu)}$$

Here l_g and $\bar{c}_g$ are the gas-and-ion mean free path and the average heat-motion velocity.   The factor $\sqrt{2}$ takes account of the persistence of forward motion.

For the electrons in a plasma, $m \ll M$, so that

$$g_e = \frac{8}{3\pi} \frac{e}{m_e} \frac{l_e}{\bar{c}_e}, \text{ and } v_e = \frac{8}{3\pi} \frac{e}{m_e} \frac{l_e}{\bar{c}_e} F \qquad \text{(788 esu)}$$

The electrons are so light that they exhibit negligible persistence of forward motion, as indicated by the fact that Equation (788) does not contain the $\sqrt{2}$ factor.

The quantities $\bar{c}_g$ and $\bar{c}_e$ in these mobility and drift velocity equations depend on the average ion and electron energies respectively, and therefore on the electric field strength. The nature of this dependence is described in the next two sections.

In the derivation of Langevin's mobility equation it is assumed that the colliding particles are perfectly elastic spheres resembling billiard balls. This assumption is only a first approximation to the truth. Furthermore, the expressions given for mean free paths in Section 132 are only approximately correct, because they too are based on the assumption of spherical gas particles. Therefore Equations (786), (787), and (788) predict ion and electron drift velocities that are only in approximate agreement with the results of experiment. Extensive discussions of mobility measurements and of more elaborate mobility equations are to be found in various scientific books and periodicals.

234. Drift Velocities of Plasma Electrons. Suppose that electrons are released into a gas at atmospheric or near-atmospheric pressure, in the presence of a moderately strong electric field. It is presumed that the gas particle random energies are small enough so that the electrons do not receive appreciable amounts of energies from the gas particles (that is, the likelihood of a gas particle's delivering energy to an electron at collision is extremely small). It is also presumed that there are no random variations in potential of the type discussed in Section 231 in connection with the randomizing of electron energies, and that collisions between electrons are very infrequent. With these restrictions in force all of any individual electron's energy comes directly from the electric field.

Under these conditions an individual electron's energy might be expected to grow in the manner diagrammed in Fig. 165. Collisions of electrons with gas particles should presumably result in (a) elastic conversion of forward motion into randomly directed motions, and in (b) average loss at collision of the small fraction of the existing energy specified by Equation (769), Section 218. The individual energy losses at collision would of course be very small. However, they should occur so very frequently, because of the high gas concentration, that an electron's *average* energy should be limited to a definite value that is often called the *terminal energy*. At the terminal energy, the average energy acquired from the field per centimeter of electron advance should be just equal to that lost to gas particles at elastic collisions, per centimeter of advance.

Morse, Allis, and Lamar's equations for velocity distribution and drift velocity, stated in Section 231, apply for the set of conditions just postulated, and Equations (779.3) and (779.4) express the terminal energy mathematically. A study of Equation (779.1) shows that there is a considerable spread of the successive values of an electron's energy above and below the terminal energy. Thus if the zig-zag line in Fig. 165 were continued indefinitely it would gradually and repeatedly rise and fall, covering altogether a rather wide range of energy values. An accidental sequence of many short free paths, terminated chiefly by "head-on" collisions, results in a gradual decline in the energy; an accidental sequence of many long free paths, terminated chiefly by glancing collisions, results in a gradual rise in the energy. Of course sequences intermediate between these two extreme types occur much more frequently than either extreme.

The factor ϵ^{-R^4} in Equation (779.1) indicates that an electron's energy is very unlikely to grow to values very much larger than the most probable, while the factor R^2 indicates a considerable spread of energy values in the range below the most probable. With sufficiently high terminal energy, the high-energy excursions may rise far enough to permit excitation or ionization.

If by means of interchange of energy among the electrons themselves, either directly or by way of

Fig. 165. An electron's average energy restricted to a "terminal" value by elastic collisions with gas particles.

local random potential variations, the distribution becomes Maxwellian, the average electron energy may be described as $\frac{3}{2} E_{eT}$ electron volts, where E_{eT} is the voltage equivalent of the electron temperature (see Sections 71 and 112). E_{gT} is the voltage equivalent of gas temperature.

The following paragraphs contain a derivation of an expression for the electron temperature, and for the corresponding electron drift velocity, based on the assumption of a Maxwellian electron velocity distribution, and on the supposition that exciting and ionizing collisions account for a negligible amount of energy transfer out of the electron stream. This derivation is much easier to follow than that leading to

Morse, Allis, and Lamar's equations, and has the merits of illustrating principles rather well, besides indicating the nature of the influence of gas particle temperature.

Let s stand for the average electron advance in the direction of the field between collisions. s will be called the electron *mean free advance.* There are $1/s$ collisions per centimeter of advance. The average loss of energy at each collision is $\frac{3}{2}E_{e^T}$ multiplied by f from Equation (769). $\frac{3}{2}E_{g^T}$ and $\frac{3}{2}E_{e^T}$ are used for E_g and E_e in that equation, so that the ratio E_g/E_e becomes E_{g^T}/E_{e^T}. The equality between the energy gained from the field per centimeter of advance and that lost by elastic collisions per centimeter of advance is expressed mathematically as follows:

$$Fe = \frac{1}{s}\left(\frac{3}{2}E_{e^T}e\right)\frac{8/3}{m_g/m_e}\left(1 - \frac{E_{g^T}}{E_{e^T}}\right) \tag{789}$$

Rearrangement leads to the following form:

$$Fs\left(\frac{m_g}{m_e}\right) = 4(E_{e^T} - E_{g^T}) \tag{790}$$

The mean free advance s is the fraction

$$s = \frac{\text{centimeters of advance per second}}{\text{total collisions per second}} = \frac{v_e}{\bar{c}_e/l_e} \tag{791}$$

The use in Equation (791) of v_e as obtained from Equation (788) leads to the following expression for s:

$$s = \frac{8}{3\pi}\frac{e}{m_e}\frac{l_e^2}{\bar{c}_e^2}F \tag{792 esu}$$

The factor $\bar{c}_e$ in Equations (788), (791), and (792) is the average velocity among Maxwellian electrons. As in any Maxwellian distribution

$$E_{e^T}e = \frac{1}{2}m_e\alpha_e^2 = kT_e \tag{793 esu}$$

Here α_e and T_e are the characteristic velocity and the temperature of the electrons, and k is Boltzmann's gas constant. $\bar{c}_e$ is related to α_e in accordance with Equation (389). Combination of Equations (793) and (389) shows that

$$m_e\bar{c}_e^2 = \frac{8}{\pi}E_{e^T}e \tag{794 esu}$$

Therefore, from Equation (792)

$$s = \frac{1}{3}\frac{Fl_e}{E_{e^T}}l_e \tag{795}$$

Combination of Equations (790) and (795) leads to the following

quadratic equation for E_{eT}:

$$E_{eT}(E_{eT} - E_gT) = \frac{1}{12}\overline{Fl_e}^2\left(\frac{m_g}{m_e}\right) \tag{796}$$

The solution for E_{eT} is

$$E_{eT} = \frac{E_gT}{2} + \sqrt{\frac{E_gT^2}{4} + \frac{\overline{Fl_e}^2(m_g/m_e)}{12}} \tag{797}$$

The general drift velocity expression for electrons, Equation (788), can be put into the following form by the use of Equation (794):

$$v_e = \frac{1}{3}\sqrt{\frac{8}{\pi}\frac{e}{m_e}}\frac{l_e}{\sqrt{E_{eT}}}F \tag{798 esu}$$

The equation for drift velocity in terms of field strength and gas particle temperature is then obtained by using Equation (797) for E_{eT} in Equation (798). A convenient form is *

$$v_e = \frac{4}{3\sqrt{\pi}}\sqrt{\frac{e}{m_e}}\frac{\sqrt{Fl_e}}{\sqrt{\dfrac{E_gT}{Fl_e} + \sqrt{\left(\dfrac{E_gT}{Fl_e}\right)^2 + \dfrac{1}{3}\left(\dfrac{m_g}{m_e}\right)}}} \tag{799 esu}$$

In practical units this is *

$$v_e = \frac{3.16 \times 10^7\sqrt{Fl_e}}{\sqrt{\dfrac{E_gT}{Fl_e} + \sqrt{\left(\dfrac{E_gT}{Fl_e}\right)^2 + \dfrac{1}{3}\left(\dfrac{m_g}{m_e}\right)}}} \tag{800 p}$$

235. Proportionality of Drift Velocities of Electrons and Ions to the Electric Field Strength or Its Square Root.

The derivation leading to Equations (799) and (800) in the previous section is based on an analysis of Maxwellian *electron* behavior. However, the principles used in the analysis are rather general, and if reasonable care is employed the results can be used to estimate the drift velocities of Maxwellian ions as well as of electrons. Expressions for drift velocities of ions must be based on Equation (787) rather than on Equation (788). The practical-unit form of the general expression for ion drift velocity that is arrived at by using Equation (787) is

$$v_i = \frac{3.16 \times 10^7}{\sqrt{\dfrac{m_g}{m_e}}}\frac{\sqrt{2Fl_g}}{\sqrt{\dfrac{E_gT}{Fl_g} + \sqrt{\left(\dfrac{E_gT}{Fl_g}\right)^2 + \dfrac{1}{3}}}} \tag{801 p}$$

This of course corresponds to Equation (800) for electron drift velocity.

* These results agree with those stated by Loeb, on page 609, Chapter XI, of his *Kinetic Theory of Gases*,[V] except that Compton's mobility equation is used here.

An important general principle regarding drift velocities can be rather inaccurately but briefly stated as follows: a charged particle's drift velocity is proportional to

(a) The *first power* of the electric field strength when the ratio

$$\frac{\text{Average forward energy of charged particles}}{\text{Average random energy of neutral gas particles}} \tag{802}$$

is very small.

(b) The *square root* of the electric field strength when that ratio is very large.

Equations (799), (800), and (801) form the basis for this principle, Fl/E_gT being considered a measure of the ratio of forward charged particle energy to random gas particle energy. The principle can be much more accurately and specifically stated as follows:

(a) The drift velocity of a charged particle is proportional to the *first power* of the electric field strength when

$$\left(\frac{E_gT}{Fl}\right)^2 \gg \frac{1}{3}\frac{M}{m} \tag{803}$$

(b) The drift velocity of a charged particle is proportional to the *square root* of the electric field strength when

$$\left(\frac{E_gT}{Fl}\right)^2 \ll \frac{1}{3}\frac{M}{m} \tag{804}$$

In these relations, as in Equation (786), m and M are the masses of the charged particles and of the gas particles respectively, and l is the charged particle mean free path. As will appear in the next section, plasma ions and plasma electrons illustrate respectively these two extreme kinds of dependence of drift velocity on field strength.

236. Drift Velocities of Plasma Ions and Electrons. The conditions governing the motions of plasma *ions* are usually such as to satisfy Equation (803), because for plasma ions $m = M = m_g$. The physical effect of the equality of mass is to cause whatever energy the ions acquire in excess of average gas particle energy to be rapidly lost to the gas particles, so that the ion temperature never rises appreciably above the gas temperature. Thus in the ionic counterpart of Equation (797) the second term in the radical is insignificant under plasma conditions.

When Equation (803) is satisfied, Equation (801) reduces to the following form, which can be used for estimating the ion drift velocity in a plasma:

$$v_i = \frac{3.16 \times 10^7}{\sqrt{m_g/m_e}}\frac{Fl_g}{E_gT} \tag{805 p}$$

The conditions governing the motions of plasma *electrons* are usually such that Equation (804) is satisfied, because for plasma electrons

$M/m = m_g/m_e$, which is very large.   When Equation (804) is satisfied, Equation (800) reduces to the following form:

$$v_e = \frac{3.16 \times 10^7}{\left(\dfrac{1}{3}\dfrac{m_g}{m_e}\right)^{\frac{1}{4}}} \sqrt{Fl_e} = 4.16 \times 10^7 \sqrt{\frac{Fl_e}{\sqrt{m_g/m_e}}} \qquad (806 \text{ p})$$

It is interesting to note that Equation (806) specifies a drift velocity that is nearly the same as that given by Morse, Allis, and Lamar's expression, Equation (779.6), the difference amounting to between 6 and 7 per cent.   A similar comparison of the average energies predicted by the two methods can be obtained by considering E_gT to be negligible in Equation (797).

The average energy is then

$$\frac{3}{2} E_eT = 0.433 \, Fl_e \sqrt{\frac{m_g}{m_e}} \qquad (806.1 \text{ p})$$

This result should be compared with Equation (779.4).   The close agreements in the *average* predictions by the two methods is in sharp contrast with the fact that *Morse, Allis, and Lamar's distribution indicates very much smaller rates of excitation and ionization than does the Maxwellian distribution,* for the same average energy.

With the exception of Equation (798), all the equations for electron drift velocity in this and the preceding section, also those for E_eT, are based on the assumption that elastic collisions account for all the transfer of energy out of the electron stream.   For plasmas in gases at or near atmospheric pressure this assumption may correspond well enough to true conditions to make Equation (806) lead to fair estimates of electron drift velocities.   It does not correspond at all well to true conditions in low-pressure plasmas that have high electron temperatures.   Equation (798) is valid whether the electrons' energies are lost by elastic or by inelastic collisions.

If exciting and ionizing collisions account for a substantial part of the energy transfer out of the electron stream, additional terms to account for such transfer must be included on the right-hand side of Equation (789).   Section 238 contains a statement of the nature of the additional terms.

Of course F is made larger by the inclusion of the additional terms; conversely, a given field strength corresponds to a smaller electron temperature when there are exciting and ionizing collisions than when there are not.   Excitation and ionization do take place in any real plasma; therefore the actual electron temperature that corresponds to any given set of values of Fl_e, E_eT, and m_g/m_e should be less, and the drift velocity

greater, than the values predicted by Equations (797) and (800) or (806).

Of course results obtained by using the expressions for drift velocities that appear in this and the preceding section are subject to the inaccuracies discussed in the last paragraph of Section 233. For example, actual electron mean free paths are dependent to a very considerable extent on electron energies, in the manner described and tabulated in Table XV.

237. Mobilities of Townsend Current Ions and Electrons. The conditions governing the drift of Townsend current *ions* (see Section 221) are more likely to satisfy Equation (804) than to satisfy Equation (803), because F and l are both relatively large in a Townsend current region, and the gas is cool. If neither extreme condition is satisfied, the ion drift velocity is expressed by Equation (801).

The mobility of an individual Townsend current *electron* varies as its energy increases along the rising front of a saw-tooth (Figs. 154 and 156, Section 218). The energy is made random *in direction* as fast as acquired, because the electrons are so light that the persistence of forward motion is negligible. However, each electron's energy has always the *magnitude* E electron volts that corresponds to height along the saw-tooth. In a Townsend current the electron energy magnitudes are not scattered into a random energy distribution.

Thus the total velocities of Townsend electrons are in general randomly directed but do not have random magnitudes. Therefore the average total electron velocity, to be used for $\bar{c}_e$ in Equation (788), is just the total velocity c that is proportional to the square root of E. Mathematically

$$\frac{1}{2} m_e c^2 = Ee, \quad \text{so that} \quad c = \sqrt{\frac{2Ee}{m_e}} \qquad (807 \text{ esu})$$

As this c is to be used for $\bar{c}_e$ in the denominator of Equation (788), the drift velocity of a Townsend current electron must vary *inversely as the square root of its height along a saw-tooth*. An overall average drift velocity can be obtained by determining the time average of the varying drift velocity along the front of a saw-tooth of average height.

238. Rate of Ion Production and of Ion Loss in a Plasma. One of the requirements for the persistence of a plasma is that the electron temperature must be high enough to produce new ions fast enough to replace those lost through the boundaries. The number of new ions produced per cubic centimeter of plasma per second is the product of two factors, which are: (1) the number of collisions with ionizing energy per cubic centimeter per second; (2) the probability p_I that ionization will occur in case of a collision with ionizing energy.

Ordinarily only a few per cent of collisions involving electrons with ionizing energy result in ionization, so that p_I is usually a moderately small fraction.[98] Under some circumstances, and in some gases, as many as 50 per cent of the collisions with ionizing energy result in ionization. In the present discussion p_I will be treated as a constant, although it is in reality some more or less complicated function of the excess of the electron's energy above the ionizing energy. Brode[98] gives a very interesting discussion of ionization and excitation probabilities. The quantity here called p_I is the ratio of Brode's quantity P_i, called by him the probability of ionization, to his probability of collision, P_c (see Table XV).

The quantity dp_c in Equation (383) is the fraction of the total number of Maxwellian gas particles or electrons that have velocities between c and $c + dc$. It is also the fractional part of any one second within which any one gas particle or electron possesses velocities between c and $c + dc$. During this fraction dp_c of one second an electron makes $dp_c \cdot c/l_e$ collisions, all with the energy corresponding to the velocity c. These are the only collisions occurring during the second with that energy, for during all other parts of the second the electron has other values of velocity and energy.

If c and dp_c are expressed in terms of R $(= c/\alpha_e)$,

$$\left. \begin{array}{l} \text{The number of collisions with} \\ \text{velocities between } c \text{ and } c + dc \\ \text{made by an electron in one} \\ \text{second} \end{array} \right\} = \frac{c \, dp_c}{l} = \frac{\alpha_e}{l_e} \frac{4}{\sqrt{\pi}} R^3 \epsilon^{-R^2} dR \quad (808)$$

To obtain the number of collisions with ionizing energy per second per electron this expression must be integrated between R_I, corresponding to ionizing energy, and infinity. (See Table VIII.) The product of the expression so obtained by p_I is the number of new ions produced per second by each electron. Mathematically:

$$p_I \frac{\alpha_e}{l_e} \frac{2}{\sqrt{\pi}} [\epsilon^{-R_I^2} (R_I^2 + 1)] = \begin{array}{c} \text{The number of new ions produced} \\ \text{per second by each electron} \end{array} \quad (809)$$

Multiplication of this quantity by N_e, the electron concentration, gives the number of new ions produced per cubic centimeter of plasma volume, per second. Of course $R_I^2 = E_I/E_{eT}$; see Group IV, Section 110.

Suppose the plasma to have a circular cross-section, of radius a centimeters. Suppose also that ions are lost only by diffusion (see Section 228). The diffusion current density of ions across the plasma boundary is just the random ion current density J_{ri} within the plasma; see Equation (454). The *number* of ions lost by this diffusion, per square

centimeter of plasma surface area, is $J_{ri}/e = N_i\alpha_i/2\sqrt{\pi}$. And of course the electron and ion concentrations N_e and N_i are the same, that is, $N_e = N_i$.

The electrons will be assumed to have a Maxwellian distribution. This assumption undoubtedly leads to too large an estimate of the rate of ion production, but permits the principles to be studied without the use of cumbersome mathematical expressions.

The surface area of one centimeter length of plasma is $2\pi a$ square centimeters, and the volume πa^2 cubic centimeters. The requirement that the rate of ion production within the plasma must equal the rate of ion loss through the boundaries is therefore expressed by the following equation:

$$\pi a^2 N_e \frac{\alpha_e}{l_e} \frac{2}{\sqrt{\pi}} [\epsilon^{-R_I^2}(R_I^2 + 1)] = 2\pi a \frac{N_e \alpha_i}{2\sqrt{\pi}} \tag{810}$$

Equation (810) reduces to the following expression for $\sqrt{E_{gT}}$:

$$\sqrt{E_{gT}} = \sqrt{\frac{m_g}{m_e}} \frac{2a}{l_e} \left[\frac{\epsilon^{-\frac{E_I}{E_{eT}}}(E_I + E_{eT})}{\sqrt{E_{eT}}} \right] \tag{811}$$

239. Energy Transfer in Low-Pressure Plasmas. It was stated in Section 236 that additional terms should be included in the right-hand side of Equation (789) if the effects of ionizing and exciting collisions are to be accounted for. These effects are of predominating importance in low-pressure plasmas that operate with high electron temperatures.

It will be convenient in deriving the forms of the additional terms to use the symbol α' to describe, as in Sections 218–220, the number of ionizing collisions per centimeter of electron advance in the direction of the electric field. It will also be convenient to use the symbol β' to describe the number of exciting collisions per centimeter of an electron's advance. A Maxwellian electron velocity distribution will be assumed, for the reasons stated early in the previous section.

The energy transferred out of the electron stream at each ionizing collision is $E_I e$ ergs, and at each exciting collision at least $E_E e$ ergs, where E_I and E_E are the ionizing and least exciting potentials of the gas particles, in statvolts. In this discussion the simplifying assumption will be made that just $E_E e$ ergs are transferred out of the electron stream at every exciting collision. This is reasonably near to the truth for gases such as neon and argon, in which there is not very much spread between E_E and E_I.

In order to take account of the effects of exciting and ionizing collisions

Equation (789) should be rewritten as follows:

$$Fe = \frac{1}{s} \frac{4(E_{eT} - E_{gT})e}{\dfrac{m_g}{m_e}} + \beta' E_E e + \alpha' E_I e \qquad \text{(812 esu)}$$

Each electron advances v_e centimeters per second, therefore produces $\alpha' v_e$ ions per second. From Equation (809), then using $v_e l_e = s\bar{c}_e$ from Equation (791),

$$\alpha' = \frac{\alpha_e p_I}{v_e l_e \sqrt{\pi}} [\epsilon^{-R_I^2}(R_I^2 + 1)] = \frac{p_I}{2s}\left[\epsilon^{-\frac{E_I}{E_{eT}}}\left(\frac{E_I}{E_{eT}} + 1\right)\right] \qquad (813)$$

The form of this expression is convenient for use in Equation (812).

The use of Equation (795) for s leads to the following expression for α':

$$\alpha' = \frac{3p_I}{2l_e} \frac{E_{eT}}{F l_e}\left[\epsilon^{-\frac{E_I}{E_{eT}}}\left(\frac{E_I}{E_{eT}} + 1\right)\right] \qquad (814)$$

This dependence of α' on field strength is considerably different from that existing in a Townsend current region.

An indication of the general nature of the effect of exciting collisions can be obtained by assuming a constant probability,[98] p_E, of a collision whose energy is greater than E_E, being inelastic. It will be assumed further that each such collision results in a transfer of $E_E e$ ergs of energy, with the exception that for collisions with energies above E_I, ionization rather than excitation results from a fraction p_I/p_E of the inelastic collisions. Actually the probability of excitation is some more or less complicated function of the energy, and the energy transferred by an inelastic collision may lie between $E_E e$ and $E_I e$ ergs.

The bracket in Equation (814) resulted from integration of Equation (808) between the limits R_I and infinity. If instead R_E (corresponding to E_E) and infinity are used, and p_E is used instead of p_I as the probability factor, the number of inelastic collisions is obtained. If there is then subtracted a term like the bracket of Equation (814), to account for ionizing collisions, an expression for the number of exciting collisions is obtained. In this way the following expression for β' is obtained:

$$\beta' = \frac{p_E}{s}\left\{\epsilon^{-\frac{E_E}{E_{eT}}}\left(\frac{E_E}{E_{eT}} + 1\right) - \frac{p_I}{p_E}\epsilon^{-\frac{E_I}{E_{eT}}}\left(\frac{E_I}{E_{eT}} + 1\right)\right\} \qquad (815)$$

This expression is in convenient form for use in Equation (812). It can of course be easily converted into a form similar to Equation (814).

The use in Equation (812) of Equations (813) and (815) for α' and β',

and of Equation (795) for s, leads to the following relation between Fl_e and E_{eT}:

$$\frac{1}{3}\overline{Fl_e}^2 = \frac{4E_{eT}(E_{eT} - E_gT)}{\dfrac{m_g}{m_e}}$$

$$+ p_E E_E\left[(E_E + E_{eT})\epsilon^{-\frac{E_E}{E_{eT}}} - \frac{p_I}{p_E}(E_I + E_{eT})\epsilon^{-\frac{E_I}{E_{eT}}}\right]$$

$$+ p_I E_I(E_I + E_{eT})\epsilon^{-\frac{E_I}{E_{eT}}} \qquad (816)$$

The three terms in the right-hand side of Equation (816) are respectively proportional to the amount of energy transferred out of the electrons stream (1) by elastic collisions, which tend to raise the temperature of the gas, (2) by exciting collisions, which result in immediate radiation of energy at specific wave lengths, and (3) by ionization of gas particles. The energy transferred by ionization is carried to the plasma boundaries by the ions, and released by recombination outside the plasma. If recombination takes place on a material boundary wall, the wall is heated by the recombination energy. If recombination takes place in the flame outside an open-air plasma, the recombination energy is radiated in specific wave lengths.

In a low-pressure plasma with a high electron temperature the first part of the middle term is likely to represent most of the energy transfer. By omitting all but this first part of the middle term, and solving the resulting expression for E_{eT}, the following approximate expression is obtained for E_{eT} in a low-gas-pressure plasma having a relatively high electron temperature:

$$E_{eT} = \frac{E_E}{\log\left[\dfrac{3E_E(E_E + E_{eT})}{\overline{Fl_e}^2}\right]} \qquad (817)$$

By using this equation for E_{eT} in Equation (798), an approximate expression for electron drift velocity in a low-pressure plasma is obtained. In working with Equation (817) numerically, a first estimate of the magnitude of E_{eT} can be made by dropping out the E_{eT} in the logarithm.

240. Energy Input to the Plasma. The energy input in ergs per second per cubic centimeter of plasma is FJ_e (esu), where J_e is the electron drift current density. In a cylindrical plasma of radius a, carrying current I, the total energy input per centimeter of plasma length is

$$FI = F\pi a^2 J_e = F\pi a^2 N_e e v_e \qquad (818 \text{ esu})$$

By the use of Equation (798) for v_e, and subsequent use of Equation (816) solved for F^2, an energy input expression can be obtained which contains one term for each of the three types of energy transfer.

Presumably the energy represented by the first of these three terms is, in an open-air arc, radiated in substantial accordance with Equation (310). In a plasma surrounded by physical boundaries this first-term heat energy may pass out of the plasma by heat conduction. The energy represented by the last of the three terms is released by recombination outside the plasma, or at its boundaries.[LL 167]

241. Static Arc and Glow Volt-Ampere Curves; Empirical Relations. Fig. 166a is intended to represent a typical set of static arc and glow volt-ampere curves, for a certain electrode geometry and material, and in a definite gas at definite values of pressure and ambient temperature.[99] As indicated in the figure, some overlapping in the voltage and current ranges of the two types of discharge is to be expected. The term " static " is used to signify that these curves apply only in direct-current or moderate-frequency alternating-current circuits.

Considerable effort has been directed toward the determination of empirical formulas for arc volt-ampere curves.[100] The following expression seems to have considerable merit:

$$E_a = A + \frac{B}{I_a{}^n} \qquad (819\,\mathrm{p})$$

Here E_a is the arc voltage, I_a the arc current, and A, B, and n empirical constants. The exponent n is found to vary directly with the boiling or sublimation temperature T_b of the *anode* material, according to the relation

$$n = 2.62\,T_b \times 10^{-4} \qquad (820)$$

The quantities A and B both vary with arc length, B more so than A. Apparently A is primarily a characteristic of the electrode materials, and B of the gas in which the arc plays. If both electrodes are hot enough to supply vapor to the arc, that from the anode rather than from the cathode tends to become active in the discharge. This behavior may be associated with the fact that the electric field tends to drive ions from the anode toward the cathode.

242. Arc and Glow Stability; Oscillating Arc Circuits. Suppose that the arc-and-glow apparatus whose volt-ampere curves appear in Fig. 166a is used in the circuit illustrated in Fig. 166b. Two typical " load lines " for this circuit have been dotted in on Fig. 166a. A, B, C, D are the intersections of one of them with the arc and glow characteristic curves. The coordinates used are the reverse of those on which

triode plate characteristics are usually drawn, so that the voltage E_B is the intercept of the load line with the vertical axis, and E_B/R_L that with the horizontal axis.

The voltage equation of this circuit is of course

$$E_B = iR_L + L\frac{di}{dt} + e_a \qquad (821\ p)$$

Here e_a symbolizes the instantaneous arc voltage, iR_L the voltage drop in the resistance, and $L\ di/dt$ that in the inductance. Suppose that the current has at a particular moment the value identified on the figure as i_1. The corresponding values of i_1R_L and of e_a are determinable

(a)   Volt-ampere diagram.   (b)   Circuit diagram.
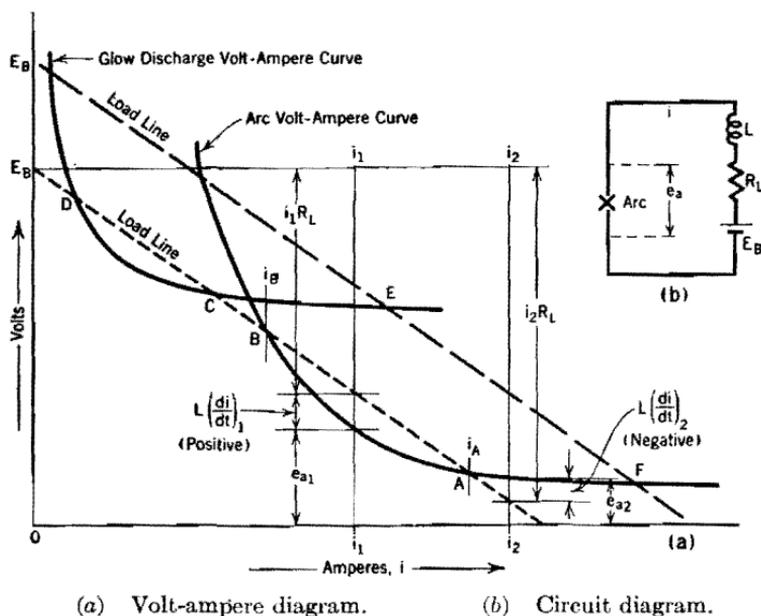
FIG. 166.   Static arc and glow discharge characteristic volt-ampere curves.

directly from the figure. Evidently the gap between i_1R_L and e_a is the value that $L\ di/dt$ must have in order to satisfy the circuit equation.

If the current has at any instant the value i_1, the quantity $L\ di/dt$ is not at that instant zero, so that the current must be *changing*. Since $L\ di/dt$ is positive, the change must be in the nature of an increase of current. The increase must continue until the current has the value i_A, for only then does $L\ di/dt$ vanish. If the initial current is i_2, the rate of change of current is negative, and the current decreases to the value i_A.

Evidently the intersection marked A describes a point of *stable operation*, for if any momentary departure from that point occurs, the circuit promptly forces the current back to it again. The effect is the same

regardless of the magnitude of L, except that if L is large, changes occur slowly.   Any real circuit has some inductance.   If a laboratory setup is made like Fig. 166b, except that the inductance is omitted, the value of L is not zero; it is only very small, so that the point of stable operation is reached very quickly.

Intersection B is not a point of stable operation.   Although $L\, di/dt$ has no value there, any slight increase in current to a value greater than i_B makes $L\, di/dt$ positive, requiring travel along the curve to A.   Similarly any slight decrease to a value less than i_B trips off a continued current increase.   Points B and D describe conditions of unstable equilibrium, resembling the state of a billiard ball balanced on the tip of a cue; A and C represent conditions of stable equilibrium.

The analysis of equilibrium requirements that has just been given indicates why load resistors are needed to maintain steady operation of direct-current arcs.   The load line of a zero-resistance circuit would be horizontal, the only intersections being like that at B which is a point of unstable equilibrium.   With alternating-current arcs ballast *reactors* are used.

The dashed load line has a stable intersection with the glow discharge curve (point E) at a rather large current value.   Operation at E is stable as far as the circuit is concerned, but is subject to abrupt, unheralded transfer to the point F, with simultaneous appearance of a hot " cathode spot " on the negative electrode.

If an arc is connected to a resonant " tank circuit," by an arrangement similar to that illustrated in Fig. 134a, sustained pulsations of arc current may occur if the arc is operated under conditions that give the volt-ampere curve a negative slope in the operating range.   The circuit analysis and equations for such an oscillating arc circuit are the same as for a dynatron oscillator; see Section 166.   The frequency of pulsation of arc current must be low enough so that the cyclic current-voltage locus lies along or near to the static characteristic curve rather than along or near to a high-frequency characteristic.   High-frequency operation is discussed in the next section.

243. Voltage and Current Variations in a High-Frequency Pulsating Arc.   During operation at any point along a static volt-ampere curve the energy input to an arc equals the energy dissipated by it.   The processes of energy intake and deliverance are associated with energy storage.   For example, recombination energy is released just as fast as ionization energy is introduced; but in the interval between ionization and recombination each ion represents the storage of E_I electron volts of energy in the plasma.   The random energies of the gas particles, electrons, and ions all represent stored energy.   The total amount of

energy stored per unit length of arc of course depends partly on the plasma cross-section.

If either the voltage or ballast resistance R_L in a circuit like that of Fig. 166b is changed, operation shifts to some new point along the static volt-ampere curve. The electron-ion concentration, the gas-and-ion temperature, the electron temperature, and the plasma cross-section may all change to new values that correspond to the requirements for energy equilibrium at the new operating point. All these changes affect the amount of energy stored in the arc. Thus a shift to a new point along the static volt-ampere curve is always associated with a change in the amount of energy that is stored in the arc.

(a) Volt-ampere diagram.   (b) Circuit diagram.
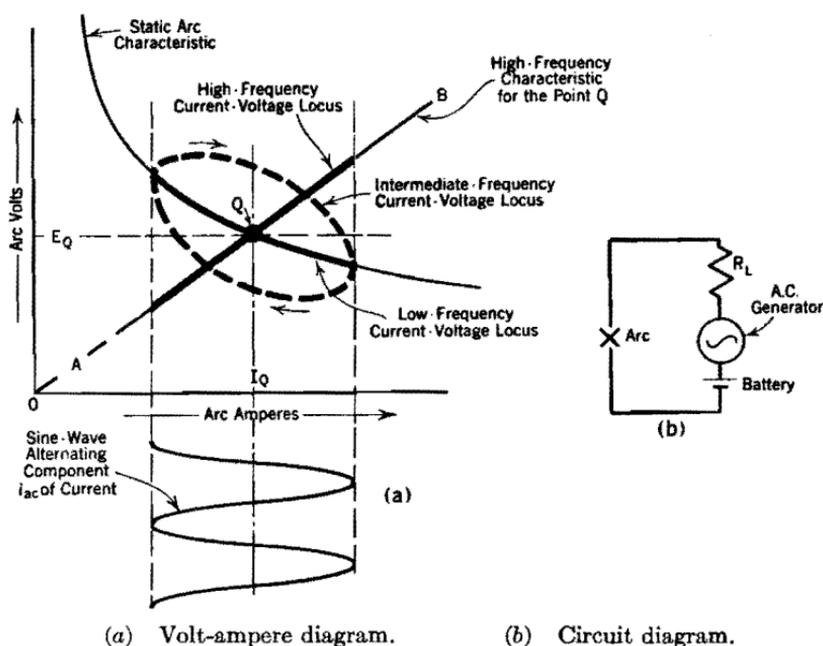
FIG. 167. Arc current-voltage loci at high, intermediate, and low frequencies.

Time is required to introduce or remove energy, so that the change from one point to another along the static curve cannot take place instantly. Static or equilibrium-condition curves like those in Fig. 166a are obtained experimentally only if the current is altered slowly enough so that readjustments to the new energy and ion balance conditions can take place as rapidly as the changes in current do.

The cyclic variations of current and voltage in 60-cycle atmospheric-pressure pulsating arcs follow the static arc characteristics reasonably closely. Those in 600-cycle arcs depart considerably from the static curves.

The plasma of an arc has been compared to a metallic conductor. The comparison is especially apt when the arc current is unidirectional but rapidly pulsating, as in Fig. 167a. I_Q in that figure is the average value of current, and i_{ac} the alternating component of current. If the current pulsates very rapidly, the electron and ion concentrations and temperatures, and the arc cross-section, remain steady at values corresponding to the I_Q, E_Q condition. Under these circumstances the plasma resistance remains constant, and the plasma gradient is proportional to the current. If the arc is long enough so that the voltage drop in the plasma is a large part of the total, the high-frequency dynamic characteristic becomes practically a straight line directed away from the origin,[101] as AB in Fig. 167a.

At an intermediate frequency the current-voltage locus pursues a loop, as suggested in the figure.[101] During the down-swing of current, the existing plasma is one originally provided for a larger current. During the up-swing, it is one originally provided for a smaller current. At any given current the plasma gradient must of course be greater for current growth than for current decay. During growth the higher gradient is building in energy toward a high-current plasma; during decay it is permitting the plasma to shrink, so that full energy maintenance is not necessary.

244. Voltage and Current Relations in Alternating-Current Arcs; Reignition. If the current in an arc is alternating, not pulsating, there is a reversal of current at the end of each half-cycle. Ordinarily the arc voltage is a small fraction of the total circuit voltage. The magnitude, wave form, and phase position of the current are controlled by the circuit external to the arc. At each moment the arc *voltage* has the value that is specified by the arc characteristic for the current flowing at that moment. Most power arcs operate at currents so large that the characteristic curve is practically flat, so
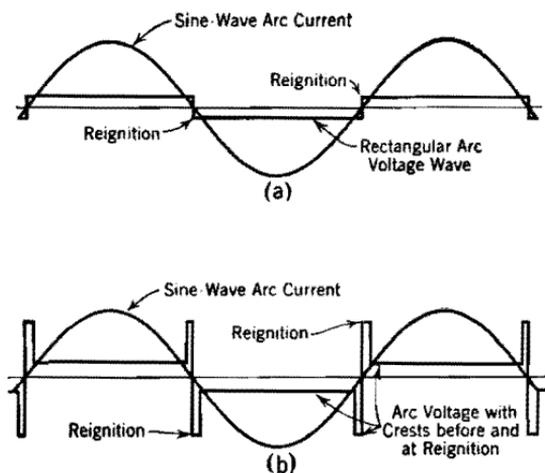
FIG. 168. Wave forms of voltage and current in alternating-current arcs.

that the voltage across the arc has a rectangular wave shape, as illustrated in Fig. 168a. Many alternating-current arcs have high-voltage

peaks of short duration at the beginning and end of each half-cycle, as illustrated in Fig. 168b.[102]

Critical examination of arc behavior immediately before and after the current in the old direction vanishes has shown that the process of reignition of the arc in the new direction is very complicated, and does not follow at all the behavior predicted from simple static characteristic curves. For example, an ordinary metallic electrode that has been serving as an anode will not develop a " cathode spot " for electron release in the new direction until a potential of at least 250, sometimes 400 or more, volts appears across the electrodes in the new direction. Westinghouse " Deion " (not deion grid) circuit breakers make use of this fact.[103]   A rise in voltage requires an appreciable time, even in the simplest circuits, so that there is always a brief zero-current period at the end of each half-cycle.

A 110-volt circuit never reaches the required reignition value, so that it is impossible to operate a 110-volt alternating-current arc between ordinary metallic electrodes.   Electrodes of very refractory materials, such as carbon and tungsten, become and remain hot enough to supply electrons freely thermionically; this behavior permits reignition at relatively low voltages after current zero.   Hence it is possible to operate arcs with such electrodes on 110-volt alternating-current circuits.

PROBLEMS

Chapter XIX

1. As explained in Section 231, local random variations in charged-particle concentrations may result in local random variations in plasma potential.   Suppose that, at a certain moment, a local region by accident comes to have the one-dimensional sinu-
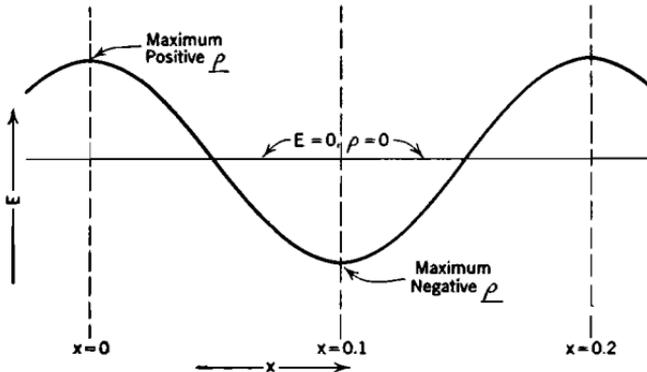
Fig. D.   Local variation of potential in a plasma.

soidal potential variation illustrated in Fig. D.   Assume that the average positive ion concentration is 10^{12} ions per cc, and that at the crest of the potential curve shown in Fig. D the electron concentration is 0.1 per cent less than this, and at the valley

0.1 per cent greater. Find the potential difference, in volts, between crest and valley.

2. Equations (799) and (800) are based on the assumption that the electron energies are scattered into a Maxwellian random distribution. Show why and how these equations must be modified to describe velocity of drift of electrons whose energy magnitudes are not scattered, each electron's energy remaining at approximately the terminal value continuously.

3. (a) In Fig. 156b, what is the slope of the rising front of each saw-tooth, in electron volts per l-unit?

(b) Show that the slope in volts per l-unit is in fact the quantity Fl_e, field strength multiplied by the electron mean free path.

(c) Calculate the average mobility of electron drift taking place under conditions as in Fig. 156b, assuming that all saw-teeth terminate at the excitation energy and begin again at zero energy. (See Section 237.)

(d) Same as (c), but for Fig. 156c. Estimate the height of an average saw-tooth, and base calculations on such a saw-tooth.

4. In a certain plasma there are 10^{17} ions per cc, and of course the same number of electrons per cc. Gas-atom concentration is 2×10^{18} per cc, gas-atom temperature, 1500° K, electron temperature 15,000° K. Ionizing potential of the gas is 10 volts; 2 per cent of the collisions with ionizing energy cause ionization. Radius of gas atoms is 3×10^{-8} cm. Atomic weight of the gas particles is 150.

(a) Find the average electron random velocity.

(b) How many collisions do all the electrons in a cubic centimeter make in one second? (See Section 238.)

(c) How many collisions with ionizing energy do all the electrons in a cubic centimeter make per second?

(d) How many new ions are produced per cubic centimeter per second?

(e) How many ions per square centimeter per second diffuse across the boundaries of this plasma?

(f) If the plasma is confined to operation within a rectangular cross-section, d centimeters wide, $2d$ centimeters long, what must d be in order to maintain equilibrium between ion production and loss of ions by diffusion through the boundaries?

5. Suppose that in a certain plasma, in order to produce ions rapidly enough to make up for loss by diffusion through the boundaries, the electron temperature must be 15,000° K. The atomic weight of the gas is 40; there are 3×10^{18} atoms and 10^{17} ions, also electrons, per cc. Gas-and-ion temperature is 2500° K. Gas-particle radius is 2×10^{-8} cm.

(a) Determine the gas pressure in millimeters of mercury.

(b) Determine the electron mean free path.

(c) Use Compton's mobility equation (see Section 233) to determine the electron mobility.

(d) Determine from the reasoning employed in Section 234 what electric field strength would be required to maintain the stated electron temperature if all the energy transfer out of the electron stream were due to elastic collisions with gas particles.

(e) Determine the corresponding electron drift velocity by using the mobility determined in (a), and check your result against those given by Equations (800) and (806). Which of the two contrasting extremes described in Section 235 does this situation approach most closely?

(f) Determine the electron drift current density.

(*g*) Determine the drift velocity of the *ions* in this plasma by the use of the proper form of Compton's mobility equation, and check the result against that obtained by using Equation (805). Which of the two extremes described in Section 235 does this situation approach most closely?

(*h*) Determine the ion drift current density.

6. Conditions as in Problem 5. The ionizing potential of the gas used is 16 volts, and ionization results from 6 per cent of the impacts of electrons with ionizing or greater energy.

(*a*) Find the number of new ions produced per cubic centimeter per second by each electron, and by all electrons. (See Section 238.)

(*b*) Find what the radius of the circular cross-section of this plasma must be if the rate of ion production is to equal the rate of ion loss.

(*c*) If this plasma is required by physical boundaries to operate with a rectangular cross-section that is *d* centimeters wide and 2*d* centimeters long, find for what value of *d* the rate of ion production will equal the rate of ion loss by diffusion.

(*d*) Find for both (*b*) and (*c*) conditions the total electron drift current through this plasma, using the value of drift current density determined in Problem 5.

7. Conditions as in Problems 5 and 6. Assume that the least excitation potential of the gas is 11 volts, and that excitation results from 12 per cent of the collisions of electrons having greater than 11 electron volts of energy.

(*a*) Determine the field strength necessary to maintain equilibrium in this plasma when all three kinds of energy transfer out of the electron stream are taken into account. This requires the determination of α' and of β'. Compare your result with that obtained in part (*d*) of Problem 5.

(*b*) Find the electron drift velocity, electron current density, and total plasma drift current, corresponding to the field strength as found in (*a*). Assume a circular plasma, radius as determined in part (*b*) of Problem 6.

8. How many watts per centimeter of length are removed from the plasma of Problem 7 by

(*a*) Outwardly diffusing ions, each carrying E_I electron volts of energy.

(*b*) Radiation of the " arc spectrum " type, as a direct result of electron excitation.

(*c*) Radiation and convection of heat resulting from the high temperature of the gas.

9. Answer questions as in Problem 5 for the following set of conditions: electron temperature 25,000° K, gas-and-ion temperature 1000° K, atom concentration 3×10^{15} per cc, ion and electron concentrations 10^{14} per cc. Gas is the same as that in Problems 5, 6, 7, and 8.

10. For the conditions stated in Problem 9, make the determinations asked for in Problem 6.

11. Carry out for the conditions of Problem 9 determinations of the kind asked for in Problem 7.

12. Determine for the conditions of Problem 11 the various rates of energy transfer out of the plasma, as classified in Problem 8.

13. The static arc characteristic for a certain gas and electrode arrangement passes through the following points (1) $I_a = 10$ amperes, $E_a = 40$ volts, (2) $I_a = 50$ amperes, $E_a = 15$ volts. Boiling temperature of the anode material is 2000° K.

(*a*) Evaluate the quantities A and B in Equation (819).

(*b*) Plot the static arc characteristic on a sheet of graph paper.

(*c*) If this arc arrangement is part of a circuit like that of Fig. 166*b*, and if $E_B =$

220 volts, find the value of R_L which will result in an arc current of 50 amperes, also that needed to produce an arc current of 10 amperes.

14. Select the constants for a circuit like that of Fig. 134a to permit use of an arc, with static characteristics as in the previous problem, for obtaining dynatron oscillations at a frequency of 300 cycles per sec. The oscillations are to be of small amplitude about a point Q for which $E_a = 30$ volts.

15. Find mathematically the low-frequency, also the high-frequency, alternating-current resistance of the arc of Problems 13 and 14 at the point for which $E_a = 30$ volts.

16. Show that the intermediate-frequency loop, Fig. 167a, must have vertical tangents at the intersections with the static characteristic.

17. Determine the most probable total velocity, and the most probable total energy, for the velocity distribution described by Equation (779.1). Plot the velocity distribution curve, and the energy distribution curve.

CHAPTER XX

PLASMA BOUNDARY REGIONS

245. The Cathode Spot and Cathode Fall Space of a Glow Discharge.
The boundary regions that surround plasmas were classified in Section
228. The active boundary regions of a glow discharge will be discussed
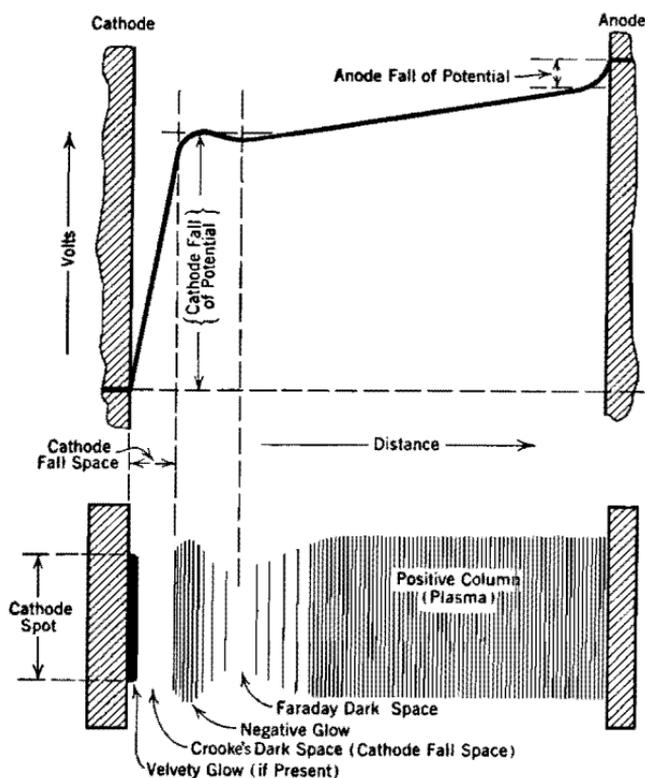
(a) Potential distribution. (Compare with Fig. 164a.)
(b) Illustrative light pattern. Positive column, negative glow, and
velvety glow are often brightly but differently colored.
FIG. 169. Glow discharge properties.

in connection with Fig. 169, which illustrates a typical glow discharge
potential variation and light pattern.

The cathode fall space, which emits little or no light, extends outward

some distance from the cathode spot area. The termination of the cathode fall space on the side toward the plasma is clearly marked by the edge of the negative glow. Beyond the negative glow there can usually be observed a less brilliant region called the Faraday dark space, which merges into the luminous positive column or plasma. The cathode spot is sometimes covered by a luminous layer that is very aptly described as the "velvety glow."

Probably the velvety glow, cathode fall space, negative glow, and Faraday dark space should all be considered as parts of the plasma boundary region through which the electron stream enters from the cathode. The properties of these bright and dark regions are described in detail in various books and technical papers.[104, LL 133, MM]

The boundary region between cathode and plasma of either arc or glow discharge accommodates the mechanism necessary for procuring emergence from the cathode surface of enough electrons to provide the current demanded by the circuit. In a glow discharge the cathode is continually bombarded by positive ions which have been accelerated by passage through the cathode fall of potential. Each of these on striking the cathode steals an electron from it and becomes a neutral gas particle. The part of the current accounted for by this mechanism is called the *positive ion current* at the cathode.

The ion bombardment of the cathode results in the ejection from it of additional electrons, in much the way that electrons by hammering the plate of a screen-grid tube produce secondary emission. The flow toward the plasma of the electrons so released constitutes the *electron current* at the cathode. The total current through the cathode fall space is the sum of the electron and positive ion currents.

The thickness of the cathode fall space (distance it extends outward from the cathode) is approximately one electron mean free path. Thus the edge of the negative glow marks the termination, with resultant excitation and ionization, of the first free flight of the electrons ejected from the cathode by bombardment. A set of equilibrium requirements and a least-energy limitation, remotely similar to those discussed in connection with the plasma, determine for any required current the normal values of cathode fall of potential and of area of cathode spot.

As in the plasma, the equilibrium conditions could be satisfied by a discharge cross-section (in this case cathode spot area) of almost any value. For one particular area, corresponding to what is called the *normal* current density, the cathode fall of potential is a minimum. This minimum is the value that exists normally, and is called the *normal* cathode fall of potential. It provides current passage with the least

energy.   In laboratory glow discharges at gas pressures of a few millimeters of mercury, cathode spots of a few square centimeters area usually pass normal currents measured in milliamperes through normal cathode fall potentials of a few hundred volts, the cathode fall spaces extending outward a few millimeters from the cathodes.

As the current through a glow discharge increases, the cathode spot enlarges, the current density and cathode fall of potential remaining at the normal values.   If the electrode area is restricted, so that the current density is compelled to be abnormally large, the cathode fall of potential also becomes abnormally large, and an *abnormal glow discharge* is said to exist.

The normal value of the cathode fall of potential of a glow discharge is determined by the nature of the gas and the material of the cathode, but is unaffected by changes in gas pressure.   Normal cathode fall voltages range between 75 and 400 volts.[LL 136]   Cathodes made of the alkali metals, sodium and potassium, permit a low cathode fall of potential with any gas, because they have small work functions.   Glow discharges in the inert gases, helium, neon, and argon, have low cathode fall potentials with any cathode material.   The neon-argon combination has a 75-volt cathode fall of potential, and the mercury vapor-iron combination is at the other extreme with a normal cathode fall of potential of about 390 volts.   These voltage values are considerably modified by the presence of small amounts of impurities, either on the surface or in the gas.   Slepian gives empirical formulas for determining the abnormal cathode fall in terms of the normal cathode fall and abnormal current densities.[LL 137]

The distinguishing feature of a glow discharge is the special mechanism it provides for releasing electrons from the cathode at the rate demanded by the circuit.   Therefore no such thing as a glow discharge with a thermionic cathode can exist, because the thermionic cathode provides its own very different electron release mechanism.   Of course if a circuit demands more than the temperature-limited thermionic current, a mechanism similar to that in a glow discharge may arise to make up the difference.

246. The Effect of Changes of Gas Concentration in a Glow Discharge; Similitude.   In electrical discharges in which ionization and the obstruction of the movements of charged particles are due to collisions with other particles, the mean free path is a yardstick by which the discharge lays out its own pattern.   The statement just made expresses briefly what is often called the " Principle of Similitude "[LL, MM, OO] for electrical discharges that are dependent entirely on collisions for the interruption of charged particle motions.

According to the principle of similitude the pattern of a discharge remains geometrically *similar* throughout wide changes in gas concentration.   Only the scale on which the pattern is laid out changes as the gas concentration changes.   A kindred principle of similitude is used in the analysis of hydrodynamic and aerodynamic *scale effects* (effects due to changes in scale).

The occurrence of Fl, volts per mean free path, rather than just F, volts per centimeter, in the various equations of Sections 218–221 and 234–239 illustrates the fact that the mean free path serves as a unit of distance in the determination of the properties of a discharge.   Similarly Equation (795) expresses the mean free advance s as a fraction of l_e, and Equation (811) contains the ratio a/l_e rather than just the distance a alone.

The cathode fall space of a low-pressure normal glow discharge provides an excellent illustration of similitude.   The nature of the gas and the cathode material determine the voltage that appears across the cathode fall space.   The external circuit can be made to maintain the current constant; the mean free path is varied by changing the gas pressure.

When the gas concentration is increased in a 1 to 3 ratio, the electron mean free path is reduced in a 3 to 1 ratio, so that the outward extent of the cathode fall space becomes one-third of what it was formerly. In accordance with similitude the cathode spot diameter also shrinks to one-third of its former value, and the cathode spot area to one-ninth of its former value.   However, the current is unchanged, so that the current density must be nine times its former value.   Thus current density varies inversely as the *square* of the electron mean free path.

In a type of laboratory demonstration that is frequently used the electron mean free path of a glow discharge is made to grow from a few tenths of a millimeter up to several centimeters, by a reduction in gas pressure.   The associated increase in extent of the cathode fall space is made evident by receding of the negative glow and Faraday dark space to greater and greater distances from the cathode.   At a sufficiently low pressure, the cathode fall space may occupy the entire distance between electrodes many centimeters apart.

The mean free path can be used as a yardstick in measuring distances along the potential distribution curve that describes potential variations outward from the surface of the cathode.   Therefore this curve must have the same *shape* regardless of the scale used in laying it out. The curve in Fig. 170 that bears the legend $E = f(x)$ illustrates a type of potential variation that might exist originally.   Now suppose the gas

concentration is tripled, so that the yardstick shrinks to one-third of its original value. The new curve is the one labeled $E = f(3x)$. It is evident that:

(a) The electric field ($-dE/dx$) at any given value of potential is three times its original value.

(b) Therefore the force that accelerates electrons and ions from or to the cathode is three times its former value; yet *the velocity of the charged particles at any given potential is the same as before*. This is obvious for the electrons, because most of them traverse the entire cathode fall space without collision, so that Equation (154) applies. It is also true of the ions, in spite of their short free paths, for although the force accelerating ions between collisions is three times what it was originally, the distance between retardations due to collisions is one-third its former value. Driving
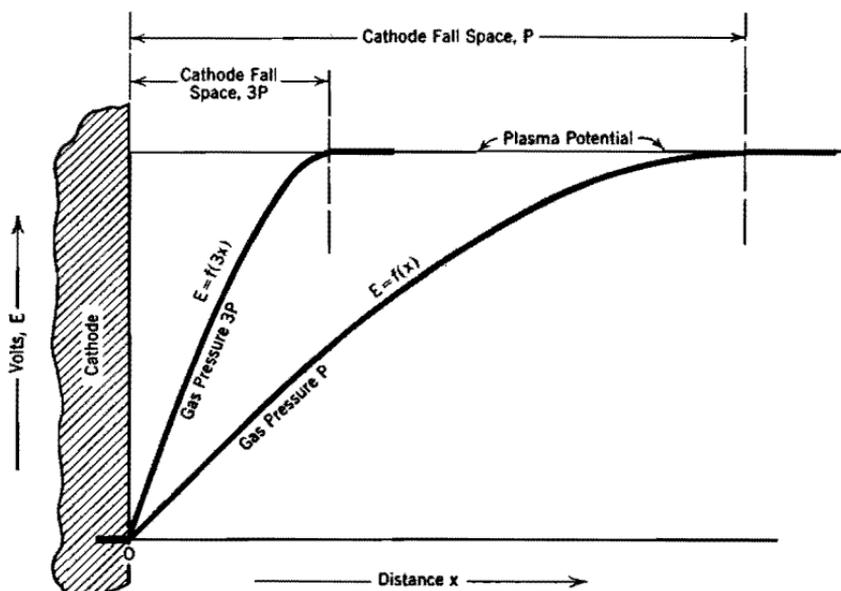
FIG. 170. Change in potential distribution within glow discharge cathode fall space as a result of increase in gas pressure, illustrating the principle of similitude.

force and obstruction to motion are both tripled so, as with the electrons, the velocity of ions at any given value of *potential* along the new curve is the same as along the old one. This is indicated mathematically by the fact that Fl_e remains constant (see Section 234). Since ion velocities are unchanged, the energy of impact at cathode surface is unchanged.

(c) The scale factor (3 in this illustration) occurs twice in obtaining the second derivative of the potential, so that the flexion, therefore the *space-charge density* (by Poisson's equation) is nine times as great in the new situation as in the old. Current density is the product of velocity by space charge density ($J = pv$) so that the current density is increased in the ratio 1 to 9. As the total current is unchanged, the cathode spot area must be reduced in the ratio 9 to 1, and its diameter in the ratio 3 to 1, the original scale factor.

It is interesting to note that while the gas particle concentration is increased by the scale factor 3, the space-charge density and ion concentration are increased by 9 times, that is, as the *square* of the scale factor. The per cent ionization is tripled. With a sufficiently large increase in pressure the increase in per cent ionization undermines the similitude relation. The nature of the discharge behavior changes as soon as a major portion of the particles present are accelerated by the force of the field; only a few gas particles then stand still waiting to be hit by moving ones.

247. Sputtering of Cathode Material. The bombardment of the cathode surface by positive ions of a glow discharge is often severe enough gradually to wear away the material of the surface. The wearing-away process is called " cathode sputtering."[105] It is uncertain whether the surface material is originally dislodged as small particles or as individual atoms. It is certain, however, that it diffuses through the interior of the tube containing the discharge as though consisting of individual atoms, that is, in the manner to be expected of a gas. After considerable time a deposit of cathode material appears over the interior walls of the tube. Surfaces in the immediate neighborhood of the cathode are more rapidly affected than remote ones.

Sputtering has certain definite practical results. It wears away the electrodes, giving them limited life. It forms an opaque deposit on the interior of the glass walls, thereby cutting down the passage of light, and increasing the chance of electric leakage along the interior walls. In the process of deposition the sputtered material takes with it some of the gas in the tube, so accelerating the process of " clean-up," or disappearance of gas from the tube.

The major problem in the early days of the construction of glow-discharge light sources was the avoidance of blackening of the tube walls as a result of sputtering.^{ZZ}

248. The Cathode Spot and Cathode Fall Space of an Arc. The recognized criterion for distinguishing an arc from a glow discharge is the magnitude of the cathode fall of potential.[99] In an arc the cathode fall of potential is approximately the ionizing potential of the active gas or vapor, hence almost invariably less than 25 volts, but in a glow discharge it is at least three times that, and often several hundred volts. In an arc, as in a glow discharge, the electron emergence from the cathode is limited to a well-defined area called the cathode spot, but cathode-spot current densities in arcs are usually very much larger than in glow discharges. An arc cathode spot one square millimeter in area may pass scores or hundreds of amperes, depending on the gas pressure.

The details of the mechanism by which electron emergence from an arc's cathode spot is obtained are not fully understood.[107, 108] The mechanism is not the same for all arcs, but is in all cases very different from that in glow discharges. Thermionic emission of the usual sort cannot be considered the prevailing mechanism, for copper, brass, aluminum, nickel, and many other metals that are used for arc electrodes vaporize at temperatures below those necessary for appreciable thermionic emission. Furthermore, an arc can exist momentarily on a practically cold surface; in the Westinghouse " Deion " circuit breakers arcs carrying many thousands of amperes are driven by magnetic fields so rapidly over copper electrodes that the electrode surfaces are scarcely marked. The heating of the cathode spot must be considered an incidental result rather than a necessary part of the electron-releasing mechanism. It is, however, true that cathode spots on refractory materials, such as carbon, tungsten, molybdenum, and platinum, can and often do become so hot that thermionic emission supplies enough electrons to provide the required current.

Whatever its details may be, the mechanism that is contained within the cathode fall space of an arc performs in general two more or less distinct functions: (1) It results in the release of electrons from the cathode, and (2) it gives the electrons ionizing energy.

If the cathode is an electrically heated electron-emitting surface in its own right, as in various types of thermionic mercury-vapor rectifiers, the electron-release part of the cathode-fall-space mechanism is taken care of independently of the action of the arc. The electron emergence is not then limited to a well-defined cathode spot, but occurs over as much of the thermionic surface as is exposed to the plasma. The emission current density is much less than the current density in a typical cathode spot. The emission from oxide-coated thermionic surfaces is materially increased by the presence of an electron-accelerating electric gradient; see Section 107. The cathode fall space of a thermionic-cathode arc provides such a gradient, so that the current density at a thermionic arc cathode can be greater than it usually is in high-vacuum devices with similar cathodes.

The second function of the cathode-fall-space mechanism, that of giving the electrons ionizing energy, must be performed whether the electrons come from an electrically heated thermionic surface or from a hot spot maintained by the arc. Therefore the cathode fall of potential has approximately the same value for either type of cathode. If a discharge contains a high percentage of metastable states, electrons can acquire ionizing energy from a cathode fall of potential amounting only to a moderate fraction of the ionizing potential of the gas. Ionization

then occurs as a two-stage process, the first step excitation to a metastable state, the second step removing the electron from the atom. Only the second of these steps need *necessarily* occur as a result of cathode-fall-space energy, for metastable atoms produced anywhere in 'the plasma may conceivably diffuse into the cathode fall space. Probably both steps usually do occur in the cathode fall space.

The distance that the cathode fall space extends outward from the cathode is so short as to defy measurement. This is true in low-pressure as well as in high-pressure arcs. This fact is surprising, for these hot cathode spots exist at gas pressures so low that the mean free path of an electron in the plasma is a matter of millimeters or even of centimeters. The probability that there can be an appreciable number of collisions of any kind within or even near the very thin arc cathode fall space seems very remote. Yet the fact that the cathode fall of potential is the ionizing potential of the active gas indicates that the cathode fall space is terminated on the plasma side, at some very short distance from the cathode, by the occurrence of many ionizing collisions. A glow discharge whose cathode fall space (approximately the electronic mean free path) is several millimeters may abruptly collapse into an arc whose cathode fall space is obviously a very small fraction of a millimeter.

Possibly the reason for the extremely small thickness of the cathode fall space is that the electronic mean free path therein is very much less than in the adjacent plasma. The plasma electric field drives ions toward the cathode spot. On arrival there each ion is neutralized and becomes a gas particle. Of course gas particles do not accumulate indefinitely. Presumably they diffuse outward because a difference between the gas particle concentration in the cathode fall space and that in the plasma is built up. This difference in concentration must ultimately become great enough to remove neutralized gas particles as fast as ions enter from the plasma. In order to accomplish this purpose the concentration in the cathode fall space may have to be greater than that in the plasma by a factor of 10 or even 100. The electron mean free path in the cathode fall space would then be shorter than that in the plasma by the same factor.

Any discussion of the mechanism of electron release from a hot-spot cathode is necessarily speculative, for it is not known definitely what the true mechanism is. The explanation of the observed high cathode-spot current densities that has received the most serious consideration is that advanced by Langmuir.[108] He suggested that the cathode fall space may be so thin that the gradient is steep enough to reduce the work function to perhaps a volt or a fraction of a volt. A large thermionic

current can then flow at a moderate electrode temperature. The reduction of work function comes about in the manner described in Section 100. The ion concentration in and thickness of the cathode fall space that are required by this hypothesis are in reasonable accord with observations and common sense.

249. The Anode Fall Space. In passing out of the plasma into the anode the main electron stream passes through the " anode fall space " within which it traverses the "anode fall of potential." Electrons are receivable at any part of the anode that is exposed to the plasma. There is no limitation to an anode spot whose area is independent of the plasma cross-section. The electrons "fall into " the anode, that is, the work function aids their entrance.

The anode fall of potential may be either positive or negative, and is usually not more than a few volts. The anode is not an ion source, and the electric field tends to drive ions away from it. The consequent scarcity of ions that exists near the anode may require the main electron stream to pass from plasma to anode as a space-charge-limited current (described in Sections 50 and 51), which demands a considerable electron-accelerating potential. On the other hand, the large random electron current density in the plasma tends toward a high electron diffusion current into the anode if the area of exposure of anode surface to plasma is very large. A small negative anode fall of potential may then be required to limit the rate of electron entrance to the anode to the current called for by the circuit.

Heat energy is released by the entrance of electrons into the anode, so that in high current density arcs the area of anode exposed to the plasma reaches a very high temperature. The amount of heat released at the anode surface is the product, in watts, of the arc current by the sum of the anode surface work function, the average temperature energy of the arriving electrons ($\frac{3}{2} E_{eT}$), and the anode fall of potential (if positive). If the anode fall of potential is negative it is merely omitted from the sum. The cooling action that must be associated with a negative anode fall of potential affects the adjacent plasma rather than the anode.

250. Sheaths ("Inactive Boundaries"). "Inactive " boundary regions envelop all the surfaces of a plasma except those through which the main electron stream enters and leaves. The inactive boundary regions are usually referred to as *sheaths,* or *positive ion sheaths.* They are called sheaths because they surround (form sheaths around) any objects, such as grids or exploring electrodes, that intrude into plasmas. They are called positive ion sheaths because they are practically free

from electrons, but contain positive ions in concentrations nearly the same as the positive ion concentrations in adjacent plasmas.[109]

Where there are no material bounding surfaces, as in an open-air arc, the inactive boundary region (flame) that surrounds the plasma may occupy a much greater volume than the plasma, and merges gradually into the surrounding atmosphere.

The extent of a plasma may be limited by its own equilibrium requirements, as in open-air arcs, or by physical boundaries, such as the glass walls of a tube, the arc chutes of an air circuit-breaker, the oil of an alternating-current·oil circuit breaker, or the grid of a thyratron.   The analysis of conditions within the sheaths that bound a plasma will be presented under the following headings:

(1) *Current-Carrying Sheaths:* a study of conditions within the type of sheath that separates a plasma from a conducting boundary surface which is electrically connected into the circuit of the discharge. The connection is usually made through a high resistance and battery. The sheath between plasma and grid of a grid-controlled gas-filled tube, and that between any plasma and an exploring electrode, or probe, correspond to this description.   The analysis accordingly follows the principles developed by Langmuir [109] and others in the course of studies of electrical discharges by means of probes.

(2) *Insulating Sheaths:* a similar study except that the boundary surface is insulated from the main electric circuit, so that no current can pass through the sheath.   The material of the boundary wall may be either insulating or conducting.   The sheath between a plasma and a " floating " grid (one having no external connections), and that between the plasma of a laboratory glow discharge and its enclosing glass tube, illustrate these conditions for conducting and nonconducting walls respectively.   This type of sheath will be treated as a simple special case of the first type.

(3) *Flaming Sheaths:* a study of the boundary region surrounding a plasma whose cross-section is determined by its own equilibrium conditions.

251. Current-Carrying Sheaths; Probes.   Suppose a plasma to be contained within a hollow metal cylinder, the cathode being at one end, the anode at the other.   The surrounding metal envelope constitutes a third electrode.   Fig. 171a is intended to represent the volt-ampere curve for such a third electrode.   The essential properties of this electrode are that (1) it is cylindrically symmetrical relative to the plasma axis, and (2) its surface can as a first approximation be treated as a plane surface, because the radius of curvature is large.

In many thyratrons (grid-controlled gaseous thermionic rectifiers, see Fig. 175, page 482) the grid structure is approximately of the nature just described.   It is essentially a *sheet* of metal in spite of the many perforations through it.   Even the part of the grid that lies across the direct cathode-to-anode path is symmetrical relative to the plasma and is essentially plane-surfaced.   Thus Fig. 171a indicates the general nature of the volt-ampere response of such a grid.

The properties of plasmas are often investigated by the introduction into discharges of small exploring electrodes, called *probes*.[92]   Probes may have various shapes; for example, fine wires, spheres, flat plates,

(a)   Volt-ampere curve for probe or grid introduced into a plasma.
(b)   Circuit diagram.
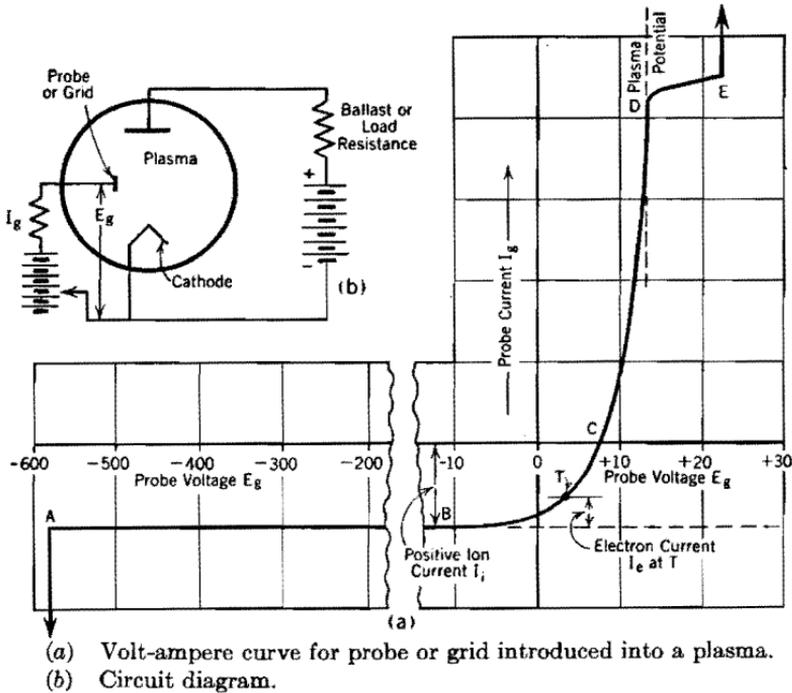
Fig. 171.   Properties of a current-carrying sheath adjacent to a probe or grid.

rings, and hollow cylinders may be used.   Fig. 171a is in general typical of the volt-ampere characteristics obtained from probes, though there may be striking deviations from certain details of the curve with probes so small that their surfaces are not even approximately plane, or with probes that are not placed symmetrically with respect to an axis of the plasma.

The circuit diagrammed in Fig. 171b can be used to obtain experimentally the probe characteristic (or grid current-grid voltage) curve, Fig. 171a.   All voltages in Fig. 171a are measured relative to the cathode.

For reasons stated farther on, the bend at D, Fig. 171a, indicates that at that point grid and plasma have the same potential. Fig. 171a is quantitatively false in that the point D should in reality be several *hundred* times as high above zero current as the horizontal portion AB is below it. The placement of D along the voltage scale is, however, approximately correct in the figure.[92]

Fig. 172a contains a typical sheath potential distribution curve for any value of grid voltage between A and D. There is no mystery about the general shape of this potential distribution curve. The experimental set-up requires that the grid be lower in potential than the plasma. The potential line between plasma and grid must therefore slope downward from one to the other. It cannot bend abruptly from horizontal to downward slope within the gas, for that would indicate a layer of surface charge in free space, which would be absurd. The potential line must bend downward gradually. However, the bend back to horizontal at the grid surface occurs abruptly, for there can be no potential gradient within the metal of the grid, and there can be a charge on its surface.

In the ensuing discussion it will be assumed that the third electrode, for which Fig. 171a contains the volt-ampere curve, is a hollow cylinder surrounding, concentric with, and bounding the plasma. It may extend for only a part of the *length* of the plasma, but the plasma does not extend outside of it. The control grids in many types of *shield grid* thyratrons (see page 499) have just this type of construction.

In general, the flow of *ions* to the plasma face of the sheath adjacent to such an enveloping electrode depends on (a) the random ion current density in the plasma (see Section 127) and on (b) the radially outward movement of ions in response to gentle radial gradients, that Tonks and Langmuir have shown to exist.[93]

If the gas pressure and temperature are high, the random ion current density predominates. If, as in most mercury vapor rectifiers (including thyratrons, ignitrons, etc.), the gas pressure and temperature are low, the important ion movement is that due to the radial field. The ions then fall freely from points of origin, under the influence of whatever potential structure exists. The random ion current density depends on the ion concentration and temperature; the radial ion flow due to the field depends on the rate of ion generation and on the radius to the inner face of the sheath. All of these quantities are determined by the *nature and extent of the plasma*. Thus the positive ion current to the plasma face of the sheath depends entirely on the properties of the plasma, and *not at all on the potential of the third electrode*.

All the *ions* that for one reason or another approach the sheath pass

on through it to the probe or grid, for the sheath potential curve is to them a hill down which they must fall. Each carries with it energy equal to its ionizing potential, which is more than enough to draw out of the grid surface, in spite of the work function, the electron needed to neutralize it. So each ion is neutralized on impact with the grid surface, and drifts away again as a neutral gas particle.

Thus it appears that the plasma face of the sheath is an "ion sink."[97] Any plasma positive ion whose motion brings it to the edge of the sheath falls to the grid and recombines there.

(a) Sheath adjacent to a   (b) Sheath adjacent to a
     conducting surface.   non-conducting surface.
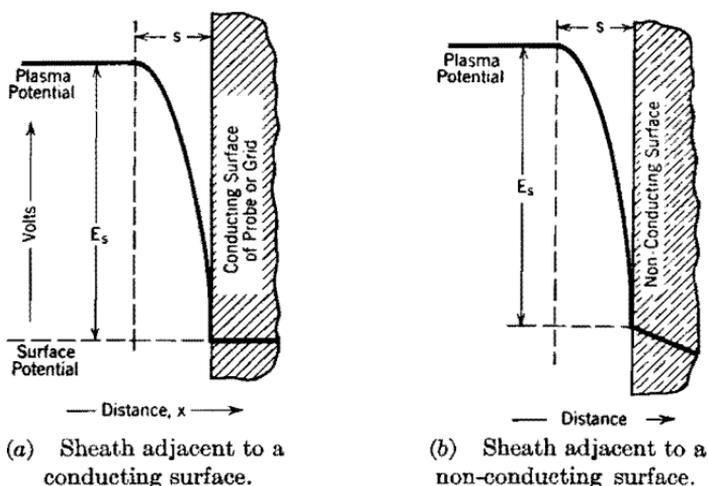
Fig. 172.   Potential distribution in positive ion sheaths adjacent to surfaces.

The sheath potential variation has of course a very different effect on the *electrons* brought by the random electron current density. These are turned back into the plasma, except that a few with exceptionally high x-directed energies may succeed in penetrating to the grid. Thus the sheath is practically free from electrons, so that the ions in the sheath produce space charge.

The velocities of the ions increase while falling through the sheath, so that the positive space-charge density produced by them must become progressively less as the probe surface is approached. Just at the plasma face the positive ion concentration must be the same as it is in the adjacent plasma; as it is less at all points nearer the grid than at the plasma face, its average value in the sheath must be slightly less than that in the plasma. This point is made in order to emphasize the fact that the sheaths are called "positive ion sheaths" *because of the absence of electrons from them*, not because of any special congregation of ions within them.

Between points A and B, Fig. 171a, the grid potential is so low that

no electrons are able to penetrate through the sheath; hence within this range of voltages the current to the grid is just that corresponding to the *ion* flow to the grid from the plasma. As this rate of ion flow is not affected by variations in the grid potential, the volt-ampere curve between A and B is horizontal.

When the grid potential rises above values in the neighborhood of B, the potential difference between the two faces of the sheath becomes small enough so that an appreciable fraction of the incoming electrons are able to penetrate to the grid. The resulting electron current flow is opposite in polarity to that due to ion flow. As the total grid current is the algebraic sum of that due to ions and that due to electrons, the growth of the electron current is indicated by a rise in the curve. At some such point as T the electron and ion currents combine algebraically to give a net grid current which may be smaller numerically than either alone. At C the two contributions are equal, so that the total grid current is zero.

At D the grid and plasma potentials are equal, so that neither electrons nor ions experience any force action in passing through the sheath; in fact, there is no sheath. Rise of grid potential above that of the plasma, as from D toward E, produces only a moderate increase in electron current. The entire random electron current density is already entering the grid, and is unaffected by the potential of the grid. Hence there is a more or less sharply defined bend in the grid volt-ampere curve at the plasma potential; this fact is used to determine experimentally what the plasma potential is at any particular point in an arc or glow discharge.

It might be expected that an *electron* sheath would form as the grid potential rises above the plasma value, just as a positive ion sheath forms when it falls below the plasma value. There is in fact a tendency toward such behavior. However, for reasons outlined in a later section, the thickness of a sheath varies approximately inversely as the square root of the mass ratio of the kind of particle that constitutes its space charge. The electron sheath should therefore be hundreds of times as thick as the positive ion sheath, for the same sheath potential difference. Thus a rise in grid potential to just a few volts above that of the plasma may make an electron sheath extend clear across the plasma, and may thereby materially alter the properties of the arc or glow as a whole.

At E the third electrode abruptly becomes a new anode of the discharge, and can then no longer properly be called a probe or grid.

The third electrode abruptly becomes a new anode when raised to a few volts above plasma potential, but it becomes a new cathode only when pushed to several hundred volts below the original cathode's

potential, as at A, Fig. 171a. The difference is attributable primarily to the fact that electrons pass freely into a metal, but are removed from it with difficulty.

The objective facts of consequence as to the volt-ampere curve as a whole are that:

(a) Between A and E, Fig. 171a, the grid is neither cathode nor anode of the arc or glow into which it is introduced, but carries a small current whose variation with changes in grid voltage depends on the properties of the sheath, so of the plasma;

(b) At A or E the grid becomes abruptly a new cathode or new anode respectively. The current to the new cathode or anode thereafter depends primarily on properties of the circuit external to the arc or glow of which it is a part.

252. Current Densities in Current-Carrying Sheaths.

The symbols listed below, having reference to the physical situations illustrated in Figs. 171 and 172, will be used in presenting the equations relating positive ion sheath properties to plasma properties. Maxwellian distributions of both ions and electrons will be assumed (see Section 231).

In the plasma:	Of electrons	Of positive ions	Of gas particles
Random current densities........	J_{re}	J_{ri}	
Concentrations ($N_e = N_i$)........	N_e	N_i	N_g
Temperatures, degrees Kelvin	T_e	T_g	
Voltage equivalents of temperature.	E_{eT}	E_{gT}	
Characteristic random velocities ..	α_e	α_g	
Mass per particle..............	m_e	m_g	

Plasma potential (relative to cathode) E_{plasma}
Probe or grid potential (relative to cathode)....... E_g
Potential difference between the two faces of the sheath, that is, between plasma and grid....... E_s
Area of sheath exposed to the plasma............. A
Thickness of the sheath........................ s
Current density and total current carried to grid by positive ions falling through the sheath ($I_i = AJ_i$)............................. J_i, I_i
Current density and total current carried to grid by electrons penetrating through the sheath ($I_e = AJ_e$).............................. J_e, I_e

The electrons and ions in the plasma are assumed to have Maxwellian random velocity distributions. Therefore, from Section 127,

$$J_{re} = \frac{N_e e \alpha_e}{2\sqrt{\pi}} \tag{822}$$

$$J_{ri} = \frac{N_i e \alpha_g}{2\sqrt{\pi}} \tag{823}$$

The easiest way of determining the α's *numerically* is to use the following relations:

$$\alpha_e = 5.95 \times 10^7 \sqrt{E_e T} \qquad (824\ p)$$

$$\alpha_g = \frac{5.95 \times 10^7 \sqrt{E_g T}}{\sqrt{m_g/m_e}} \qquad (825\ p)$$

However, the *ratio* of the α's is of primary importance. This ratio is most satisfactorily expressed by use of the following original definitions of the α's

$$\tfrac{1}{2}m_e\alpha_e{}^2 = kT_e \qquad \tfrac{1}{2}m_g\alpha_g{}^2 = kT_g \qquad (826)$$

Here k is Boltzmann's gas constant; see Section 70. By the use of these original definitions, and of the fact that $N_e = N_i$, the following expression for the ratio of the J's and α's is obtained

$$\frac{J_{re}}{J_{ri}} = \frac{\alpha_e}{\alpha_i} = \sqrt{\frac{T_e}{T_g}\frac{m_g}{m_e}} \qquad (827)$$

Thus the random electron current density in the plasma is ordinarily several hundred times as large as the random ion current density, chiefly because the mass ratio is always very large.

All of the ions, but only a fraction of the electrons, pass on through the sheath to the grid. Penetration of high-velocity electrons through the sheath to the grid takes place in accordance with Equation (458), Section 128. In the present symbolism:

$$J_e = J_{re}\epsilon^{-\frac{E_s}{E_e T}} \qquad (828)$$

The total electron current to the grid is therefore

$$I_e = AJ_{re}\epsilon^{-\frac{|E_s}{E_e T}} \qquad (829)$$

The ion current to the grid includes all the ions arriving at the plasma face of the sheath, so that

$$I_i = AJ_i \qquad (830)$$

253. Measurement of Electron Temperature. Equation (829) can be used to evaluate the electron temperature from an experimentally determined volt-ampere curve like that in Fig. 171a, in a manner outlined in the following paragraph.

The ion current is not affected by sheath potential, so that its value at some such point as T, Fig. 171a, can be determined by prolonging the line AB to the right beyond the B region. This prolongation is shown dotted in the figure. The total current and the ion current for any such point is then known, so that the electron current can be

evaluated. The logarithms of the electron currents for various such points are then plotted in the manner suggested by the following logarithmic form of Equation (829):

$$\log I_e = \log A J_{re} - \frac{1}{E_{eT}} E_s \qquad (831)$$

This equation represents a straight line, if $\log I_e$ and E_s are the variables. The slope of this straight line is $-1/E_{eT}$.

It is desirable to save work by plotting a curve of values of $\log I_e$ against corresponding values of E_g rather than of E_s. Fig. 173 illus-
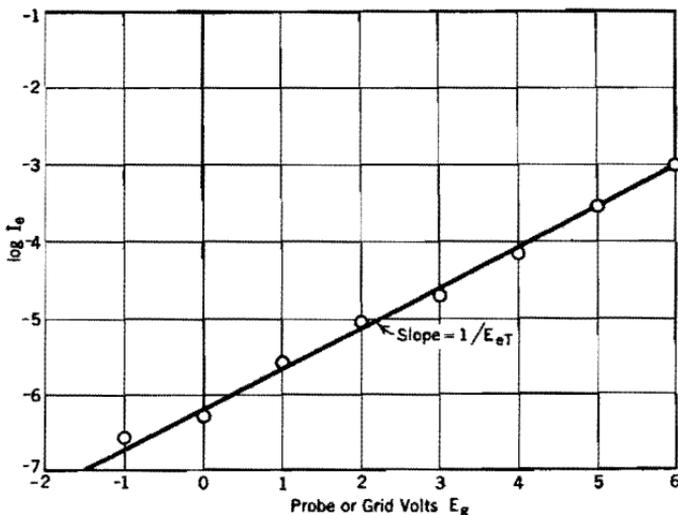
FIG. 173. Experimental determination of the electron temperature in a plasma.

trates the type of graph that results from such handling of experimental data. The points plotted lie along a line whose slope is $+1/E_{eT}$ rather than $-1/E_{eT}$. This reversal of sign is caused by the fact that

$$E_g + E_s = E_{\text{plasma}} \qquad (832)$$

Of course E_{plasma} is constant for any given discharge and probe location.

The points on graphs similar to Fig. 173 lie along a straight line only if the electrons have a Maxwellian distribution. The belief that the electrons in many low-pressure discharges do have Maxwellian distributions is based on the observation that in many cases such points do lie along straight or nearly straight lines. All measurements of this type indicate electron temperatures considerably above gas-and-ion temperatures.[110] The contrast between the two temperatures is more marked in low-pressure than in atmospheric-pressure arcs.

254. Sheath Thickness; Shut-off Grids. It is possible to predict from plasma properties what the thickness of the sheath should be,

for any given value of sheath potential difference that is not too small. The method used takes advantage of the fact that the positive ions pass through the sheath under conditions very similar to those existing in high-vacuum devices that carry space-charge-limited electron currents; see Sections 48, 49, 50, and 51.   In particular, (a) current-carriers of one sign only (ions, for the sheath) are present; (b) these carriers start with negligible velocities from a plane at which the potential gradient is zero (plasma face of the sheath), (c) they fall freely to the receiving electrode, for the sheath thickness is ordinarily a small fraction of the ionic mean free path.

It is fairly obvious that these conditions are satisfied for grid voltages between A and B, Fig. 171a, because within that range the electron current is negligible and the negative space charge in the sheath therefore also negligible.   For grid voltages between B and D the electron current has a measurable value, so that there must be some electron space charge in the sheath.   However, the electrons traverse the sheath so rapidly that the space charge due to them is negligible unless the electron current is many times the positive ion current.   For this reason the analogy between high-vacuum space-charge-limited electron flow and ion flow through the sheath is valid for grid voltages well up toward D.

Within the range of grid voltages for which this analogy is valid Equation (236) can be adapted to express the relation between the positive ion current density in the sheath, and sheath thickness and potential, for a plane-surfaced grid or probe, as follows:

$$ J_i = \frac{\sqrt{2}}{9\pi} \sqrt{\frac{e}{m_g}} \frac{E_s^{\frac{3}{2}}}{s^2} \qquad\qquad \text{(833 esu)} $$

The corresponding expression in practical units is

$$ J_i = \frac{2.336 \times 10^{-6}}{\sqrt{m_g/m_e}} \frac{E_s^{\frac{3}{2}}}{s^2} \qquad\qquad \text{(834 p)} $$

These two expressions are not equations for the determination of the ion current density.   The magnitude of the ion current density to the sheath is determined entirely by conditions within the plasma.[98] Equations (833) and (834) are in reality statements of the relation between s and E_s for given random ion current density.   It is therefore convenient to rearrange Equation (834) into the following form:

$$ s = \sqrt{\frac{2.336 \times 10^{-6}}{J_i \sqrt{m_g/m_e}}} E_s^{\frac{3}{4}} \qquad\qquad \text{(835 p)} $$

This equation shows that sheath thickness may be expected to increase nearly but not quite in proportion to an increase in sheath potential difference.

A low-pressure plasma is distinctly more luminous than the sheaths associated with it, because no electrons are present within the sheaths to produce excitation of gas particles.   The plasma face of a low-pressure sheath is therefore marked by a change in luminosity.   With suitable experimental arrangements it is possible to measure the distance from the grid surface to the plane at which this observable change in luminosity takes place.   Equation (835) has been experimentally verified by such measurements.[109]

Some grid-controlled gas-filled rectifier tubes (thyratrons) have grids made of sheet metal with many perforations.   A grid current-grid voltage curve for such a tube may show a slight but distinct slant between points similar to A and B, Fig. 171a.   In order to explain this, it will first be noted that as the grid is made more negative, the sheath potential difference increases, and the thickness of the sheath around individual grid wires also increases, in accordance with Equation (835). As the sheaths become thicker, the area they expose to the plasma becomes larger.   The grid current therefore becomes slightly larger as the grid voltage is made more negative, even though the ion current density remains constant.   Therefore the volt-ampere line is not quite horizontal in the AB range.

In general the grid in such a tube surrounds the main current-carrying part of the plasma, as well as having a section interposed between cathode and anode.   In some designs the plasma extends through the holes in the grid to the glass walls of the tube, though no appreciable drift current flows in this outer part of the plasma.   If the grid potential is made progressively negative, the sheaths around adjacent grid wires become continually thicker, until the sheaths finally completely fill the holes.   The region outside the grid is then masked from that inside, so that the plasma is restricted to the inner region, as is indicated by a pronounced and sudden reduction in the luminosity of the outer region. At the same time the grid current falls to perhaps half its former value, because only the inner faces of the sheaths, instead of both inner and outer ones, are exposed to the plasma.

Considerable effort has been directed toward the development of tubes in which grids interposed between cathodes and anodes may be used to interrupt the main current flow, instead of merely initiating it, as in present-day devices (see Chapter XXI).   Such current interruption can be accomplished by driving the grid sufficiently negative to make the grid sheath block the grid holes through which the plasma drift current

normally flows. As yet no devices containing such shut-off grids have been made that are effective enough to serve any useful purpose.[113]

255. Measurement of Ion Concentrations. The portion of the volt-ampere curve between A and B, Fig. 171a, permits determination of the ion current density J_i to the sheath. If the plasma conditions are such that the random ion current predominates in J_i, the ion concentration can be very easily calculated by the use of Equation (823), providing the gas-and-ion temperature is known. If the ions fall freely from the point of origin, a much less simple relation exists between J_i and N_i, but it has been determined for standard geometries, so that again N_i can be calculated.[93]

If a probe consisting of a very fine wire is used, not all the ions that reach the plasma face of the sheath strike the probe, because some of them pursue orbits around the probe that bring them back into the plasma. The proper employment of such probes permits direct determination of the ion concentration regardless of the nature of the predominating ion flow.[109] It is in general reasonable to expect a plasma to contain between 10^{10} and 10^{17} ions per cubic centimeter, depending on the gas pressure.

256. Insulating Sheaths. An insulating sheath is one between a plasma and a material boundary wall into which no current can flow, either because the material of the wall is nonconducting, or because, though conducting, the wall is insulated from the main arc circuit. In either case the net current through the sheath must be zero. If the boundary wall is a conductor, each ion on arrival draws out from the wall the electron needed to neutralize it, and each electron that penetrates through the sheath enters the wall. If the wall is nonconducting, each ion after arrival waits on the surface of the wall until an electron coming through the sheath happens to arrive near enough to permit recombination.

Either way, the sheath must correspond to that required for operation at the zero-current point C, Fig. 171a. This definitely determines what the potential difference between the two faces of the sheath must be, for there is only this one point on the curve at which the current due to penetrating electrons is just the same as the random ion current. At this one point $I_e = I_i$.

It is found experimentally that after a glow discharge has ceased there remains a residual negative charge on the glass walls of the enclosing tube. In order to explain this, Fig. 172a may be modified to the extent shown in Fig. 172b, to represent the potential distribution for a sheath adjacent to an insulating boundary. In Fig. 172b the potential line has a moderate slope within the wall itself. This slope might be either

positive or negative, depending on the relative potentials of plasma and surrounding objects.   It will, however, always be much less than the slope of the potential line at the surface of the wall, so that there is always a sharp bend, and a corresponding negative charge, at the surface of the wall.   Because the wall is an insulator, this surface charge is retained for some time after the plasma has ceased to exist.

257. Flaming Sheaths around Unrestricted Plasmas. Perhaps the most important difference between current-carrying and insulating sheaths on one hand and the flaming sheaths that surround open-air arcs on the other hand (see Fig. 164b) is in the manner and place of recombination of ions.

Any plasma ion that enters any of these sheaths is lost from the plasma; it is bound to recombine.   In a sheath adjacent to a wall, the recombination takes place at the wall.   In flaming sheaths the lost ions continue their outward diffusion for varying distances before recombining. Such sheaths do not have sharply defined outer boundaries; they merge gradually into the surrounding atmosphere.   Probably the two-stage mechanism described in Section 230 accounts for most of the recombination that takes place in flaming sheaths.

The electrons that take part in the recombination, like the positive ions, diffuse outward from the plasma into the flaming sheath.   This diffusion takes place in spite of the force due to the potential gradient in the sheath.   The electrons are able to move against the force due to the electric field because of their very considerable temperature energies. Because of the two-stage nature of the recombination process, flaming sheaths may contain negative ions in appreciable concentrations.

The true potential distribution curve for a flaming sheath must merge gradually into the space-charge-free potential distribution pattern established by the circuit as a whole for points remote from the arc. Both positive and negative space charge may exist in a flaming sheath, the latter enveloping the former.   There must always be a hollow cylinder of positive space charge immediately surrounding the plasma. If the arc occupies a low-potential position in the circuit, there may have to be a hollow cylinder of negative space charge around the positive space charge region.   The necessity for the outer envelope of negative space charge is lessened by the cylindrical symmetry that most arcs exhibit, because in cylindrical geometry a potential distribution line can have upward flexion in regions that do not contain space charge (see Section 9).

PROBLEMS

Chapter XX

1. Immediately after a cyclic current zero, in an alternating-current arc (see Fig. 168), there is a brief period during which the region between the electrodes is essentially a plasma in which no current is flowing. The voltage gradient in the plasma is zero, and whatever potential difference exists between the electrodes appears between the two sides of a positive ion sheath, similar to that represented by Fig. 172a, adjacent to an electrode. The electrode that bounds one face of the sheath is about to become the new cathode.

(a) If the positive ion concentration within the sheath is uniform at a value 10^{16} ions per cc, how far from the electrode must the sheath extend in order to support a potential difference of 700 volts?

(b) What is the corresponding gradient adjacent to the electrode surface?

2. In Fig. 169 the potential gradient is negative between the middle of the negative glow and the middle of the Faraday dark space. What mechanism can account for the flow of electrons from the negative glow to the Faraday dark space in spite of the negative gradient? What does your answer imply as to the relative electron concentrations in the negative glow and the Faraday dark space?

3. A gaseous electrical discharge exists in sodium vapor, the sodium vapor temperature being 806° K. The electron temperature is 29,000° K. There are 10^{12} positive ions per cc. Assume that *random* ion flow to sheaths predominates.

(a) Find the current density to a boundary whose potential is so low relative to the plasma that a negligible number of electrons reach it.

(b) Find the current density to a boundary whose potential is 10 volts below that of the plasma.

(c) Find the potential difference between the plasma and an insulated boundary.

4. Tests on a certain mercury-vapor thyratron show that as the grid voltage is decreased the grid current reaches a constant value of about 114 milliamperes (direction of current flow from cathode to grid inside the tube) at points below about −12 volts relative to cathode, also that the grid current is zero when grid voltage is −2 volts. Assume gas-and-ion temperature to be 800° K, and a grid area of 20 sq cm to be exposed to the plasma. Plasma potential is 16 volts above that of the cathode. Find:

(a) The electron temperature.

(b) The thickness of the sheath adjacent to the grid when the grid voltage is −100.

5. To what electron temperature does the straight line in Fig. 173 correspond?

CHAPTER XXI

MERCURY-VAPOR RECTIFIERS

258. Mercury-Vapor Rectifiers Essentially High-Speed Switching Devices. The name *rectifier* may properly be applied to any electronic device that can pass current freely in the " forward " direction, but does not permit current to flow in the " inverse " direction. Rectifying devices having low pressure electric arcs in mercury vapor as the essential current-carrying elements are used for a great variety of engineering services.[PP, QQ, RR, TT, WW, R]

Ratings of mercury-vapor rectifying devices range from a few milliamperes at a hundred or so volts to a few thousand amperes at a few thousand volts. The basic principles of operation are the same regardless of the size, type of cathode, or purpose for which used. All of them, regardless of the names they bear or the specific uses to which they are put, *are essentially high-speed switching devices*. The simplest forms provide closed circuits for forward current and open circuits for inverse current.

It is important to keep constantly in mind the sharp contrast between the current-" throttling " action of high vacuum devices and the simple switching action of gas and vapor rectifiers. A throttle controls rate of flow; a switch either denies passage or permits flow at an unlimited rate. A mercury-vapor rectifier has no more control over the *amount* of instantaneous current flow through it than a knife switch has.

The circuit through a mercury-vapor rectifier is closed when the arc exists, open when it does not exist. The closing and opening can be controlled by creating conditions that permit or forbid the existence of the arc. The voltage drop between anode and cathode makes a mercury-vapor arc an imperfect switch, in that it uses some power, which (1) might otherwise be more usefully employed, and (2) demands that the device itself be able to dissipate power. In addition to having this technical imperfection, mercury-vapor rectifiers are more expensive than simple switches, and often require auxiliary equipment which costs money, occupies space, and requires more or less attention. This is true of radio-receiver rectifier tubes with their auxiliary filament transformers, as well as of 1,000-ampere railway-service steel-tank rectifiers with their vacuum pumps and arc-initiating and keep-alive rigs.

However, these various mercury-arc switches can open and close cir-

cuits repeatedly and with extreme rapidity, and the time and rate of opening and closing can, if desirable, be made conveniently subject to the will of the operator. The cost and complications of mercury-vapor rectifiers are for many purposes more than compensated for by the extremely rapid, quiet, and easily controlled circuit opening and closing that they can accomplish.

259. Classifications as to Types and Uses. Mercury-vapor rectifiers can be classified as to the type of cathode employed. Thus there are:

(A) Mercury-pool rectifiers. In a rectifier of this type the electrons enter the arc through a hot wandering cathode spot on the surface of a pool of mercury. The pool must contain enough mercury to permit continuous evaporation at the cathode spot. The evaporated mercury returns to the pool by gravity after condensation on the walls. The cathode spot may be initiated:

(a) By a mechanical arc-initiating rig (often a tilting arrangement) which breaks the electrical contact between the pool and an auxiliary electrode or well of mercury while current is flowing. An arc is drawn as the contact is broken. After initiation in this manner the cathode spot is maintained by having the arc play in turn to each of two or more anodes during their respective positive half-cycles. Only by having a common pool serve two or more anodes can the cathode spot be kept alive in an alternating-current circuit.

(b) By an igniter rod of carborundum or similar material that extends from above to a little below the mercury surface, in the manner illustrated in Fig. 174. A momentary auxiliary rod current (usually a few amperes) sufficient to establish a rod gradient in the neighborhood of 100 to 150 volts per centimeter will initiate a hot cathode spot at the rod-mercury junction, in the manner described later in Section 272. Rectifiers that employ igniter-rod initiation are called *ignitrons* or *ignitron rectifiers*. In ignitron rectifiers each anode may have its own cathode pool in its own separate chamber, because arc initiation may take place anew at the beginning of each positive half-cycle of arc. It is not necessary to keep a cathode spot continuously alive.[114]

Igniter-type rectifiers were first developed by the Westinghouse Electric and Manufacturing Company, and the word ignitron was originally a Westinghouse trade name. It has, however, come into general use as a term descriptive of igniter-type rectifiers regardless of manufacture, and the Westinghouse Company has recently reached a decision to relinquish trade-mark rights to the word for the sake of the convenience of the general public.

There are also:

(B) Thermionic mercury-vapor rectifiers, in which the electrons enter through the oxide-coated surfaces of electrically heated cathodes. In the larger devices of this type, included in the category of industrial electronic tubes, cathodes are used that approach more or less closely to the inward-radiating ideal described in Section 78. Their electron-emitting efficiencies are therefore very much higher than for the cathodes

in high-vacuum devices. The advantage inherent in high emission efficiency is somewhat offset by the inconvenience of having to wait for from one to several minutes after turning on cathode heating power before drawing anode current.[111]

Evaporation from droplets of mercury, introduced during manufacture, supplies the necessary mercury vapor in thermionic mercury-vapor rectifiers.
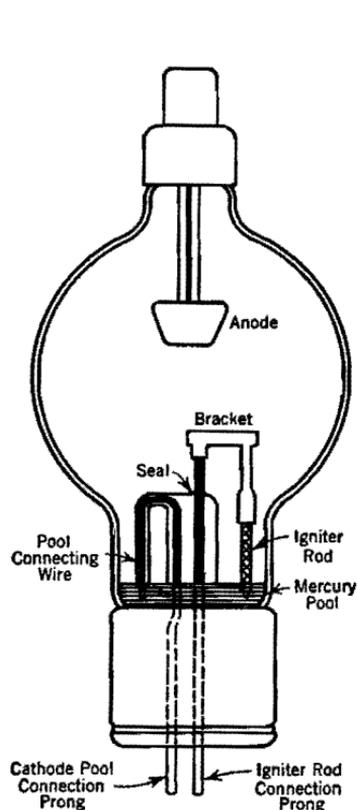
Fig. 174. Electrode arrangement in a Westinghouse Type KU-637 ignitron.
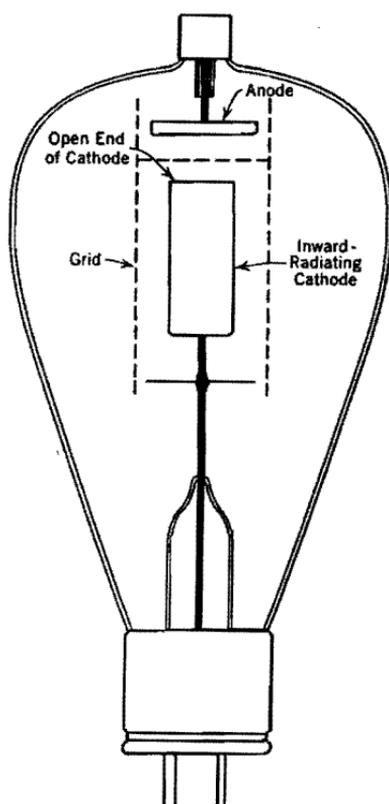
Fig. 175. Typical electrode arrangement in a thyratron.

Mercury-vapor rectifiers can be classified according to the manner in which arc initiation can be controlled. The operator may be enabled to compel or forbid *initiation* of an arc by providing:

(1) Grids interposed between cathodes and anodes, in either thermionic or mercury-pool-type apparatus. Fig. 175 illustrates a typical arrangement of parts in a grid-controlled mercury-vapor rectifier, usually called a *thyratron*.[111] In order for thyratron grids to function, the cathodes must always be ready to liberate electrons freely. Thermionic cathodes obviously meet this requirement. In mercury-pool

apparatus equipped with grids it is met by having auxiliary "keep-alive" arcs, operating to auxiliary anodes, maintain hot cathode spots at all times, whether anode current is flowing or not. With either type of cathode a grid may serve to compel or forbid, at the will of the operator, entrance of electrons into the accelerating field between grid and anode. Once electrons enter this field, the arc is immediately initiated and the anode circuit closed; the grid cannot reopen it again. Grid control of arc initiation is discussed in detail in Sections 268 and 269.

Grid-controlled gaseous rectifiers were first developed by the General Electric Company, and the word thyratron was originally a General Electric trade name. It has, however, come into general use as a term descriptive of grid-controlled gaseous rectifiers, particularly those with thermionic cathodes, and the General Electric Company has recently reached a decision to relinquish trade-mark rights to the name for the sake of the convenience of the general public.

(2) Igniter rods connected to suitable auxiliary igniting circuits, in ignitron rectifiers. Arc initiation by means of igniter rods is discussed in detail in Sections 272 and 273.[114]

Mercury-vapor rectifiers of these various types are used for:

(a) Power rectification, that is, conversion of electric power from alternating to direct current.

(b) Switching and control of alternating-current machinery.

(c) Providing finely graded, efficient control of average power flow by passing current during controllable fractions of each alternating-current half-cycle; rectification may or may not be simultaneously provided.

(d) Permitting control of the flow of substantial amounts of power by small-energy impulses, as from light- or sound-sensitive devices, in either direct or alternating-current circuits.

(e) Miscellaneous "trick" services, as for counting impulses and for providing intermittent condenser charge and discharge in stroboscopic and cathode-ray sweep circuits.

(f) Inversion of electric power from direct to alternating current.

Applications employing single-phase alternating current, either in the input or output or both, ordinarily provide for "full-wave" operation, that is, operation during both halves of each alternating-current cycle. For full-wave single-phase *rectifying*, one cathode and two anodes in a common glass or metal envelope may be used. Apparatus for full-wave *switching*, or for current control without rectification, must have two separate envelopes each with its own cathode and anode. In either case current passes through one arc in positive half-cycles, and through the other arc in negative half-cycles.

260. Commercial Classification. Probably the most familiar classification of mercury-vapor rectifiers is a semicommercial one, which distinguishes in a general way between the fields of engineering work in which the various types are used. Thus there are:[112, R, QQ, TT, WW]

I. Large and expensive polyphase steel tank "mercury-arc rectifiers," rated in scores or hundreds of kilowatts, each one requiring vacuum-

pumping and other auxiliary apparatus. These are used for recti-
fying large amounts of power, as for example in substations that supply
direct current to the trolley wires of electric railways. They are invari-
ably mercury-pool devices, and may or may not have either control
grids or igniter rods. They are usually six-phase rectifiers, operated
from three-phase power supply circuits. In the type of six-phase rectifier
most commonly used there are no igniter rods, and six anodes operate
from a common mercury-pool cathode as illustrated in Fig. 176. The
arc transfers from one anode to another, then to the next, and so on,
but has its cathode always on the one cathode pool.
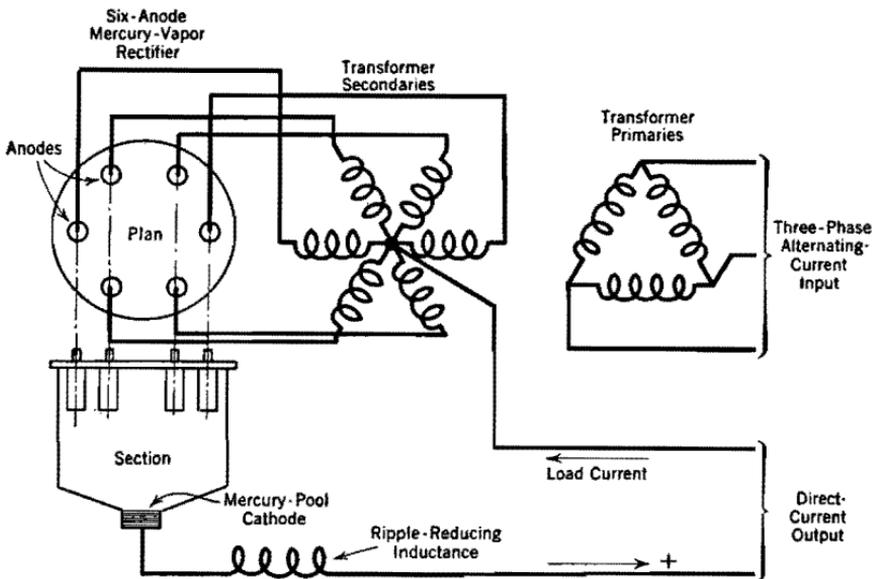
Fig. 176. Circuit diagram of six-phase mercury-arc rectifier and transformers.

Rectifiers of this type are used only for rectification (possibly in the
future for inversion) and not for switching or control services. They
compete with synchronous converters, and the advantages are by no
means all with either kind of apparatus. The most serious problem in
the development of satisfactory mercury-arc rectifiers has been the
avoidance of *arc-back*, that is, internal short-circuiting as a result of an
arc striking between anodes.

II. Devices, usually described as "industrial electronic tubes," which
are used for a great variety of rectifying, switching, and current-control
purposes in industrial plants and power stations. Industrial electronic
tubes are manufactured by the large electric manufacturing companies,
and are available in ratings from a few tenths to a score or more amperes,

and may have either glass or metal envelopes. The following types are of special interest:

(1) Thermionic mercury-vapor rectifiers without either grids or igniter rods.

(2) Mercury-vapor thyratrons, that is, grid-controlled thermionic mercury-vapor rectifiers. The name thyratron is also applied to similar special-purpose low-rating tubes which employ argon instead of mercury vapor, in order to avoid variations of operating characteristics with changes in temperature. Fig. 175 illustrates a typical arrangement of parts in a thyratron.

(3) Ignitrons, that is, igniter-controlled mercury-pool type rectifiers, illustrated in Fig. 174.

(4) Older mercury-pool type glass-tube rectifiers, of moderate ratings, having two or more " anode arms " and tilting arrangements for initiating the arc. Devices of this type have been made with and without grids; when made with grids they have been called thyratrons. Devices of this type have been largely superseded by equipment of more modern design.

A considerable part of the present-day interest in electronic developments is centered around the possibilities for use of thyratrons and ignitrons, Classes (2) and (3) above.[112]

Thermionic thyratrons, Class (2) above, make use of by far the most flexible combination of type of cathode and method of control. It is not, however, an especially cheap or durable combination. Thermionic cathodes are expensive, require continuous power input for heating, are easily damaged by overloads, short circuits, and continued vibration, and have limited life at best. Mercury-pool cathodes are not expensive, have large reserve capacities, and are mechanically rugged and long-lived. It is nevertheless more satisfactory to use grids in thermionic than in mercury-pool-type devices, because auxiliary filament-heating circuits are easier to handle, more familiar to users, and consume less power than the auxiliary keep-alive arc circuits that must be used in connection with grid-controlled mercury-pool apparatus. All modern thyratrons have thermionic cathodes, although in polyphase power conversion apparatus grids are sometimes used with mercury-pool cathodes.

Industrial electronic tubes of the ignitron type, Class (3) above, have very distinct advantages in regard to cathode ruggedness. However, igniter-rod control incorporates the practical weakness of momentarily demanding considerable power from the igniting circuit at each ignition, whereas grids can be actuated from familiar types of low-energy auxiliary circuits.

In general, thyratrons are very convenient and flexible, but not as cheap and rugged as might be desired. Ignitrons are rugged and economical, especially in the larger ratings, but lack the flexibility of control possessed by grid-actuated devices.

In some applications where substantial amounts of current are to be

controlled, ignitrons carry the main current, but their ignition is controlled by thyratrons. Such an arrangement makes use of the flexibility of grid control and of the economy and reliability of the mercury-pool cathode construction. It also reduces the overload and short-circuit hazard on the thermionic cathodes by placing them in auxiliary rather than in the main circuits. As the thyratrons may have small average ratings, they may be relatively inexpensive, and periodic replacements of them to insure continuity of service may be made at moderate cost.

III. Thermionic rectifiers developed primarily for use in the radio and communication fields, but extensively used for a great variety of laboratory purposes and in measuring and recording apparatus. Ratings are from a few score to a few hundred milliamperes; the RCA-82 and RCA-83 tubes are examples of this class of equipment. At present none of the tubes in this class employ mercury-pool cathodes or igniter rods; they are all primarily small-current devices.

261. Rectifying and Filtering. Simple rectification can be accomplished without either grids or igniter rods. A simple rectifier usually has a single cathode with as many anodes (often called plates in low-rating devices) as the circuit requires. Two anodes are needed for full-wave single-phase operation, six for the six-phase operation which is customary in three-phase circuits.

Full-wave single-phase rectification is accomplished by circuits of the general type illustrated farther on in Fig. 185, Section 274. The switching function performed by the rectifying elements in that circuit reverses the polarity of the voltage applied to the load in each alternate half-cycle. In the simplest type of operation no filter is used, the load is purely resistive, and a *sine loop* voltage, illustrated in Fig. 186*b*, appears at the load terminals. This sine loop voltage is unidirectional but pulsating.

The rectifying process is usually completed by means of a filter circuit which permits brief periodic energy storage. Some of the energy delivered during the peaks of periodic power input is held over in the filter and delivered during the valleys of power input.

In electronic rectifying circuits this energy storage is accomplished by means of condensers and inductances. The simplest filter consists simply of a low-resistance, high-inductance choke in series with the load. Fig. 186*a*, *b*, and *c* contrast the alternating-current rectifier input voltage, the sine loop voltage at filter input, and the direct-current voltage at the load when such a simple filter is used under load conditions that permit practically perfect filtering.

Fig. 177*a* illustrates the voltage waves of the three-phase power supply for the six-phase rectifying arrangement of Fig. 176.[PP,QQ,RR]

The solid line in Fig. 177b is the corresponding voltage between transformer neutral and rectifier cathode, and Fig. 177c the voltage across the load if the "ripple" is completely suppressed by the inductance in series with the load. In general, in a circuit containing n rectifier anodes, each anode is in service during the $1/n$th of each cycle for which its circuit voltage is greater than that of any other anode. Therefore the greater the number of anodes, the smaller is the ripple to be filtered out.
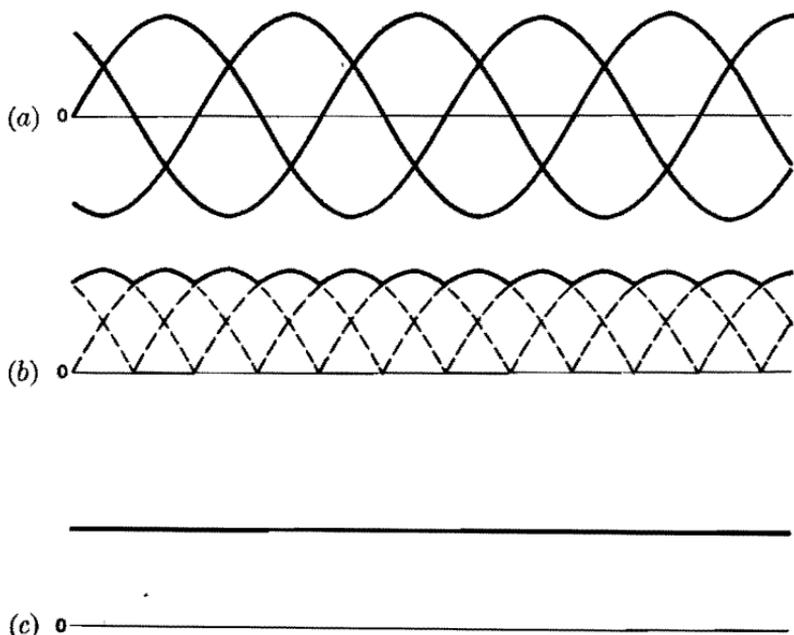
(a)   Three-phase alternating-current input voltage.
(b)   The heavy wavy line is the voltage between transformer neutral and rectifier cathode; it is also the load voltage at no load.
(c)   Voltage at load, full load operation.

FIG. 177.   Rectified and filtered voltages for the six-phase circuit of Fig. 176.

262. Alternating-Current Switching; Control of Average Current.

Switching, control of current magnitude, relaying, and inversion can be accomplished by the use of thyratrons or ignitrons.

Neither grids, as in thyratrons, nor igniter rods, as in ignitrons, will stop the arc, that is, interrupt forward current flow, once it is established. Some laboratory progress has been made toward the development of a circuit-opening type of grid, along the lines discussed in Section 254.

The lack of circuit-opening or "shut-off" control [113] is no great handicap in alternating-current switching applications, for in all alternating-

current circuits the current automatically goes to zero twice in every cycle. Each new half-cycle presents an opportunity to open the circuit permanently by preventing the restarting of the arc. A 60-cycle circuit offers 120 such opportunities in every second.

Thyratrons and ignitrons cannot be used to regulate the value of instantaneous current, but they can be used to control the average value of current. Auxiliary grid or igniter-rod circuits can be provided which initiate arcs at controllable phase points in every half-cycle. In this way current flow is made to take place during an adjustable fraction of each half-cycle. The *average* main current value depends on how large this adjustable fraction is, and can be controlled by controlling the phase-angle lag of the auxiliary circuit. This method of current control is described in some detail farther on in Section 281, in connection with Figs. 192–195.

263. Inverse Voltage Rating of Rectifiers; Arc-Back. As ordinarily used in alternating-current circuits, a rectifier's anode is strongly negative relative to the cathode in alternate half-cycles. The *inverse voltage* rating specifies the maximum safe instantaneous value of this negative anode voltage. All rectifiers, large and small, with and without grids and igniter rods, are given inverse voltage ratings, although the word "inverse" may not always be used.[133]

In multianode rectifiers failure in the inverse direction may result from unwanted initiation of a cathode spot on an idle anode, called "arc-back."[115] This results in an internal short-circuit between anodes.

In spite of the placement of the several anodes in "anode arms" or partially partitioned-off cells, the presence in the main chamber of the plasma that is necessary to carry current to the active anode sometimes results in arc-back to an idle one. Grids enclosing the anodes discourage arc-back, but produce some increase of arc drop.

Arc-back does not occur in polyphase rectifiers of the igniter type, because the anodes are located in separate chambers. Each chamber has its individual cathode pool. Each arc is completely extinguished during inverse half-cycles, and initiated at the beginning of each forward half-cycle by its own rod. Because the arc in each chamber is completely extinguished during inverse half-cycles the cathode-to-anode spacing can be made very small without incurring the risk of formation of an inverse arc. The arc drop and associated power loss are correspondingly small.

Full-wave single-phase rectifiers of the small-current type (Class III, Section 260) often have the two arcs of each tube located in a common chamber, as an aid to compactness and economy of construction. In the design of such tubes attention must be given to the necessity of avoiding

arc-back. For example, in the RCA-83 tube, each filament is so completely enclosed by its own plate structure that extensive reaching-out of the plasma of the active arc toward the idle anode is very much discouraged.

The following conception of the mechanism of arc-back has been helpful to the author:

The plasma that is essential to the passage of current to the active anode tends to spread into all parts of the arc chamber, and to form sheaths at all surfaces. Each idle anode is, during part of each cycle, at a very low potential relative to this plasma. During these cyclic periods the " current-carrying sheath " that forms adjacent to an idle anode (see Sections 251 and 252) terminates in a very steep gradient at the outer face of the sheath (the face adjacent to the idle anode surface). There exists then a correspondingly high probability of initiation of a cathode spot on the idle anode surface. It is likely that this probability may be increased by the presence of impurities or mercury droplets on the anode surface.

Experiments by Slepian[115] have shown that the presence of the remote arc in a chamber having two or more anodes gives rise to a definite momentary probability of arc-back, the probability being much greater for large than for small inverse voltages, and dependent on various operating conditions and design details. Those experiments form part, but only part, of the basis for the author's belief that there exists a definite probability of formation of a cathode spot at the outer face of any current-carrying sheath, if the voltage gradient at the outer face exceeds some definite critical value. This probability is presumably a systematic function of the voltage gradient at the outer face of the sheath and so of the potential difference across the sheath and the ion concentration within it. A mathematical analysis of a similar probability of ignition of the arc between igniter rod and pool appears in Section 273.

264. Forward Voltage Rating. The *forward voltage rating* of a rectifier specifies the maximum safe instantaneous voltage when the anode is positive, but the anode circuit held open by the rectifier. Since only grid- or igniter-controlled rectifiers (thyratrons or ignitrons) are expected to maintain open circuits in spite of forward anode voltages, only these types are given forward-voltage ratings.[133]

Failure of a thyratron to maintain properly an open circuit in the forward direction can result if the grid does not limit to a sufficiently small rate the entrance of electrons into the anode field. Various types of abuse, particularly overloads that spatter cathode material onto the grid, making it a weak *source* of free electrons when heated by prox-

imity to the discharge, can cause gradual decrease of the maximum safe forward voltage.

Failure of ignitrons to maintain properly an open circuit in the forward direction can result only from direct internal spark-over.

265. Current Ratings. Mercury-vapor rectifiers of all kinds are given average (not root-mean-square) current ratings, which specify the greatest average currents that can be carried continuously without overheating any of the parts. Such ratings are made subject to specified limits of ambient (i.e., surrounding-atmosphere) temperature. As with all electrical devices the heating depends on the rate at which energy must be dissipated.

In addition to filament heating power, if any, a rectifier must dissipate the power represented by the product of current and cathode-to-anode voltage drop. This is rarely less than 5 or more than 25 volts. Although this " arc drop " has a tendency to rise when the current decreases to very small values, it is usually affected very little by current variations within the normal operating range. Therefore the power to be dissipated is proportional to the average rather than to the root-mean-square current, so that the long-time continuous rating is expressed in average amperes. Ratings also usually express in one form or another the permissible short-time overload currents.

Power loss in a rectifier is proportional to the arc drop as well as to the average current. Therefore any modification in design that increases the arc drop, such as the inclusion of grids either to provide grid control or prevent arc-back, necessitates a reduction in the current rating. A smaller current produces the same heating effect in the modified design that the original rated current does in the original design. The small cathode-to-anode spacing permissible in ignitron rectifiers having anodes in separate chambers favors a small arc drop and small power dissipation.

Thermionic cathode gaseous rectifiers of the industrial type are usually given:[133]

(1) *Average current* ratings, together with specifications of the maximum time, usually between 10 and 30 seconds, over which the average is to be taken. Tubes overheat if operated continuously at greater than the rated average currents.

(2) *Maximum instantaneous* or *crest* ratings, which specify the maximum value the current should be permitted to reach in each duty cycle. The crest rating is likely to be several times the average current rating.

The electron emission from the cathodes of thermionic mercury-vapor rectifiers is considerably aided by the voltage gradient at the cathode surface, in the manner suggested in Section 248. During

perfectly normal duty cycles the cathode fall of potential may have to rise considerably above the ionizing potential of the gas in order to set up a gradient steep enough to release electrons as fast as the circuit demands.

The cathode fall of potential is a measure of the energy with which incoming ions strike the cathode surface. Progressive destruction of the coating on the surface by ion bombardment sets in if this energy exceeds a critical value, which usually corresponds to an arc drop between 20 and 25 volts. The crest current rating must be small enough to keep the energy of ion bombardment always below the critical value.[113]

If the current continues, either steadily or intermittently, to exceed the crest rating, the coating on the surface is gradually destroyed. As a result (a) the maximum safe instantaneous current becomes smaller, (b) the life of the tube is shortened, and (c) cathode material may be scattered on the grid, making the grid surface a partial emitter, and thereby introducing a tendency toward false current initiation.

Mercury-vapor rectifiers of all types are likely to be subjected on rare occasions to momentary "surge currents" as a result of short-circuits in associated apparatus. Although short-circuit currents are interrupted as quickly as possible by fuses or circuit-breakers, the interruption is never instantaneous. It is desirable to employ rectifying devices that will not be destroyed by surge currents before circuit-interrupting devices have time to operate. In order to permit estimates as to ability to withstand occasional severe short-circuit currents without destruction, rectifiers are sometimes given *surge current ratings*.

The smaller thermionic cathode rectifiers (those developed for radio and communication work) are given simply maximum average current ratings that take care both of temperature and cathode life restrictions. Such ratings are usually expressed in milliamperes.

266. "Clean-Up" of the Conducting Gas. The concentration of gas particles has an important influence on tube drop and on the control characteristics of rectifiers containing grids, because both of these properties depend on the electronic mean free path. The mercury-atom concentration in a mercury-vapor rectifier is affected by the temperature of the condensed mercury in the tube, in the manner described in the next section. In some rectifiers an inert gas (helium, argon, or neon) is used instead of mercury vapor, in order to avoid variation of gas particle concentration, and so of tube properties, with changes in temperature.

If the current in a tube containing an inert gas is kept below the maximum instantaneous rating, the amount of gas in the tube remains very nearly constant. Therefore the internal gas concentration also remains

very nearly constant. Operation above the maximum instantaneous or crest current rating starts disintegration of the cathode and at the same time causes rapid " clean-up," that is, removal of the gas.[116, ZZ, LL] The gas that disappears does not leak out; it becomes attached to the tube walls or electrodes by some mechanism that is closely associated with the disintegration of the cathode. It may be that dislodged cathode particles capture gas particles and " nail them down " to the walls. Slow clean-up of the gas takes place during normal operation.

267. Concentration of Mercury Atoms in Mercury-Vapor Rectifiers. Clean-up does not occur in mercury-vapor rectifiers, because fresh vapor is continually provided by droplets of condensed mercury in thermionic cathode devices and by the pool in mercury-pool-type apparatus. Absence of clean-up is one of the important reasons for the widespread use of mercury vapor as the conducting medium in gaseous rectifiers.

In a closed basement containing a pool of water the humidity eventually reaches 100 per cent. This means that the water vapor pressure becomes that at which condensation into the pool takes place just as rapidly as evaporation from it. Similarly the mercury-vapor pressure inside a mercury-vapor rectifier is always that at which the rate of condensation into the pool or droplet just equals the rate of evaporation.

The rate of evaporation, therefore also the mercury-vapor pressure, depends on the temperature of the condensed-mercury pool or droplet. The relation between mercury-vapor pressure and temperature *when vapor and condensed mercury have a common temperature* is described in Fig. 178 and Table XVI. This dependence is essentially exponential in character, and is in fact very similar to the dependence of thermionic emission current on filament temperature.

Let L_g symbolize the number of mercury vapor atoms that arrive per second per square centimeter at the droplet surface. All such particles condense into the droplet. L_g also describes the rate of evaporation from the droplet, because of the evaporation-condensation equilibrium that exists. Assume the vapor and droplet temperatures to be the same. Then by analogy with Equation (454)

$$L_g = \frac{N_{Hg}\alpha_{Hg}}{2\sqrt{\pi}} = \frac{N_{Hg}}{2\sqrt{\pi}}\sqrt{\frac{2kT_{Hg}}{m_g}} \tag{838}$$

Here α_{Hg} is the random velocity that is characteristic of the temperature T_{Hg} of the condensed mercury of the droplet (degrees Kelvin). N_{Hg} is the gas concentration at which evaporation-condensation equilibrium exists *if the vapor and condensed mercury both have the temperature T_{Hg}.*

The mercury vapor and condensed mercury in a rectifier do not necessarily have the same temperature, even when the rectifier is dark.

In a thermionic rectifier the droplet is in intimate contact with the base of the tube, whereas the vapor is in intimate contact with the upper parts of the tube that may have been heated by a previous discharge. Let α_g symbolize the random velocity characteristic of the actual vapor

Temperature of Condensed Mercury, also of Mercury Vapor, in Degrees Centigrade
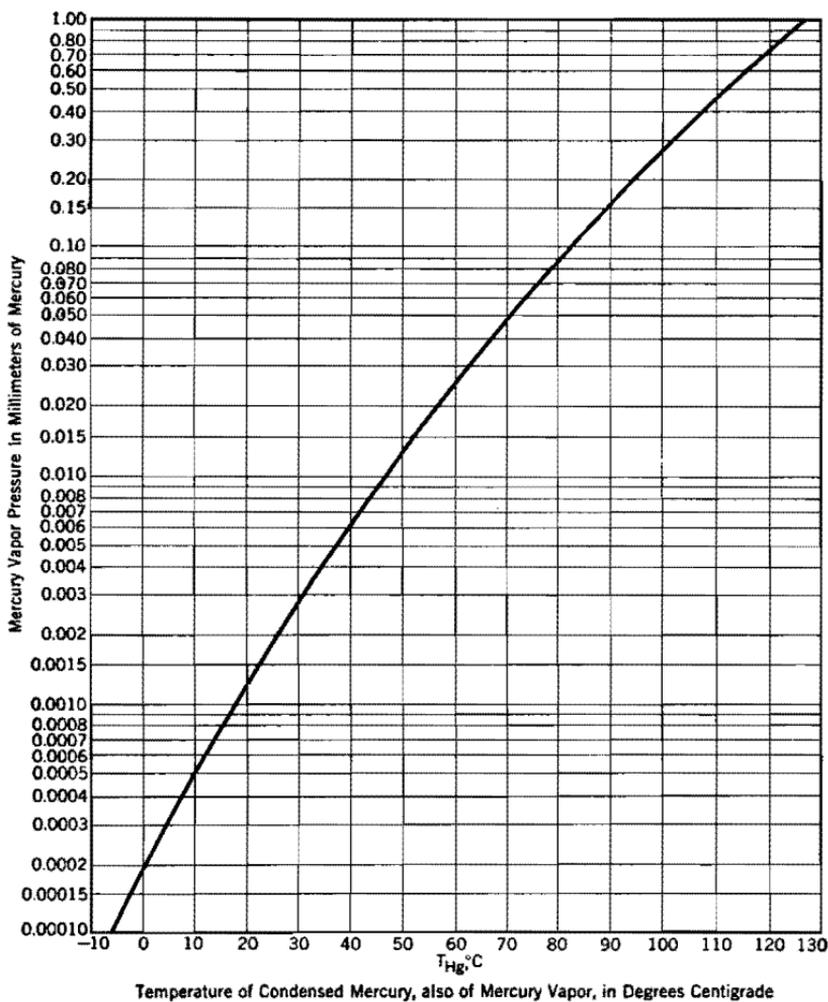
Fig. 178. Dependence of mercury-vapor pressure on temperature.

temperature T_g, and N_g the actual mercury-atom concentration in the vapor. The existence of evaporation-condensation equilibrium requires that

$$L_g = \frac{N_g \alpha_g}{2\sqrt{\pi}} = \frac{N_g}{2\sqrt{\pi}}\sqrt{\frac{2kT_g}{m_g}} \tag{839}$$

Combination of Equations (838) and (839) shows that:

$$N_g = N_{Hg} \sqrt{\frac{T_{Hg}}{T_g}} \tag{840}$$

T_{Hg} can be measured with reasonable accuracy by a thermometer at the base of the tube. If T_{Hg} is known, N_{Hg} can be calculated, from the data of Table XVI or Fig. 178, by making use of the fact that the concentration of any gas or vapor at 0° C and 760 millimeters pressure is 2.705×10^{19} gas particles per cubic centimeter, and varies directly with gas pressure, inversely with absolute temperature. In many cases only a rough estimate of the value of T_g can be made. Probably T_g is usually between 2 and 4 times T_{Hg} (degrees Kelvin) in continuously operated thermionic mercury-vapor rectifiers.

The effect on gas concentration of variations in T_g is not important, but that of variations in T_{Hg} is very important, because of the exponential nature of the dependence of N_{Hg} on T_{Hg}.

The gas concentration is of interest chiefly because it determines the length of the electron mean free path. 30 centimeters is probably a fair estimate of the order of magnitude of the electron mean free path in mercury vapor when $T_g = T_{Hg} = 293°$ K (20° C). The corresponding value when $T_g = T_{Hg} = 323°$ K (50° C) is about 3 centimeters. Mercury-vapor-particle mean free paths are presumably about one-sixth of these electron mean free paths [from Equations (486) and (487)].

268. Grid Control of Arc Initiation. Grid control of the initiation of the arc between cathode and anode is possible with both thermionic and hot-spot cathodes, although with the latter type of cathode the hot spot must be in existence, operating to an auxiliary anode, before the grid can function.

The cathode, grid, and plate of a thermionic thyratron, Fig. 175, are arranged so as to provide, *before arc initiation*, the same sort of control over the electric field adjacent to the cathode as exists in a high-vacuum triode.[111] Fig. 179a illustrates a mutual characteristic curve for a high-vacuum triode, with cut-off at plate and grid voltages of 600 and -2.25 respectively. The corresponding diagram for a typical thyratron is shown in Fig. 179b. If the grid voltage of this last-mentioned tube is brought up toward zero from strongly negative values, the arrival at a " cut-in " point, comparable with the high-vacuum " cut-off " point, is marked by an *abrupt* initiation of current flow through the tube.

This abrupt appearance of current is the result of the formation of an electric arc between cathode and anode. The value of current after the formation of the arc depends entirely on the circuit external to the tube.

The cathode-to-anode voltage after arc initiation is the normal tube drop, which is usually between 8 and 20 volts.

In general, arc initiation or cut-in occurs when the grid potential rises high enough to permit an appreciable " dark " electron current to flow to
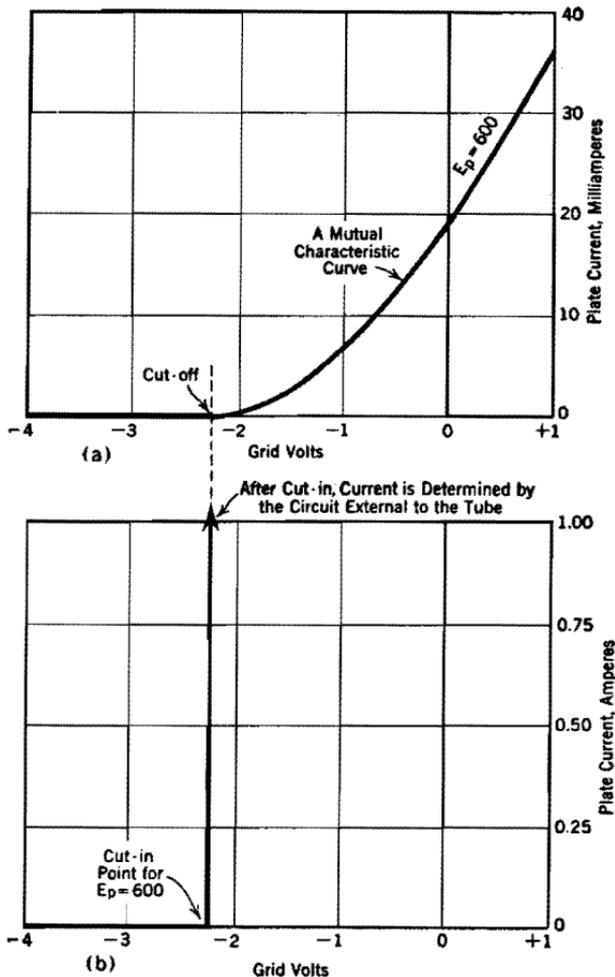
(a)   Cut-off in a high-vacuum thermionic triode.
(b)   Cut-in, or arc initiation, in a thyratron.

Fig. 179.   High vacuum cut-off and gaseous-conduction cut-in.

the plate.   This dark electron current is just the same kind of current as that flowing in a high-vacuum triode at grid-voltage values in the neighborhood of cut-off.   Discussions of the conditions that govern the magnitude of this type of current appear in Sections 99, 124, and 125. A discussion of its triggering effect appears in Section 269.

At the moment of arc initiation the voltage difference between cathode and anode decreases with extreme rapidity to the normal arc drop of the tube. The cathode-ray oscillogram reproduced in Fig. 180 records a typical collapse of plate voltage. In the operation there recorded the tube *voltage* dropped from several hundred volts to a very small value in less than half a microsecond. Ordinary circuits contain enough inductance to make the growth of tube *current* to its steady state value require from a few to a few hundred microseconds.
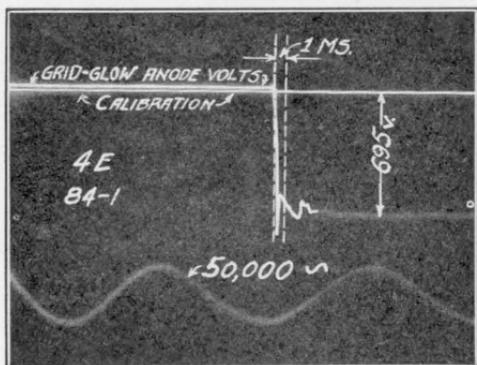
Fig. 180. Cathode-ray oscillogram of the collapse of anode-to-cathode voltage at cut-in (arc initiation) in a thyratron (also sometimes called a grid-glow tube). The irregular oscillations after collapse probably were caused by reflections in the lead-in circuit to the oscillograph.

Once the arc has been initiated, a decrease of grid potential has no effect on the current in the tube. The grid is immersed in a plasma, and a decrease of the grid potential to a low negative value merely increases the thickness of the current-carrying sheath that forms around it, in the manner described in Section 254. Grid control is reestablished only when changes in the plate circuit external to the tube bring the arc current to zero, as at the end of a half-cycle of alternating current. Any marked *rise* of grid potential after arc initiation makes the grid become a second anode, as described later in Section 270.

269. Mechanism of Arc Initiation; Grid Control Curves. It was pointed out in Section 99 that there is no such thing as mathematically defined cut-off. High-vacuum triode mutual characteristic curves like that in Fig. 179a are merely rapidly asymptotic to zero, in the neighborhood of a value of grid voltage that is called the cut-off value.

As the grid voltage is made to fall below this cut-off value, the negative potential dip outside the cathode [curves (3) and (4), Fig. 77] becomes very pronounced, and E_m correspondingly large. At each value of grid voltage in the neighborhood of cut-off the plate current is some small but definite fraction of the thermionic current, in accordance with Equation (361).

In an exactly similar way there is a very small, but definite, plate current due to electron flow at each value of grid voltage just below the cut-in point of Fig. 179a. Of course the electrons that constitute this

current produce ions along their rapid flight to the anode; these ions move relatively slowly to the grid. The positive space charge due to their presence tends to lift the bottom of the negative potential dip, because positive space charge located anywhere between the electrodes has the same kind of effect on potential distribution near the cathode that a rise in plate voltage has.

As the bottom of the negative dip is in this way lifted, the electron current is increased. The increased electron flow makes the positive ion space charge become larger, this growth in space charge lifts the bottom of the dip higher, so permitting more electrons to flow, giving rise to more positive space charge, and so on. At the definite value of dark electron current that corresponds to the cut-in point this pyramidal growth becomes unstable and accelerates itself indefinitely, the end condition being the plasma of an arc.

The electron mean free path is usually greater than the cathode-to-anode spacing, so that only a small percentage of the dark-current electrons hit gas particles during flight to the anode. Furthermore, the high plate voltage that exists prior to cut-in makes practically all electrons that do hit gas particles acquire much more than ionizing energy before they hit. Under these conditions the rate of ion production depends primarily on how many gas particles are hit per second, rather than on the electron energy at impact.

If the gas concentration is low, many electrons must flow in order to make enough hits per second to initiate a plasma. If the gas concentration is high, each electron has a considerable chance of making a hit, so that only a few electrons may need to flow in order to make the same number of hits per second. Therefore when the gas concentration in a tube is high, the critical triggering current is less, and cut-in occurs at a more negative potential, than when the gas concentration is low.

Each of the various slant lines in Fig. 101, page 265, describes combinations of a triode's plate and grid voltages that are necessary to give the plate current some definite value. The larger the definite value, the farther to the right is the slant line. The thyratron *grid-control curves*[111] illustrated in Fig. 181 are essentially just the same kind of slant lines as those in Fig. 101. They describe combinations of plate voltage and grid voltage that produce definite dark electron currents just sufficient to trigger the initiation of a plasma. High condensed-mercury temperature corresponds to high mercury-vapor concentration, so to a considerable chance of any electron's hitting an atom, so to little triggering current. Therefore the high-temperature grid-control curves lie to the left of the low-temperature ones.

In tubes that contain inert gases instead of mercury vapor the gas

concentration is unaffected by temperature.   Therefore each such tube
has a single grid-control curve rather than a family of them.

The direct applicational significance of grid-control curves can be
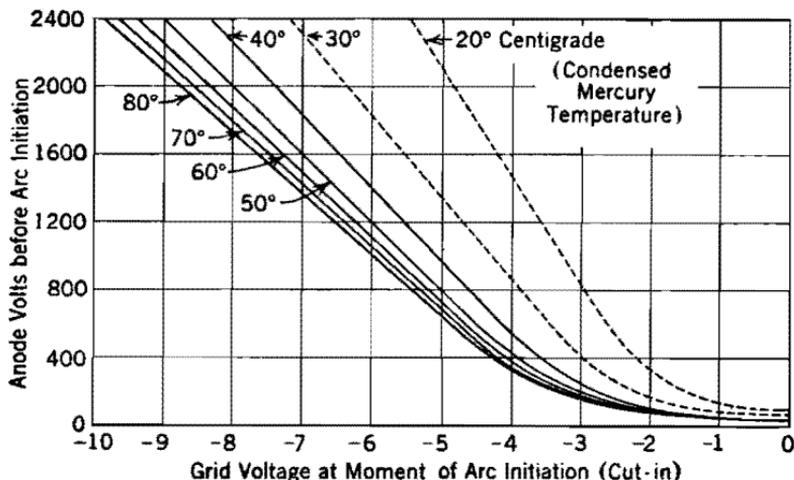
FIG. 181.   Typical grid-control curves for a General Electric type FG-17 thyratron.
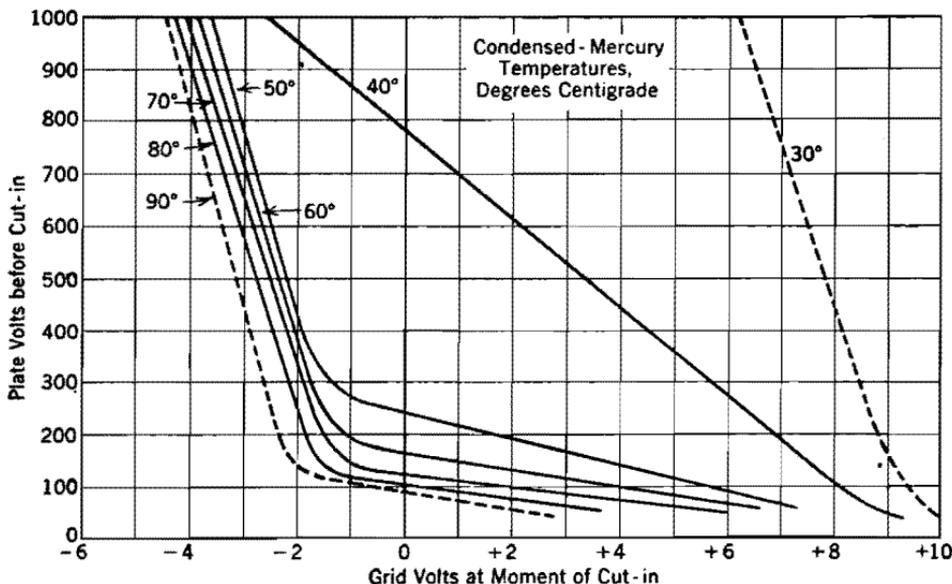
FIG. 182.   Typical grid-control curves for a General Electric type FG-67 thyratron.

made clear by a specific example.   Suppose that the tube whose grid-
control curves appear in Fig. 181 is dark, a strongly negative grid
potential holding the circuit open in spite of a steady plate potential of
800 volts.   Condensed-mercury temperature is 50° C.   If now the grid

potential is gradually raised toward zero no change is noticed either in tube appearance or in readings of meters in the plate circuit until the 800-volt point on the 50° grid control curve is reached at about -5 grid volts.   At this point the tube abruptly closes the plate circuit, as a result of the formation of an arc between cathode and anode.   Driving the grid negative will not subsequently reopen the plate circuit.

The grid-control curves of the General Electric FG-67 type of tube, Fig. 182, are less regular in shape than those in Fig. 181.   At voltage combinations in Fig. 182 that are described by points between the portion of a grid-control curve that lies below its first bend and a straight-line downward extension of its upper part, visible luminosity appears between grid and plate, and the plate current has an easily measurable but small value.   This behavior indicates that the triggering current sufficient to initiate a plasma at higher plate voltages is at lower ones able only to initiate some less intense kind of discharge which is not self-sustaining.

270. Current-Limiting Grid Circuit Resistors.   After the arc in a thyratron is initiated in the manner described in the preceding section, the grid is immersed in a plasma, and is surrounded by a current-carrying sheath of the type discussed in Sections 251–255.   Any attempt to raise the grid potential appreciably above that of the anode makes the grid become a second arc anode, as at E in the grid or probe volt-ampere curve shown in Fig. 171a.   After this happens the current in the grid circuit depends, at least partly and usually almost wholly, on the properties of the grid circuit external to the tube.   For this reason it is essential to use current-limiting resistors in the grid circuits of thyratrons. These grid resistors should be chosen so as to limit average grid currents to not more than the permissible rated values.

271. Shield-Grid Thyratrons.   The grids of thyratrons must exercise electrostatic control over all possible paths of emergence of electrons from cathode toward anode, because the production of ions along any such path can result in arc initiation.   The grids therefore have rather large areas.

In a *shield-grid* thyratron [117] the shield grid forms an enclosure around the cathode except at the anode exposure, and is kept at a low enough potential (often just the cathode potential) to prevent arc initiation anywhere except at the anode exposure.   The arc is initiated by varying the potential of the control grid, which is very small and is so designed and placed as to have electrostatic control of the dark electron flow along the direct cathode-to-anode path.

It is especially important to have thyratron grid excitation circuits behave properly during the periods when tubes are dark and grids negative but swinging up toward positive values.   Very small dark

currents flow during these periods, because of the movement to the grid of the few ions that are present even before arc initiation.   The operation of some very sensitive grid circuits is undesirably affected by these dark currents to negative grids, if ordinary single-grid thyratrons are used. Shield-grid thyratrons serve advantageously under such conditions, because their control grids have small areas, and are otherwise so designed that, when negative, dark currents to them are very small, and bright currents moderate.

A further advantage of shield-grid thyratrons is that the capacitances between control grids and other electrodes are less than in single-grid thyratrons.

272. Igniter-Rod Control of Arc Initiation.   Fig. 174, page 482, illustrates the arrangement of parts in an ignitron [114] that falls under the industrial electronic tube classification.   Ignition of the arc occurs if the potential gradient in the igniter rod at the junction with the mercury surface reaches or exceeds a critical gradient that may be expected to be between 50 and 200 volts per centimeter.   The critical gradient is likely to be considerably smaller when the rod has been heated by continuous operation, than when at room temperature.   There is some tendency for the critical gradient to show gradual but continued increase during the entire span of a tube's useful life.

The potential gradient in the rod is established by the passage of an auxiliary current of several amperes.   This current may be provided by connecting the rod to the anode through a resistance, or by an auxiliary igniting circuit.   The energy required for ignition may be very small, for although the instantaneous power in the igniting circuit may be several hundred watts, ignition is usually accomplished in a few microseconds.

When the rod gradient exceeds the critical value, an arc forms at the rod-mercury junction.   The lower terminal of this arc, a hot cathode spot, moves away from the rod along the surface of the pool.   The upper arc terminal travels upward along the rod with extreme rapidity; when it reaches the metal of the supporting bracket the rod is of course short-circuited by the arc, which now plays between bracket and pool.

Experiments have shown that the arc between cathode pool and anode forms within a small fraction of a microsecond after the cathode spot is initiated at the rod-mercury junction.[118]   However, the time required for the formation of the cathode spot after establishment of the gradient in the rod follows a simple statistical law.   The kind of statistical law followed is very similar to that found by Slepian and Ludwig to describe arc-back frequency in mercury-arc rectifiers.[115]   It seems likely that there exists for both situations a probability of cathode-spot initia-

tion which is a simple exponential function of the gradient at the location of interest. In one situation the location of interest is the rod-mercury junction; in the other the surface of the idle anode.

273. Statistical Variation of Firing-Time in Ignitrons. The mathematical analysis which appears below for the variation in time to arc ignition by an igniter rod [118] applies with slight modifications to the arc-back situation.[115] The chief difference is that the time to arc initiation by an igniter rod is a matter of microseconds, and that to initiation by arc-back is a matter of hours, days, or months. The mathematical method employed is very similar to that used in Section 132, but in the present application time rather than distance is the important variable.

Experiments have shown that for any given set of igniter-rod operating conditions there is a definite probability $p\,dt$ of ignition of an arc within a brief time interval dt, where p is primarily a function of the gradient in the rod. If P is the probability that ignition occurs later than t seconds after the application of the gradient,

$$-dP = P \cdot p\,dt \tag{841}$$

or

$$d \log_\epsilon P = p\,dt \tag{842}$$

The reasoning on which Equation (841) is based is as follows: the chance of ignition occurring at some time later than t is by definition P, and the chance of ignition in the particular interval dt *immediately following* t is $p\,dt$; the product $P \cdot p\,dt$ is therefore the chance that ignition will not occur until t seconds have passed, but will occur in dt seconds immediately thereafter. This product is therefore the negative differential change in P, negative because P must become less for longer times. The solution for P, if p is independent of time, is

$$P = \epsilon^{-pt} \tag{843}$$

If p is a function of time, the solution is

$$P = \epsilon^{-\int_{t=0}^{t=t} p\,dt} \tag{844}$$

In arriving at these expressions it is recognized that $P = 1$ when $t = 0$, that is, ignition is certain to occur later than $t = 0$.

Below a rather definite critical gradient the *momentary ignition probability* p is vanishingly small, but above the critical gradient it depends on the actual value F of the gradient, according to the relation

$$p = \frac{1}{t_0}\epsilon^{\frac{F}{F_0}} \tag{845}$$

The corresponding solution for P is

$$\log_\epsilon P = -\frac{t}{t_0} \epsilon^{\frac{F}{F_0}} \tag{846}$$

In these expressions F is the gradient along the rod, t the time delay to ignition, and t_0 and F_0 empirical constants.

The curves and lines in Fig. 183 compare the results of a series of tests with the behavior predicted by the above equations. If m is the number of ignitions that are observed to be delayed longer than time t, and M the total number of ignitions observed, a plot of m/M against t produces a staircase type of graph. The solid-line steps in Fig. 183 were located from actual test data in this way. The dotted lines represent predictions from Equation (846) when F_0 and t_0 are given the values 41 volts per centimeter and 72 microseconds respectively. These dotted lines are smooth rather than stepped because they describe the expectations in case of a very large number of observations. The deviations of the experimentally determined stepped curves from the smooth dotted ones are no greater than is to be expected from the statistical nature of the ignition process. Fig. 184 is similar to Fig. 183; however, in Fig. 184 an inverse logarithmic per cent scale is used, and the solid lines are derived from a different set of observations. The empirical constants are also different.

By letting $P = \frac{1}{2}$, Equation (846) can be used to obtain an expression for the dependence of *median ignition time* t_m on gradient, as follows:

$$t_m = t_0 \epsilon^{-\frac{F}{F_0}} \log_\epsilon 2$$
$$= 0.693 \, t_0 \epsilon^{\frac{F}{F_0}} \tag{847}$$

Equation (847) shows that the empirical constant t_0 is proportional by the factor 0.693 to the median time to be expected at zero rod gradient if the empirical law were still valid there, as of course it is not.

The expression just given for median firing-time can be rewritten in terms of the *excess* of gradient over the critical gradient F_C. To do this, Equation (846) can be rewritten as follows:

$$\log_\epsilon P = -\frac{t}{t_1} \epsilon^{\frac{F-F_C}{F_0}} \tag{848}$$

This is the same as Equation (846) if

$$t_0 = t_1 \epsilon^{\frac{F_C}{F_0}} \tag{849}$$
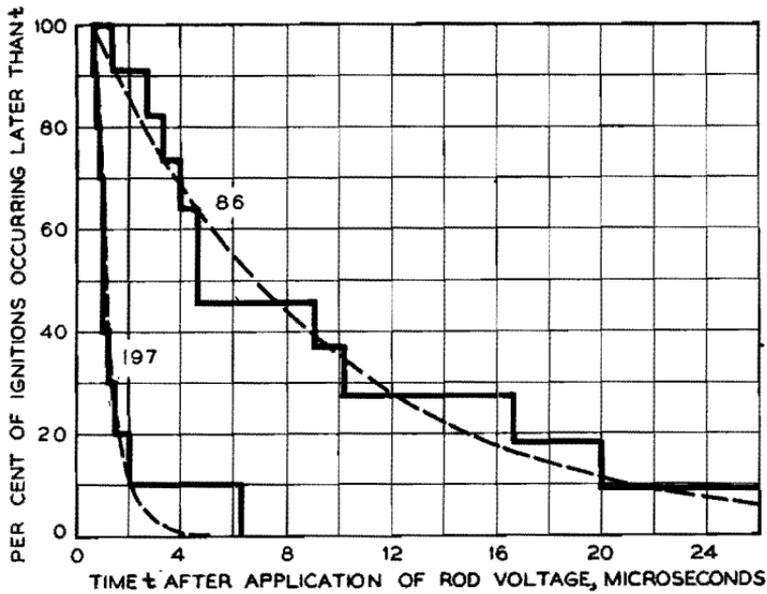
FIG. 183. Per cent of ignitions occurring later than time t, with igniter-rod gradients of 86 and 197 volts per centimeter, in an ignitron.
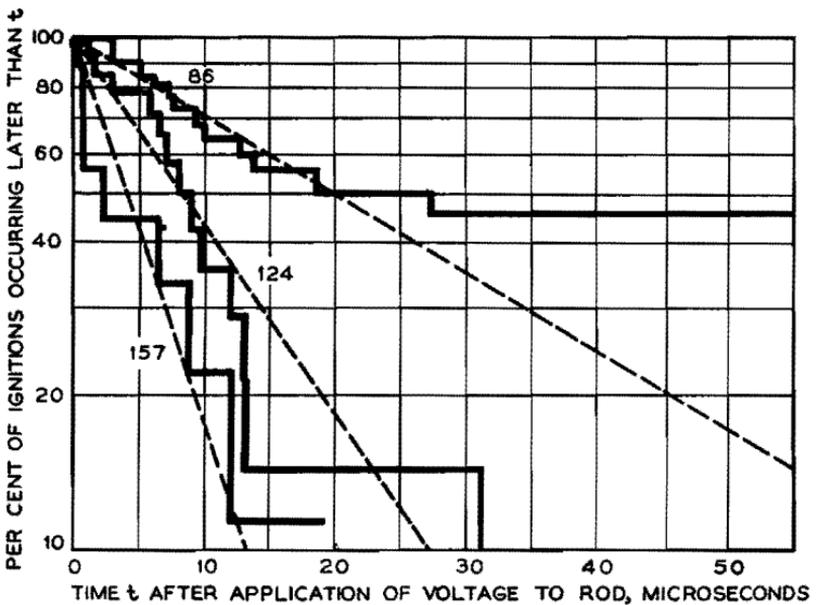
FIG. 184. Per cent of ignitions occurring later than time t, with igniter-rod gradients of 86, 124, and 157 volts per centimeter, in an ignitron. Logarithmic per cent scale.

Therefore also

$$t_m = 0.693 t_1 \epsilon^{\frac{F_C}{F_0}} \epsilon^{-\frac{F}{F_0}} \qquad (850)$$

Here t_1 is the median firing time at the critical gradient.   This expression for median firing time has the merit of using the critical gradient and the median firing-time at that gradient as bases of reference.   Both these quantities have very real physical meanings.

PROBLEMS

CHAPTER XXI

1. Fig. E illustrates the type of relation to be expected between metal case potential (relative to the cathode), and density of current flow from plasma to case, in a metal-enclosed mercury-arc rectifier.   Note the resemblance to Fig. 171a.   For Fig. E,

(*a*) What will be the potential of the case, relative to the cathode, when the case is insulated from the rest of the circuit?

(*b*) What will be the current density to the case when at cathode potential?

(*c*) Find the electron temperature.
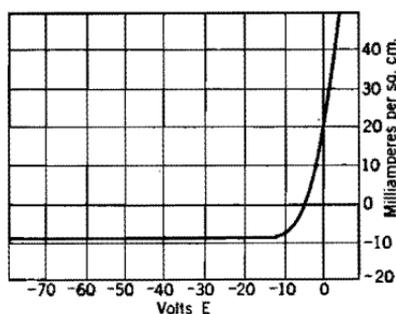
FIG. E.   Current density to the case of a mercury-arc rectifier.

2. Conditions as in Problem 1.   Determine the rate at which the plasma delivers energy to the walls of the case, in watts per square centimeter (this energy appears as heat), when the case is at cathode potential.   Plasma is 12 volts above cathode potential.   The following items should be accounted for, *each of them being calculable by multiplying a properly chosen current in amperes by a properly chosen voltage.*

(*a*) Kinetic energy acquired by the positive ions in falling through the sheath.

(*b*) Energy released by recombination of positive ions upon arrival at the surface.

(*c*) Residual kinetic energy of the electrons after penetrating the sheath.

(*d*) Heat produced by the entrance into the metal of the case, sliding down the work function hill, of the electrons not needed for recombination at the surface. Work function of the case is 4 volts.

3. Conditions as in Problems 1 and 2.

(*a*) Determine the thickness of the sheath adjacent to an idle anode whose potential is 2,000 volts below that of the plasma.

(*b*) Determine the potential gradient adjacent to the surface of this idle anode.

4. Find the concentration of mercury atoms in a mercury-vapor rectifier for which the condensed-mercury temperature is 40° C and the mercury-vapor temperature in the plasma 800° K.

5. Fig. *F* is a diagram of a circuit that employs a grid-controlled mercury-vapor rectifier to produce a saw-tooth voltage wave of the type illustrated in Fig. 41. If
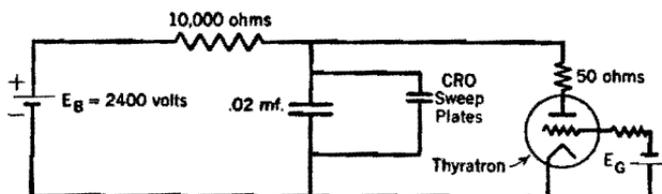
Fig. *F*.   Sweep circuit for cathode-ray oscillograph.

the tube used is an FG-17 thyratron (Fig. 181) at a condensed-mercury temperature of 40° C, and the crest of each saw-tooth is to occur at 800 volts, find

(*a*) Grid voltage required,

(*b*) Frequency,

(*c*) Approximate time required to discharge the condenser during the vertical drop from the crest of each saw-tooth.

CHAPTER XXII

SINGLE-PHASE CIRCUITS CONTAINING RECTIFYING ELEMENTS

274. Filters for Full-Wave Single-Phase Rectifiers. Fig. 185 is a circuit diagram of a type of *full-wave rectifier* which is satisfactory for supplying direct current to small and moderate loads. As in nearly all rectifying arrangements, alternating current enters through a transformer which is an integral part of the rectifier circuit. The two rectifying elements, which may be any of a variety of types, are connected on opposite sides of the center tap of the transformer secondary, in
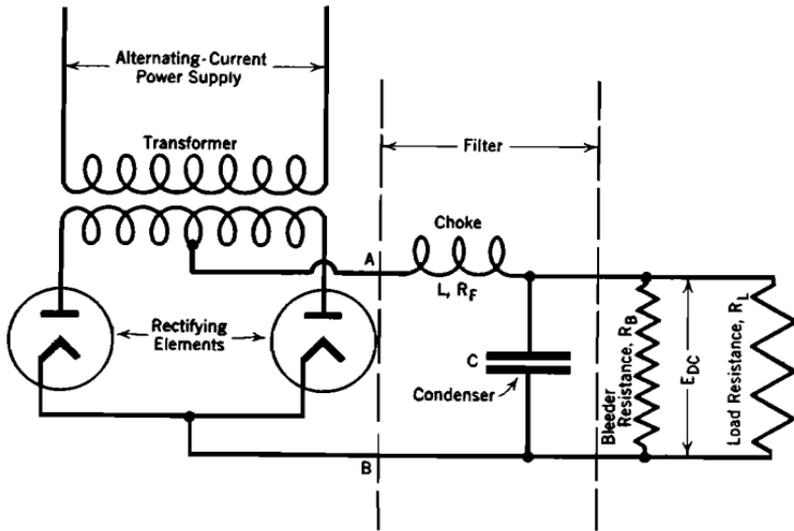
FIG. 185. Single-phase full-wave rectifier circuit with choke input filter.

order to make their anodes positive in alternate half-cycles. This makes all half-cycles have the same output polarity, as illustrated in Fig. 186*b*. If only one rectifying element is used, no center tap is needed on the transformer, and alternate half-cycles of the alternating current are completely suppressed; a circuit so arranged is called a *half-wave* rectifier.

The portion of the circuit that lies between the two dotted vertical lines in Fig. 185 is called the *filter*; its ripple-reducing function was

506

briefly described in Section 261. If the choke is placed on the input side of the condenser, as in Fig. 185, the rectifier is said to have a " choke input " filter; if on the opposite side of the condenser, a " condenser input " filter.

Choke input filters can be used successfully in any circuit of the general type illustrated in Fig. 185. Condenser input filters cannot be used in circuits in which the filter input voltage has a wave form with an abrupt rise, if the rise is produced by the action of thyratrons with oxide-coated cathodes. The abrupt change in input voltage calls for an abrupt in-

(a) Alternating-current voltage at rectifier input.

(b) Unidirectional, but pulsating, " sine-loop " voltage at filter input.

(c) Steady direct-current voltage at filter output.
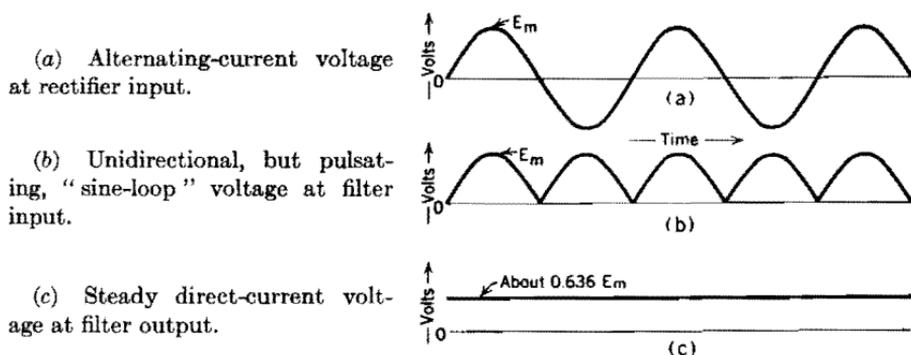
Fig. 186. Rectified and filtered voltages at full load, in a full-wave, single-phase rectifier circuit, Fig. 185. Ripple completely suppressed by the filter.

crease in the charge on the condenser; this demands a momentary charging current large enough to damage seriously an oxide-coated cathode surface, in the manner described in Section 265.

The bleeder resistance is used to improve the voltage regulation of the circuit, in a manner to be described later. This improvement is obtained at the expense of a small constant power loss.

The most interesting properties of a single-phase rectifying circuit are:

(1) The amount of direct current it can pass to the load without overburdening any of the circuit elements.

(2) The full load direct-current voltage at the load.

(3) The voltage regulation at the load.

(4) The per cent ripple in load current and voltage.

(5) Behavior under short-circuit and overload conditions.

The various parts of the circuit must have whatever current-carrying capacity corresponds to full load current at the load. In a circuit like that diagrammed in Fig. 185 the average current in the choke is the same as the load current. Each rectifying element and each half of the transformer secondary carry current half the time, therefore carry average currents half the value of that in the choke. The relation

necessary between transformer secondary voltage and load voltage will be apparent after the discussion of voltage regulation and per cent ripple.

Mercury-vapor rectifiers with oxide-coated cathodes are very often used as the rectifying elements. Cathodes of this type are rapidly and permanently damaged by severe or prolonged over-current, for reasons discussed in Section 265. It is usually desirable to protect against damage due to short-circuit or severe overload by the use of fuses.

The voltage regulation of a rectifier is defined, as for all electrical machinery, as

$$\frac{E_{\text{no load}} - E_{\text{full load}}}{E_{\text{full load}}} = \text{voltage regulation} \qquad (851)$$

The voltage values in this equation are as measured at the load terminals, that is, filter output. Alternating-current supply voltage is presumed to remain constant.

It is convenient to study the quantitative aspects of the voltage regulation and ripple in the rectifier circuit of Fig. 185 under the following three headings:

(I) Behavior with a filter that consists only of a choke, neither condenser nor bleeder resistance being used.

(II) Behavior with a filter that consists only of a condenser, neither choke nor bleeder resistance being used.

(III) Behavior with a filter that contains both choke and condenser, together with an explanation of the functioning of the bleeder resistance.

These three subjects will be taken up in order in Sections 276–279.

275. Per Cent Ripple. Filters do not deliver perfectly steady direct current to their loads; some harmonics of the frequency of the alternating-current power supply are always present. In full-wave rectifiers the second-harmonic ripples are more pronounced than ripples at any other frequencies. No fundamental-frequency voltages or currents are present in a full-wave rectifier's output, but they are very important components of the voltage and current at the output of a half-wave rectifier.

The amount of ripple at the load may be reasonably well described by:

(1) Stating the *per cent ripple*, or the *ripple fraction, r*, defined as follows:

$$\text{Per cent ripple} = \frac{e_{\text{max}} - e_{\text{min}}}{E_{\text{DC}}} \times 100 = r \times 100 \qquad (852)$$

In this definition e_{max} and e_{min} are instantaneous maximum and minimum values of the voltage at the load, so that $e_{\text{max}} - e_{\text{min}}$ is the total " swing " of that voltage. E_{DC} is the direct-current voltage at the load. The

per cent ripple can be measured from an oscillographic record of the load voltage.

(2) Or, by stating the *per cent ripple of the most important ripple voltage component*. For example, if the second-harmonic ripple is the most prominent one, as in the output of a full-wave rectifier, a satisfactory measure of the ripple magnitude is given by the *per cent second-harmonic ripple*, or by the *second-harmonic ripple fraction r″*, defined as follows:

$$\text{Per cent second-harmonic ripple} = \frac{2E_m''}{E_{DC}} \times 100 = r'' \times 100 \quad (853)$$

In this definition E_m'' is the crest value of the second-harmonic component of the load voltage. The correspondingly important definition relative to the output of a half-wave rectifier would be the per cent fundamental-frequency ripple. The per cent ripple due to the most important ripple component is usually a close approximation to the per cent ripple as defined by Equation (852), and is often much easier to predict.

276. Single-Phase Filter Circuit Behavior: (I) Filter with Choke Only. If the two rectifying elements of Fig. 185 are both conducting during all of each alternate half-cycle, the voltage E_{AB} at filter input consists of a series of periodically recurring *sine loops*, as illustrated in Fig. 186b.

Any periodically recurring voltage or current can be expressed mathematically by a *Fourier series*, consisting of a sum of a constant and sine and cosine terms of the fundamental and harmonic frequencies. The Fourier series for the sine loop voltage is an infinite one containing a direct-current term, and cosine terms of *even harmonics* of the alternating-current supply, as follows:[119]

$$e_i = \frac{2}{\pi} E_m - \frac{4}{3\pi} E_m \cos 2\omega t - \frac{4}{15\pi} \cos 4\omega t \cdots - \frac{4}{\pi} \frac{1}{(n^2 - 1)} \cos n\omega t \quad (854)$$

Only even values of n are to be used in this series. e_i is the instantaneous voltage of the sine loop, E_m its crest value, and $\omega = 2\pi f$, where f is the frequency of alternating-current power supply. Each sine loop is completed in $1/2f$ seconds.

When the filter consists only of the choke, the condenser and bleeder resistance being omitted, the two rectifying elements are both conducting during all of each alternate half-cycle, so that the sine loop voltage is the filter input voltage. The voltages corresponding to the various terms in Equation (854) produce currents in filter and load quite independently of one another. The impedance and phase angle for each

term can be expressed in the manner customary for alternating-current circuits, as follows:

$$\left.\begin{array}{l}\text{Impedance of filter and load}\\\text{for the nth harmonic}\end{array}\right\} = \sqrt{(R_L + R_F)^2 + (nX_L)^2} \qquad (855\text{ p})$$

$$\left.\begin{array}{l}\text{Power factor of filter and}\\\text{load combined, for the}\\\text{nth harmonic}\end{array}\right\} = \cos\phi_n = \frac{R_L + R_F}{\sqrt{(R_L + R_F)^2 + (nX_L)^2}} \quad (856\text{ p})$$

In these equations ϕ_n is the phase angle between the nth-harmonic voltage and current at filter input, R_L and R_F are the resistances of the load and of the filter's choke respectively, and X_L is the reactance of the choke at the fundamental frequency f. Ordinarily $R_F \ll R_L$. The load's inductance is presumed to be negligible relative to L.

The instantaneous current i_L in the load, also in the choke for the present case, is obtained by using with each term of Equation (854) the proper impedance and phase angle.   Thus

$$i_L = \frac{2}{\pi}\frac{E_m}{(R_L + R_F)} - \frac{4}{3\pi}\frac{E_m}{\sqrt{(R_L + R_F)^2 + (2X_L)^2}}\cos(2\omega t - \phi_2)$$

$$- \frac{4}{15\pi}\frac{E_m}{\sqrt{(R_L + R_F)^2 + (4X_L)^2}}\cos(4\omega t - \phi_4) - \cdots$$

$$- \frac{4}{\pi}\frac{1}{(n^2 - 1)}\frac{E_m}{\sqrt{(R_L + R_F)^2 + (nX_L)^2}}\cos(n\omega t - \phi_n) \qquad (857\text{ p})$$

Only even values of n appear in Equation (857).

The first term in Equation (857) is of course the steady direct current in load and filter, and will be represented by the symbol I_{DC}. The direct-current voltage across the load is of course $I_{DC}R_L$, that is,

$$E_{DC} = \frac{2E_m}{\pi}\frac{R_L}{R_L + R_F} = 0.636\,E_m\frac{R_L}{R_L + R_F} \qquad (858)$$

As R_F is usually small relative to R_L, the direct-current voltage at the load is approximately $0.636E_m$, which is also *the average value of the alternating-current supply voltage.*

As the load is purely resistive, it has the same impedance for any harmonic as it has for the direct current.   Therefore the per cent second-harmonic load voltage ripple is the same as the per cent second-harmonic load current ripple, and is the ratio of twice the crest voltage of the second term in Equation (857) to the first term in that equation.   This

ratio reduces to the following:

$$\left. \begin{array}{l} \text{Per cent second-harmonic} \\ \text{ripple with output choke} \\ \text{only, no condenser, full-} \\ \text{wave rectifier} \end{array} \right\} = \frac{4}{3} \frac{1}{\sqrt{1 + \left(\dfrac{2X_L}{R_L + R_F}\right)^2}} \qquad (859\ p)$$

The effects of transformer resistance and leakage reactance can be taken into account by lumping them in with R_F and X_L. Tube drop is usually of but little importance, but can be accounted for separately if desired; it has an approximately constant value at all points along each sine loop. If the full load value of R_L and the supply frequency are known, Equation (859) can be used to calculate the inductance that the choke must have to limit the per cent second-harmonic ripple to any desired value *at full load.* For example, if the ripple is not to exceed 5 per cent when the load resistance is R_L, it is necessary to make $X_L = 13R_L$.

At no load there is no current, therefore no voltage drop, in the choke, so that the complete pulsating sine loop voltage appears across the load. Thus at no load a filter consisting of a choke only has no filtering action. For that reason such an arrangement is used only when little variation in load is to be expected, and in polyphase circuits like that of Fig. 176 which have but little ripple to begin with. It has the merits of simplicity of circuit arrangement, and of economy in high-voltage circuits for which inductances are cheap and condensers expensive.

The essential features of the behavior of a filter consisting of a choke only are that (a) it reduces the ripple satisfactorily only at large load currents, and (b) the direct-current voltage at the load is approximately the average value of the sine loop voltage.

277. Single-Phase Filter Circuit Behavior: (II) Filter with Condenser Only; "Cut-Out" and "Cut-In" Points. Imagine the choke and bleeder resistance to be omitted from the circuit of Fig. 185; the filter is to consist only of a condenser. Consider first the situation at no load (R_L infinite). Suppose alternating-current voltage is applied to the circuit by throwing a switch at the instant of zero voltage (the beginning of a sine loop). Except for a small tube drop, the circuit requires the condenser potential to follow that of the sine loop exactly, up to the maximum point. *But the condenser potential cannot follow the sine loop voltage down again,* for that would require an inverse condenser-discharge current through a rectifying element, and these elements cannot pass inverse current. With the load circuit open, there is no avenue of discharge available to the condenser. It acquires a potential of E_m volts on the first rising quarter-cycle after being connected into

the circuit, then stays at that potential. Such a device is therefore a perfect filter at no load, having no ripple at all; the steady voltage at filter output terminals is just the crest value E_m of the sine loop voltage.

Now suppose that the load resistance has some finite value. Just as in the no load case, the condenser potential initially follows the sine loop voltage up to the crest. But a discharge path through the load now exists, so that the condenser potential can decrease after the crest is reached. However, the decrease cannot take place more rapidly than the flow of discharge current through the load permits, because the load is the only avenue available for the passage of condenser discharge current. See Fig. 187.

Neither can the condenser potential decrease more rapidly than the sine loop voltage does, because the rectifying elements pass condenser charging current freely.

Thus after the crest of a sine loop is passed the condenser potential *either* falls at a rate determined by the discharge circuit, *or* follows the sine loop voltage, the choice going to whichever of the two possibilities calls for the *least rapid* rate of voltage decrease.

The sine loop falls very slowly immediately after the crest is passed, so that the condenser potential at first follows the sine loop voltage away
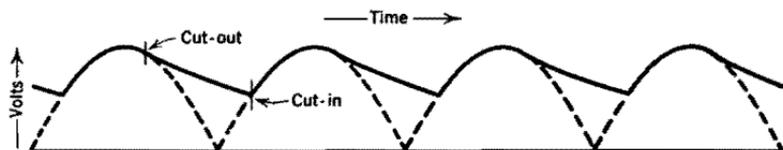
Fig. 187.   Load and condenser voltage, for a rectifier using a filter consisting of condenser only. Circuit as in Fig. 185 except that choke is omitted. Dotted lines follow the sine-loop voltage, $f = 60$ cycles, $R_L = 2,000$ ohms, $C = 4$ microfarads, $R_B = \infty$.

from the crest. The rate of sine loop decline becomes continually greater, however, but the rate of condenser discharge through the load, being proportional to the voltage, becomes continually less. Eventually a "cut-out" point is reached,[119] at which the discharge through the load tends to drop the condenser potential at exactly the rate at which the sine loop voltage is falling, so that the current in the rectifying element falls to zero. The filter and load are thereafter separated from the power supply circuit by an open circuit at the rectifying element. The separation persists until a "cut-in" point is reached as the next sine loop rises above whatever the potential to which the condenser has discharged in the interim. Fig. 187 illustrates the shape of the filter input voltage, and of the voltage at the load, for a particular set of circuit

constants.   This type of "cut-in" must not be confused with that due to the action of a grid in a gaseous grid rectifier.

It is not difficult to determine the phase angle and voltage at cut-out for this chokeless circuit.[119]   Let:

θ_1 = value of ωt (measured from beginning of a loop), at cut-out.

e_l = instantaneous sine loop voltage;

e_c = condenser potential at any moment after cut-out but before cut-in.

i_L = instantaneous load current; of course after cut-out

$$i_L = \frac{e_c}{R_L} \qquad (860\ \text{p})$$

The rates of change of e_l and e_c are:

$$\frac{de_l}{dt} = E_m\omega \cos \omega t \qquad (861\ \text{p})$$

$$\frac{de_c}{dt} = -\frac{i_L}{C} = -\frac{e_c}{RC} \qquad (862\ \text{p})$$

Cut-out is the point at which the two rates of change are equal, so, at cut-out angle θ_1,

$$E_m\omega \cos \theta_1 = -\frac{e_c}{RC} \qquad (863\ \text{p})$$

The condenser potential still has, at this moment, the sine loop value, though departure from it is imminent.   Therefore, at cut-out angle θ_1, $e_c = E_m \sin \theta_1$.   Substitution of this value for e_c into Equation (863) shows that

$$\tan \theta_1 = -\omega R_L C = -\frac{R_L}{X_C} \qquad (864\ \text{p})$$

If t_f symbolizes the cyclic period $1/f$ of the supply circuit, and t_{RC} the time constant RC of the discharge circuit, Equation (864) can be written as follows:

$$\tan \omega t = -2\pi \frac{t_{RC}}{t_f} \qquad (865)$$

Fig. 187 has been prepared to correspond to a frequency of 60 cycles, so that the duration of each sine loop is 8,333 microseconds.   The time constant of the discharge circuit is $2,000 \times 4 \times 10^{-6}$ seconds, or 8,000 microseconds.   According to Equation (866) the tangent of the corresponding cut-out angle is -6.04, so that cut-out occurs 9.4 degrees beyond the crest of the sine loop.

The behavior of the load voltage after cut-out can be estimated from the fact that the time constant of the discharge circuit is the time required for the condenser to discharge to 36.8 per cent of its initial voltage. Cut-in invariably occurs in less than one-half cycle after cut-out, for obvious reasons. In Fig. 187 the time constant is about one half-cycle, so that cut-in must occur at appreciably more than 36.8 per cent of the cut-out voltage. The exact point at which cut-in occurs can if desired be determined by making use of the fact that the charge that enters the condenser between cut-in and cut-out must equal the total charge that passes through the load between cut-out and cut-in.

It has already been pointed out that at no load there is no ripple, and that $E_{DC} = E_m$ at no load. As load current is drawn in increasingly larger amounts, cut-out occurs progressively farther down from the sine-loop crest. Therefore the valleys between crests (Fig. 187) become more pronounced. As a result the average voltage at the load falls off rapidly, and the ripple increases markedly. As the load current becomes very large, the voltage across the load approaches the sine loop pattern. Thus at heavy loads a filter consisting of a condenser only has little or no filtering action. The voltage pattern ultimately approached as the load current is increased indefinitely is in fact just the same as that at no load in the condenserless circuit of the previous section.

Even with a large condenser, the chokeless circuit of the present section accomplishes low-ripple filtering of a substantial load current only by drawing from the power supply a very large current for a small fraction of each half-cycle. Such short-time high-crest currents require thermionic rectifying elements to have high crest ratings, and represent an undesirable type of alternating-current load.

The essential features of the behavior of a filter consisting of a condenser only are that (a) it reduces the ripple satisfactorily only at no load or very small load currents, and (b), the direct-current voltage at the load is, at no load, the crest voltage of the sine loop, but decreases rapidly as the load current rises.

278. Single–Phase Filter Circuit Behavior: (III) Choke and Condenser Both Present. The complete circuit of Fig. 185, with its choke input filter, in a general way combines the properties of the circuits that have been discussed in the two preceding sections. The essential features of its behavior are that:

(a) The ripple is small at full load, and zero at no load, but is a maximum at some intermediate load.

(b) The direct-current voltage E_{DC} at the load has the value E_m at no load. As illustrated in Fig. 188, E_{DC} falls off rapidly from this no load value, as the load first increases above zero, then levels off to a nearly constant value of about $0.636 E_m$ volts.

The function of the bleeder resistance is to shift the no load point from A to a point slightly to the left of B, Fig. 188, by making the filter pass some current even when the external load current is zero. Although the use of a bleeder resistance introduces a constant power loss, it results in a marked improvement in the voltage regulation.

The reason for the steep light-load variation is that at light loads the condenser is the important filter element, so that the voltage drops off rapidly as the load rises above zero, just as in Section 277. For the same
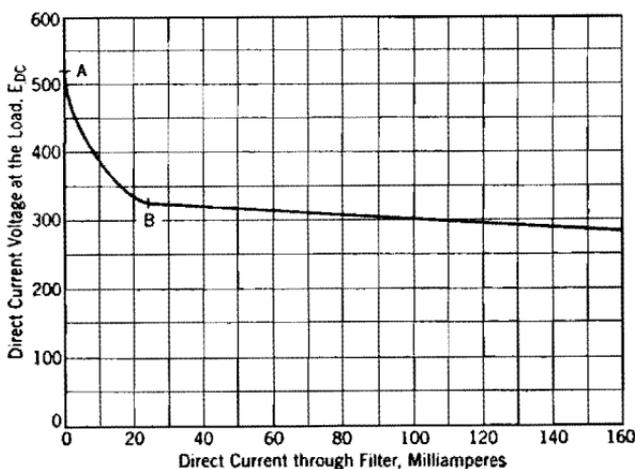
Fig. 188. Experimentally observed volt-ampere curve for a rectifier with choke-input filter, circuit diagram as in Fig. 185, $I_B = 24$ milliamperes, E_{DC} at $B = 328$ volts.

reason, cut-out and cut-in occur for loads to the left of B. At loads to the right of B the choke is the important filter element, so that the direct-current load voltage is nearly constant at about $0.636 \, E_m$ as in Section 276, and cut-out and cut-in do not occur.

The condenser helps materially to reduce the ripple at heavy loads, and the choke to reduce it at light loads, so that the joint filtering action is at all loads much more satisfactory than with either choke or condenser alone, especially when the cost of apparatus and space occupied are considered. The ripple is greatest for loads near the point B.

279. Selection of Circuit Constants for a Choke Input Full-Wave Single-Phase Filter. Suppose that a set of rectifier and filter specifications include a statement of the full load current and voltage at the load, and a statement of an upper limit to the per cent ripple at the worst point, that is, at B, Fig. 188. The problem of meeting these requirements may be approached in a variety of ways, but for illustrative purposes it is satisfactory to start by assuming willingness to locate the

bend in the voltage curve (point B), at some definite fraction of the total direct current through the filter at full load.   This amounts to a specifica-
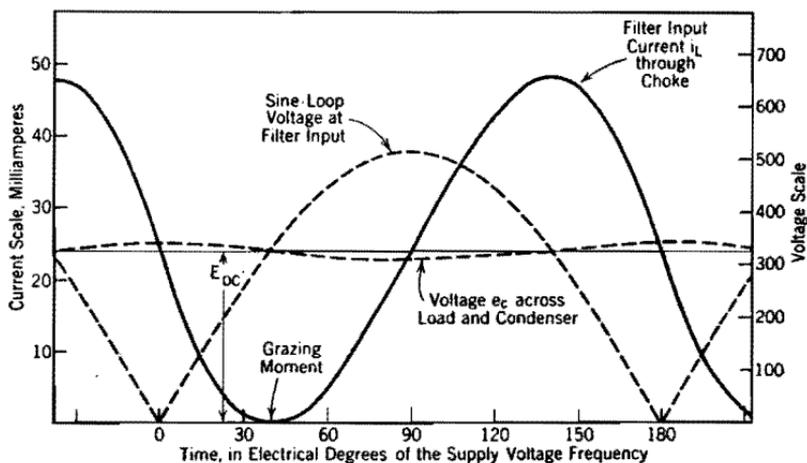
FIG. 189.   Voltage and current waves corresponding to point B, Fig. 188, with L and C chosen to give 9.4 per cent ripple at that load. $X_L/X_C = 3.8$.

tion of the value of the load resistance R_B at the point B, so that R_B will be treated as a known quantity.

The essential feature of the point B is that cut-out and cut-in occur

FIG. 190.   Circuit diagram for use in analyzing circuit behavior at [point B, Fig. 188, for the circuit of Fig. 185. R_B is shown dotted because the alternating current [through it is neglected in the circuit analysis.

for loads to the left of it, but not for loads to the right of it.   When cut-out is not occurring, as at loads to the right of B, the voltage at filter input terminals is simply the sine loop voltage, Equation (854). Therefore at and to the right of B, the magnitudes of the direct current and of the various alternating components of current, in all parts of the circuit, can be calculated from Equation (854) by the use of the circuit impedances for the various frequencies that appear in that equation.   Such treatment gives incorrect results for loads corresponding to points to the left of B, because cut-out makes the filter input voltage differ from the sine loop pattern, so that Equation (854) no longer represents the filter input voltage.

For loads to the left of B the filter input current, which is also the cur-

rent i_L in the choke, is zero during part of each half-cycle. For loads to the right of B the filter input current i_L never goes to zero, because one or the other of the rectifying elements is at all times passing forward current. For just the load B the filter input current i_L barely grazes the zero-current axis, as illustrated in Fig. 189.

Fig. 190 is a filter circuit diagram in form convenient for analysis of circuit conditions at point B. As indicated by its wave form in Fig. 189, i_L at B contains very pronounced alternating components at harmonics of the supply-voltage frequency. The ripple at the load is made small for load point B by passing the harmonic components of i_L through the condenser. In order to accomplish this, X_C must be much less than R_B; therefore the harmonic components of current in R_B must be small, and those in X_C practically the same as in X_L. For that reason R_B is shown dotted in Fig. 190, and will be neglected in the analysis of the harmonic currents and voltages for load as at B.

If X_L and X_C stand respectively for the reactances of choke and condenser at the supply frequency, the reactance X_n of the circuit of Fig. 190 to the nth harmonic is

$$X_n = nX_L - \frac{X_C}{n} = X_C\left(n^2 \frac{X_L}{X_C} - 1\right) \tag{866 p}$$

The resistance R_F of the choke will be neglected, as it is invariably small relative to R_B. The direct-current resistance of the circuit is therefore R_B at load point B.

Circuit analysis for load point B may be carried out by studying the behavior of the circuit of Fig. 190 when the sine loop voltage, Equation (854), is applied across the terminals. Voltages divide in proportion to reactances, so that the crest value of the nth harmonic voltage across X_C is

$$\frac{X_C}{X_n} \times \begin{Bmatrix} n\text{th harmonic component} \\ \text{of sine loop voltage} \end{Bmatrix} \tag{867 p}$$

The condenser voltage is 180° out of phase with the overall voltage, so that in an expression for the instantaneous load and condenser voltage, e_c, the negative cosine terms of Equation (854) become positive cosine terms, besides being multiplied by X_C/X_n. Thus, using Equation (866)

$$e_c = \frac{2E_m}{\pi} + \frac{2E_m}{\pi}\left[\frac{2}{3}\frac{\cos 2\omega t}{\left(4\frac{X_L}{X_C} - 1\right)} + \frac{2}{15}\frac{\cos 4\omega t}{\left(16\frac{X_L}{X_C} - 1\right)} + \cdots\right.$$
$$\left. + \frac{2}{(n^2 - 1)}\frac{\cos n\omega t}{\left(n^2\frac{X_L}{X_C} - 1\right)} + \cdots\right] \tag{868 p}$$

In this equation n has even values only.   The e_c voltage in Fig. 189 was plotted from Equation (868), using $X_L/X_C = 3.8$.

Of the two terms in Equation (868), the first is E_{DC} at the load, the second (bracket term) is the ripple voltage across load and condenser.
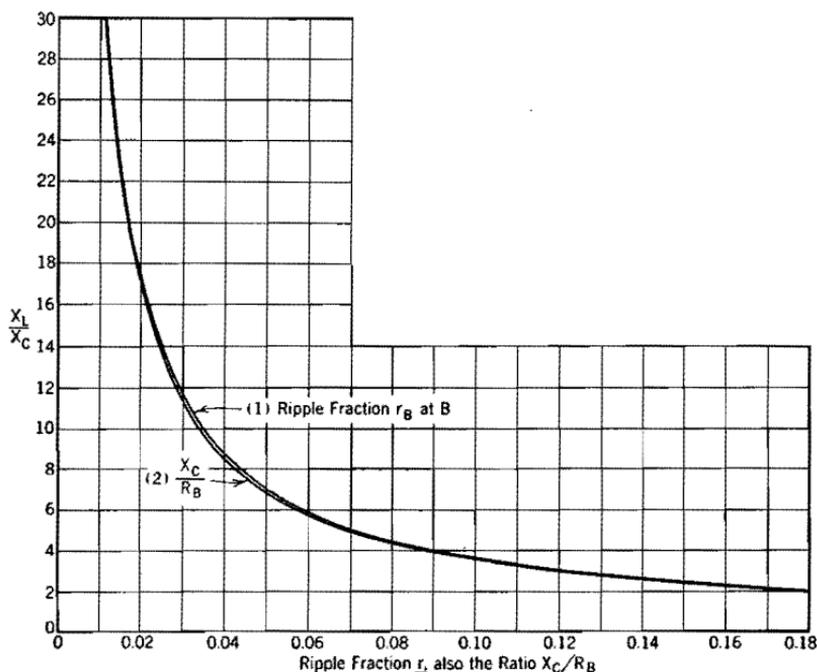
Fig. 191.   Graph of the relations between:
  (1) X_L/X_C and the ripple fraction r at B, Fig. 188, in accordance with Equation (871),
  (2) X_L/X_C and X_C/R_B, in accordance with Equation (875), all for the circuit of Fig. 185. A curve for Equation (876) would lie about midway between (1) and (2).

In order to evaluate the ripple fraction r, defined by Equation (852), the maximum and minimum values of e_c must be known. At the maximum

$$2\omega t = 0$$
$$\cos 2\omega t = \cos 4\omega t = \cos 6\omega t = \cos 8\omega t = \text{etc.} = +1 \qquad (869)$$

At the minimum

$$2\omega t = 180°$$
$$\cos 2\omega t = \cos\ 6\omega t = \cos 10\omega t = \text{etc.} = -1 \qquad (870)$$
$$\cos 4\omega t = \cos 10\omega t = \cos 12\omega t = \text{etc.} = +1$$

In an expression for the ripple swing, $e_{c\,\text{max}} - e_{c\,\text{min}}$, the terms in $\cos 2\omega t$, $\cos 6\omega t$, $\cos 10\omega t$, etc., have twice their values in Equation (868), and

terms in $\cos 4\omega t$, $\cos 8\omega t$, $\cos 12\omega t$, etc., cancel out. The mathematical expression for the ripple fraction r_B at point B (also for r at any load point for which the alternating current in R_L, Fig. 190, can safely be neglected in the circuit analysis) becomes

$$r_B = \frac{4}{3}\frac{1}{\left(4\dfrac{X_L}{X_C} - 1\right)} + \frac{4}{35}\frac{1}{\left(36\dfrac{X_L}{X_C} - 1\right)} + \frac{4}{99}\frac{1}{\left(100\dfrac{X_L}{X_C} - 1\right)} + \cdots \quad (871\ \text{p})$$

The upper curve in Fig. 191 is a graphical representation of the relation described by Equation (871).

Thus a specification of the maximum ripple fraction, which is that at load point B, is equivalent to specifying the ratio X_L/X_C, because that ratio is related to r_B in accordance with Equation (871) and Fig. 191.

The derivation, in the following paragraphs, of a further relationship between X_L, X_C, and the load resistance R_B at B will be based on the fact that at B the filter current i_L must just graze the zero current axis, as illustrated in Fig. 189. Equation (872) below is a mathematical expression for i_L at B that corresponds to Fig. 190, the alternating current in R_L being neglected, and to the fact that at B the applied voltage is as given by Equation (854). The crest values for the various terms are obtained by dividing the individual terms of the sine loop voltage equation by the proper X_n from Equation (866). As the circuit of Fig. 190 is entirely reactive to the alternating components of current, all of them lag their respective voltages by $90°$. Cosine terms in the voltage equation therefore appear as sine terms in the current equation, which is

$$i_L = \frac{2}{\pi}\frac{E_m}{R_B}\left\{ 1 - \frac{1}{\dfrac{X_C}{R_B}}\left[\frac{2}{3}\frac{2\sin 2\omega t}{\left(4\dfrac{X_L}{X_C} - 1\right)} + \frac{2}{15}\frac{4\sin 4\omega t}{\left(16\dfrac{X_L}{X_C} - 1\right)} + \cdots \right.\right.$$
$$\left.\left. + \frac{2}{(n^2 - 1)}\frac{n\cos n\omega t}{\left(n^2\dfrac{X_L}{X_C} - 1\right)} + \cdots \right] \right\} \quad (872\ \text{p})$$

Only even values of n occur in this equation.

It is convenient to abbreviate Equation (872) by using the symbol Φ for the dimensionless quantity in the square brackets, as follows:

$$i_L = \frac{2}{\pi}\frac{E_m}{R_B}\left\{ 1 - \frac{\Phi}{X_C/R_B} \right\} \quad (873\ \text{p})$$

At the moment when i_L grazes the zero axis (Fig. 189), (a) $i_L = 0$, and (b) the harmonic components add up to their maximum instantane-

ous total value, so that $\Phi = \Phi_{max}$. Therefore, at the grazing moment

$$\frac{X_C}{R_B} = \Phi_{max} \qquad\qquad (874 \text{ p})$$

Computations show that, for all values of X_L/X_C between 3 and 30, Φ_{max} occurs when $2\omega t$ is about $81°$. Therefore, in general X_C/R_B and X_L/X_C are related as follows, for values of X_L/X_C between 3 and 30:

$$\frac{X_C}{R_B} = \frac{2}{3}\frac{2\sin 81°}{\left(4\dfrac{X_L}{X_C} - 1\right)} + \frac{2}{15}\frac{4\sin 162°}{\left(16\dfrac{X_L}{X_C} - 1\right)} + \cdots \qquad (875 \text{ p})$$

Only even values of n are to be used in this equation. The lower curve in Fig. 191 is a graphical representation of the relation described by Equation (875).

Thus in general, if r_B is specified, X_L/X_C can be read directly from the upper curve in Fig. 191, and X_C/R_B subsequently from the lower curve at the selected value of X_L/X_C. However, the two curves are so close together that for most practical purposes r_B can be considered equal to X_C/R_B. Furthermore, since all subsequent terms in Equation (871) are small relative to the first one (the second-harmonic term) a very good approximation to the relations between r_B, X_L/X_C and X_C/R_B can be stated as follows:

$$\frac{X_C}{R_B} \simeq r_B \simeq \frac{4}{3}\frac{1}{\left(4\dfrac{X_L}{X_C} - 1\right)} \qquad\qquad (876 \text{ p})$$

A curve representing Equation (876) graphically would lie about midway between the two curves that appear in Fig. 191, so very close to both of them. Thus equation (876) is a very excellent approximation to the complete relations. It can be derived directly by neglecting all but the second-harmonic terms in Equations (871) and (872), then proceeding just as above. Thus for practical purposes *only the second-harmonic ripple is of consequence*, and r_B'', the second-harmonic ripple fraction at B is practically identical with r_B.

The relations that have just been derived have been based on the assumption that at load point B, $X_C \ll R_B$, and $R_F \ll R_B$. (R_F is the resistance of the choke.) Because the currents in X_C and R_B, Fig. 190, have a $90°$ phase difference, X_C can be as much as 15 to 20 per cent of R_B without seriously affecting the validity of the derivation. Corrections to take care of the most important effects of the existence of an appreciable R_F can be made by using $R_F + R_L$ instead of R_L in

Equations (872) to (875), and using $E_m \dfrac{R_L}{R_F + R_L}$ instead of E_m in Equation (868). These changes take care of the effect of R_F on the magnitude of the direct current. The alternating-current drop due to R_F is of no consequence because it occurs vectorially at right angles with that due to nX_L, and $R_F \ll nX_L$.

The overall relation between direct current through the filter and the direct-current voltage at the load is illustrated in Fig. 188. In that figure $E_{DC} = E_m$ when the direct current is zero, and $E_{DC} = 0.636\ E_m \dfrac{R_L}{R_F + R_L}$ for direct currents greater than that at B. At B, of course $R_L = R_B$. The no load point can be placed wherever desired between A and B by proper selection of the bleeder resistance. The voltage regulation depends very largely on the location of the no load point.

280. Repeating Transients. Although alternating-current voltages and currents vary sinusoidally, the nature of the variation is the same from cycle to cycle; that is, it is "steady" in nature. Hence such voltages and currents are said to exhibit *steady state* variations.

When the switch is first closed into an alternating-current circuit, or for that matter a direct-current circuit, there exists initially a set of *transient* voltages and currents.[SS] Transients ultimately die out, but the steady state currents and voltages persist indefinitely as long as the circuit constants and the nature of the applied voltage are unchanged.

The mathematical prediction of electrical transients is ordinarily accomplished by solving differential equations for currents and voltages, the differential equations being set up in accordance with Kirchhoff's circuit laws. In the process of solving any differential equation, constants of integration must be evaluated so as to satisfy the *boundary conditions* of the problem. In the theory of electrical transients the boundary conditions are usually the voltages, currents, and sometimes the rates of change of voltages and currents, that exist at the moment the transient is initiated by closing a switch.

It has been pointed out earlier that mercury-vapor and similar rectifiers are essentially switches which open and close regularly at periodic intervals. Each opening and closing of the circuit by a rectifying element, as for example at cut-out and cut-in of the preceding section, initiates a new electrical transient. The boundary conditions that must be satisfied by the solutions of the differential equations for alternating-current circuits containing rectifying circuit elements are that all voltages, currents, and rates of change of both voltage and current must begin each new cycle in just such a way that the next cycle will

begin identically. That is, the beginning and end conditions of the transients must reproduce one another periodically.

The voltage and current variations that take place in such circuits are identical in successive cycles. They may therefore be called *repeating transients* of voltage and current, because they do repeat themselves periodically, yet can be predicted by the methods used in predicting transients. However, repeating transients may also be predicted by the methods used in the analysis of steady state problems, because the fact that the voltage and current variations are identical in successive cycles satisfies the requirements for the existence of a steady state.

Thus electric circuits in which repeating transients of voltage and current exist can in general be solved either

(1) By the solution of circuit differential equations subject to mutually reproducing initial and final boundary conditions,[SS] or

(2) By the use of Fourier series for voltages and currents, and of proper impedances for the various harmonics.[119] This method is often very satisfactory even when it involves trial and error procedure.

The first method is probably the most satisfactory one for the circuit of Section 277, which depends altogether on a condenser for filtering action, while the second is by far the most satisfactory for the circuit of Section 276. In some problems the two methods can be successfully mixed, one method being used for one part, the other for other parts, of each cycle.

281. Phase-Shift Control of Thyratrons. Fig. 192a is a diagram of a circuit that makes use of a thyratron to deliver rectified current of easily adjustable magnitude to a load of resistance R_L. The magnitude of the load current is adjusted by controlling the phase position at which the grid of the rectifying element swings positive and closes the load circuit.[111, 112] Variation of phase position of grid-circuit voltage is accomplished by changing the magnitude of the variable resistance R. The heavy line in Fig. 193a illustrates a typical wave form of the resulting current in the load, if no filter is used.

A filter can be introduced between rectifying element and load if desired; the wave forms of filter and load currents then both differ markedly from that shown in Fig. 193a, but the average value of load current is not materially changed, if the filter constants are properly chosen. A condenser input filter cannot be used, for the abrupt application of circuit voltage to the condenser would pass such a large charging current as seriously to damage the cathode surface of the rectifying element.

The abrupt rise at the beginning of each current wave occurs when the grid swings positive. It is evident that if this occurs earlier in each

cycle than shown in the figure, the average load current must be increased, whereas if it occurs later in the cycle, the average load current must be decreased.

Fig. 192b is a vector diagram of the current and voltages in the grid-control circuit AGB, Fig. 192a.   The current I_{RC} in this circuit leads the voltage E_{AB} applied to the grid-control circuit by some phase angle that is dependent on the relative values of R and X_C. I_{RC} of course leads the voltage across the condenser by 90°, and is in phase with the voltage E_{GB} across the resistance R.   The grid-circuit voltage is E_{OG}; Fig. 192b shows that E_{OG} lags behind the load-circuit voltage E_{OB} by some angle ϕ that is dependent on the relative values of R and C.   As the angle AGB must always be 90° (it is the angle between a resistance and

Circuit diagram.   Vector diagram.
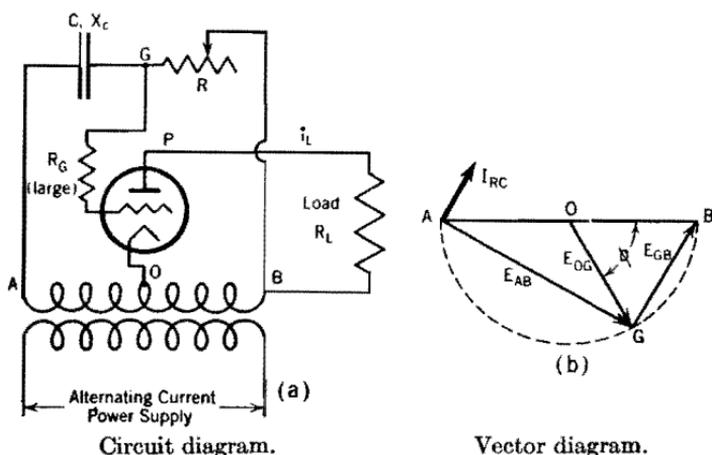
Fig. 192.   Thyratron current control by means of phase shift of grid circuit relative to power supply circuit.   Half-wave operation.

a condenser voltage for the same current), variations in relative values of R and C cause the point G, Fig. 192b, to trace an arc of a semicircle.   The grid voltage E_{OG} has therefore a constant magnitude but a varying angle of lag relative to the load-circuit voltage.

Fig. 193a illustrates the relations between grid voltage and load circuit currents and voltages.   The voltage e_{OG} is applied to the cathode-to-grid path in series with a current-limiting resistor R_G.   During the negative half-cycle of e_{OG}, no current flows between cathode and grid, because the grid is negative and the tube dark.   After e_{OG} becomes positive, the tube contains an arc, and current would flow as freely to the grid in its positive swing as to the plate, if no current-limiting resistor were provided. R_G must be at least large enough to limit grid current to the average value specified in the tube rating sheet.

Satisfactory operation of a similar nature can be obtained by replacing the grid-control-circuit condenser by a resistance, and the resistance by an inductance. The vector diagram is then the same as in Fig. 192b except that the current vector lies along the E_{AG} vector. Variation of the resistance has the opposite effect that it had before on load current, but the same range of currents is obtainable.
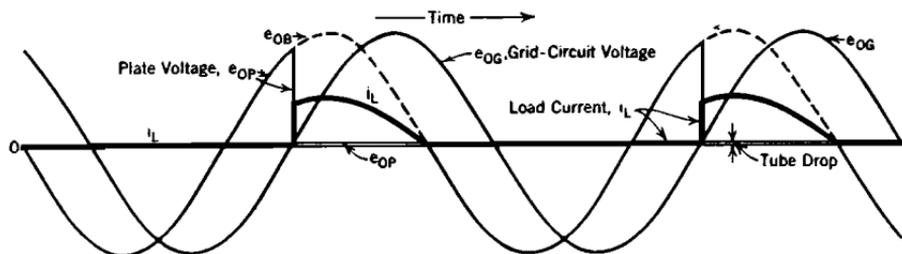
(a)   Typical voltage and current variations in the half-wave phase-shift thyratron circuit of Fig. 192.

(b)   Typical load current wave form for the full-wave rectifying phase-shift circuit of Fig. 195a.
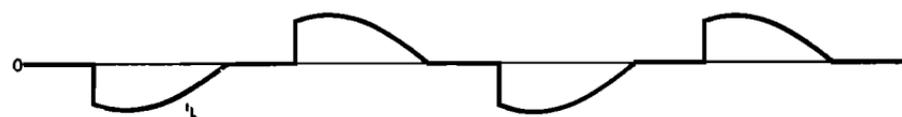
(c)   Typical load current wave form for the full wave non-rectifying phase-shift circuit of Fig. 195b.

FIG. 193.   Voltage and current waves in phase-shift thyratron current control circuits.

Fig. 194 contrasts the behavior of a circuit like that of Fig. 192a with one that is identical except that the load consists of resistance and inductance in series. With both inductive and noninductive loads, current flow begins as soon as the grid swings positive. However, with an inductive load the current grows gradually rather than abruptly, and persists well into the negative half-cycle. As soon as the current reaches zero in its normal transient course, the plate potential drops abruptly to a position along the e_{OB} voltage wave.

The circuit diagrammed in Fig. 192a provides only half-wave operation, that is, current flows through the load during a controllable portion of each alternate half-cycle. Fig. 195a is a diagram of a similar circuit arrangement that provides full-wave operation. The vari-

able resistance R provides control of current magnitude by way of phase-shift of the grid voltages of both tubes.

(a)   With resistance load. (Similar to Fig. 193a.)

(b)   With inductive load.
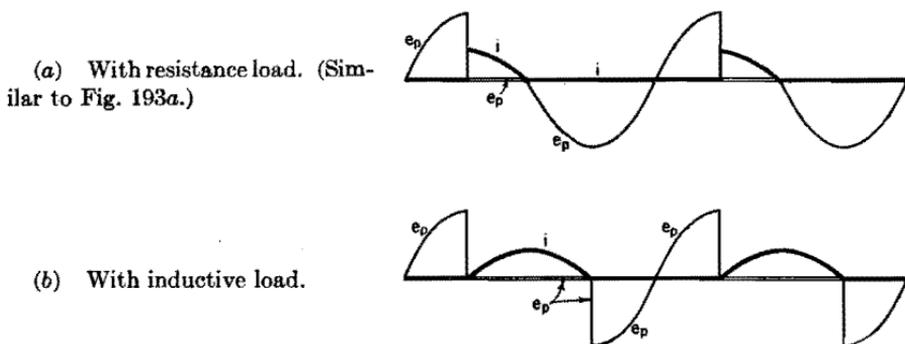
Fig. 194.   Plate voltage and current variations in a thyratron arranged to provide current control by phase shift in the grid circuit.   Arc initiated at the crest of the applied voltage wave.

The type of current control that has been described in this section is sometimes used for dimming theater lights.   In such service the direct

(a)   Direct-current output; see Fig. 193b.

(b)   Alternating-current output; see Fig. 193c.
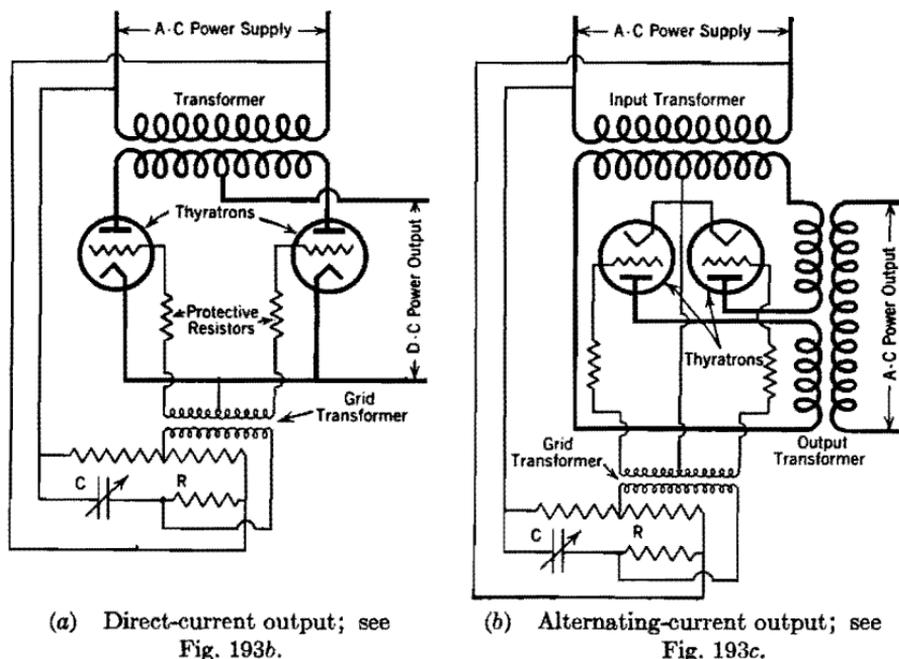
Fig. 195.   Circuit diagrams for full-wave phase-shift current control by means of thyratrons.   Variable condenser in the grid circuit shifts the phase.

current from the electronic current-control circuit passes through a coil of an iron-core reactor, while the alternating current to the lights passes

through another coil on the same reactor. When the direct current is large, the reactor core is magnetically saturated, so that the alternating-current variation in magnetizing force operates along a part of the magnetization curve that has a very gentle slope. The inductance of the core is therefore small (see Section 159) and the lights are bright. If the direct current is small, the alternating-current variation of magnetizing force takes place along a steep part of the magnetization curve, so that the inductance is large, and the lights are dim. Such an arrangement is said to employ a " saturated-core reactor." See Chamber's article.[112]

Grid control of alternating-current magnitude by grid-circuit phase shift is obtainable by the use of the circuit diagrammed in Fig. 195b.[TT 179]

282. Inversion from Direct to Alternating Current. Thyratrons and ignitrons can both be used for *inversion*, that is, for changing direct-current power into alternating-current power. Inversion must in one way or another accomplish the interruption of current flow, and one of the limitations common to thyratrons and ignitrons is that they cannot interrupt current flow. Circuit interruption is accomplished, in inverters, by special circuit arrangements which produce *commutation*, or transfer of current flow from one path to another under the control of grid circuits or igniter rods.

Most inverter circuits fall under one or the other of the three following classes: (I) Single-tube inverters,[120] (II) Parallel-type inverters employing, for single-phase operation, two or more controlled rectifying elements,[121,122] (III) Series-type inverters[123] that employ, in single-phase operation, two or more controlled rectifying elements. The operation of single-phase Class II and III types of inverters will be described briefly in succeeding sections.

Inverters are at present used chiefly for laboratory and testing purposes, as for example in connection with cathode-ray oscillograph sweep circuits and in meter-testing circuits. Inverter circuits have been given considerable study, because it seems at least possible to obtain speed control of induction motors by operating them from inverters, and to transmit large amounts of power at high direct-current voltages, if inversion to alternating-current for distribution at the load end can be successfully accomplished. These applications of inverters are realizable technically, but it has not yet been demonstrated that they will be economically justifiable.[124]

An important feature of all power applications of inverters is that large amounts of capacitance must be provided. The kva rating of an inverter's condenser must ordinarily be of the same order of magnitude as the kva rating of the load itself, unless the load operates at a low leading power factor. Capacitative action in the load serves essen-

tially the same purpose as capacitance in the inverter itself, so that less capacitance is required in an inverter that serves a leading power factor load than in one that serves a unity power factor or lagging power factor load.

283. Parallel-Type Single-Phase Inverters.[121, 122] A typical circuit diagram for a single-phase parallel-type inverter is shown in Fig. 196. The function of the circuit is to receive power from a direct-current source and deliver alternating-current power to the load. The frequency of alternation is that of the grid excitation. A great variety of voltage wave forms can exist at the load, dependent on the magnitudes of circuit constants of inverter and load. Three easily obtained types of load-current wave form are illustrated in Fig. 197, page 529.

The voltages and currents in inverters are typical examples of repeating transients. The rectifying elements change the circuit connections periodically, initiating new transients at each change, usually before the old ones die out. Operation stabilizes at a steady state condition in which the transient in each half-cycle is exactly like that in the preceding half-cycle, but of opposite polarity.

Because of the close magnetic coupling that exists between the transformer's primary and secondary in Fig. 196, the voltage across the condenser is identical at every instant with that across the load, provided the load is noninductive and the transformer turn ratio one-to-one. No matter what the turn ratio may be, the wave form of the voltage across the condenser is the same as that across the load, as long as the load is noninductive.

It is evident from Fig. 197 that the condenser discharges, then charges again to the opposite polarity, in each half-cycle. In fact, the entire circuit is simply an arrangement for producing this periodic rise and fall of condenser potential, with sufficient positiveness to force like variations in the load voltage.

The following paragraphs contain an explanation of the mechanism of commutation in the circuit diagrammed in Fig. 196. In this explanation all potentials will be expressed relative to the negative direct-current terminal, and the tube drop will be neglected.

Consider the full-load resistive case, Fig. 197b. Between moments A and B tube No. 1 is conducting, and tube No. 2 is dark, holding the circuit open by virtue of a negative grid potential. Considerably before the moment B is reached, the condenser has become charged to about twice the direct-current line voltage. It rises to twice rather than just to line voltage because the center-tapped primary transformer winding must have equal voltages in its two halves, and the voltage across one of them is the line voltage, except for the drop in the series inductance.

This drop is usually not large at the end of a half-cycle of operation of the type assumed for Fig. 197b.

At the moment B the grid of tube No. 2 swings positive; this has the same effect as closing a switch between No. 2 tube's plate and cathode. Prior to this switching operation, No. 1 tube's plate potential e_{p_1} was about twice the direct-current line voltage, for example, 400 volts if the direct-current supply is constant at 200 volts.   The switching operation ties the plate of No. 2 tube solidly to the negative direct-current terminal, so that e_{p_2} must become abruptly zero.   However, the condenser's potential difference cannot change instantaneously, and must remain for
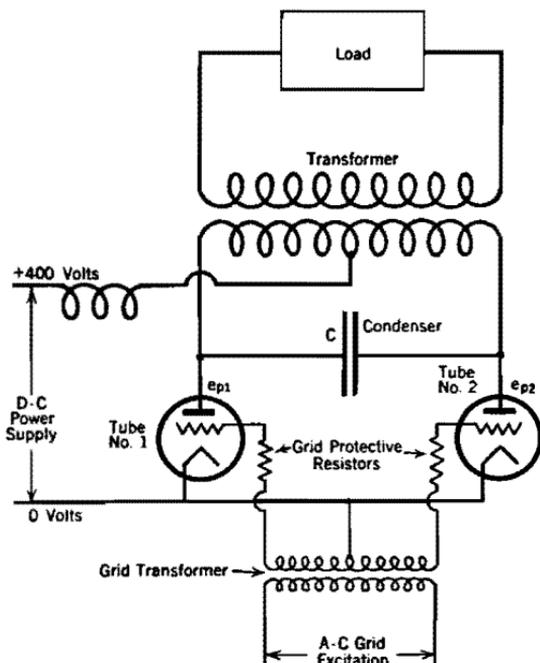
FIG. 196.   Typical parallel-type inverter circuit employing thyratrons.

the moment at about 400 volts.   The potential of No. 1 tube's plate therefore drops abruptly to about -400 volts.   Of course current flow through the tube stops at once, because current cannot pass in the inverse direction.   Current flow has been transferred from No. 1 tube to No. 2 tube by the commutating action of the circuit.

After commutation the condenser rapidly discharges around through the load circuit (to which it is coupled by the transformer) and is brought to a 400-volt charge in the negative direction.   At the end of the half-cycle the grid of No. 1 tube swings positive, closing the circuit at that tube and forcing e_{p_2} to -400 volts.   Each tube's switching action opens the circuit at the other tube.

Resistance load, light load conditions. (R_L large.)

Resistance load, heavy load conditions. (R_L small.)

Partially inductive load.
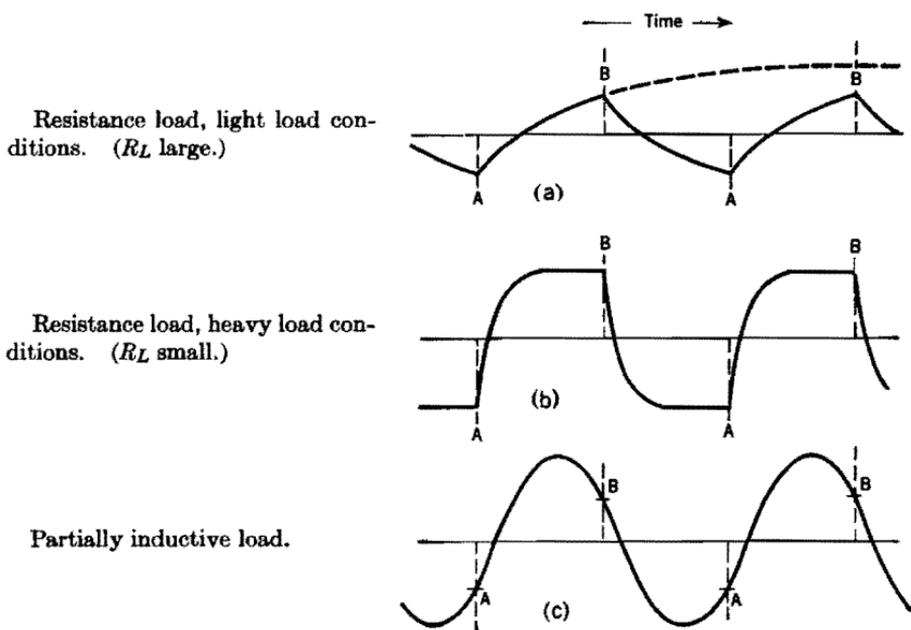
FIG. 197. Wave forms of load and condenser voltage, parallel-type inverter, circuit as in Fig. 196. At moment A No. 1 tube lights, No. 2 tube goes out; at moment B the reverse happens.

Resistance load, light load condition, as for Fig. 197a.

Resistance load, heavy load condition, as for Fig. 197b.

Partially inductive load, as for Fig. 197c.
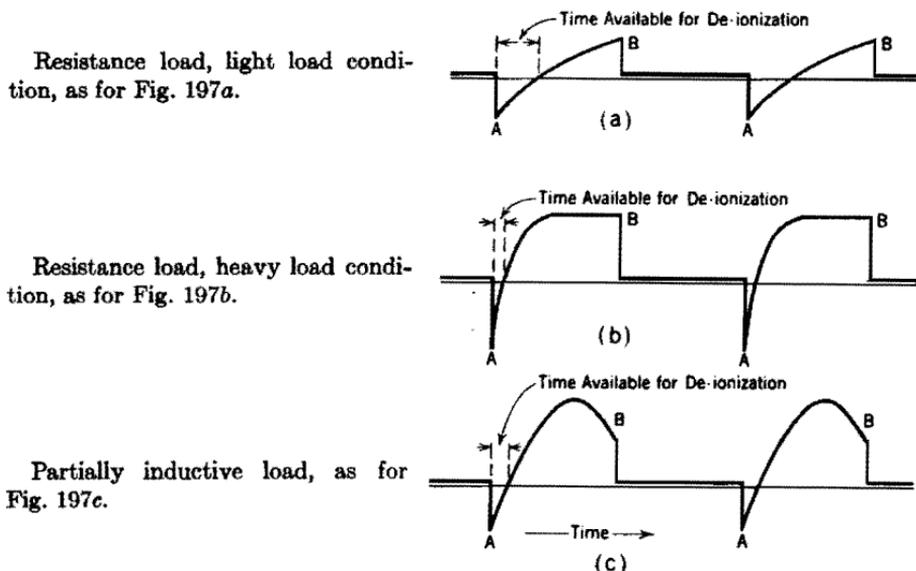
FIG. 198. Plate voltage of Tube No. 2, in the parallel-type inverter circuit of Fig. 196, for the various load conditions illustrated in Fig. 197.

If any one of the cyclic transients is for any reason not interrupted by the commutating action, the condenser slowly discharges again to zero potential, and the current in the conducting tube or tubes rises to a very large short-circuit value that is limited only by the transformer, line, and generator resistances.

A time constant $R_L C$, in which R_L is the load resistance and C the inverter's capacitance, is the primary factor in determining the rate of decay of condenser potential after commutation, so also the wave form of the voltage across the load. For example, for Fig. 197a the load current is very small, because the load resistance is large. The time constant is therefore much larger, and the discharge less rapid than in the full load condition, Fig. 197b. The slow condenser discharge at light loads tends to produce a triangular wave form, and the rapid discharge at heavy loads tends to produce a rectangular wave form.

If the load contains an important inductive element, the discharge and subsequent charge of the condenser tends to be oscillatory; instead of a time constant $R_L C$, the natural period of oscillation becomes important. For a discussion of operation into an inductive load the reader is referred to a recent article by C. F. Wagner.[122]

284. Failure of Commutation in Inverters. Commutation, or transfer of current from one tube to another, takes place when the plate of a conducting tube is suddenly driven negative. Commutation may fail to occur as it should, either: (1) because at the moment of expected commutation there is not enough charge on the condenser to drive the plate of the conducting tube negative; or (2) because the plate does not stay negative for a long enough time to permit the grid to regain electrostatic control over arc initiation, so that an arc re-forms after only momentary extinction.

The negative-plate condition in general persists only for a limited time. During this time the ions of the plasma must disappear sufficiently to permit a negative grid to keep electrons out of the field between grid and anode. The time required for the grid to regain control is called the *de-ionization time* of the tube. The de-ionization time of a thyratron depends on electrode and envelope design, and is likely to be between 100 and 1,000 microseconds.[111, 121, 122, 123]

Each set of inverter and load circuit constants keeps the plate negative for a certain definite time after driving it there; the interval during which the plate remains negative is the *time available for de-ionization*. If the de-ionization time of the tube is less than the time available for de-ionization, commutation is successfully accomplished, and the inverter continues to function properly. If at the end of any half-cycle the circuit allows too short a time, or the tube requires too long a time, for

de-ionization, commutation does not occur.   Failure to commutate properly usually results in a short-circuit on the direct-current power supply.

The time available for de-ionization is illustrated in Fig. 198 as the time between the moment No. 2 tube's plate potential is driven negative, killing the arc, until the plate potential again becomes positive. A comparison of Figs. 198a and 198b indicates that a large time-constant produces a long " time available for de-ionization," while a small time-constant produces a small time available for de-ionization.   For a discussion of the time available for de-ionization when operating into an inductive load the reader is referred to Wagner's article.[122]

285. Series-Type Single-Phase Inverters.[123]   Fig. 199 is a circuit diagram for one form of single-phase series-type inverter.   The operation of this circuit is much more easily analyzed mathematically than is that of the parallel-type inverter.   The physical arrangement of parts is a little more complicated in the series than in the parallel setup because the two cathodes are not at the same potential.   The cathode heating and grid excitation circuits for tube No. 1 must be electrically insulated from those for tube No. 2.

The operating possibilities and limitations exhibited by series-type inverters present distinct contrasts with those for the parallel type.

The natural frequency of the circuit illustrated in Fig. 199, when the load is a resistance, is given by the expression

$$f = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R_L^2}{4L^2}} \qquad (877 \text{ p})$$

In this equation R_L is the load resistance, and C and L the capacitance and inductance of the two condensers and of half the choke respectively. If this f is the same as the grid excitation frequency, the load voltage and current are practically sinusoidal, as illustrated in Fig. 200a.   The corresponding variations in potential differences of the two condensers are as shown in Fig. 200b.   As only load current can charge and discharge the condensers, each must have a maximum or minimum potential at the moment of current reversal.   Furthermore, because the condensers are in series across the direct-current line,

$$e_{C_1} + e_{C_2} = E_{DC} \qquad (878)$$

In the operation illustrated by Fig. 200 the resonant and grid frequencies are the same, so that the grids must swing positive at half-cycle intervals of the circuit's free oscillation.   Also, current reversal in the load must coincide with the moment of commutation, because the tubes carry current alternately and unidirectionally.   Therefore at moment

B, Fig. 200*a*, the grid of No. 2 tube must swing positive.   How does this cause No. 1 tube to be extinguished?
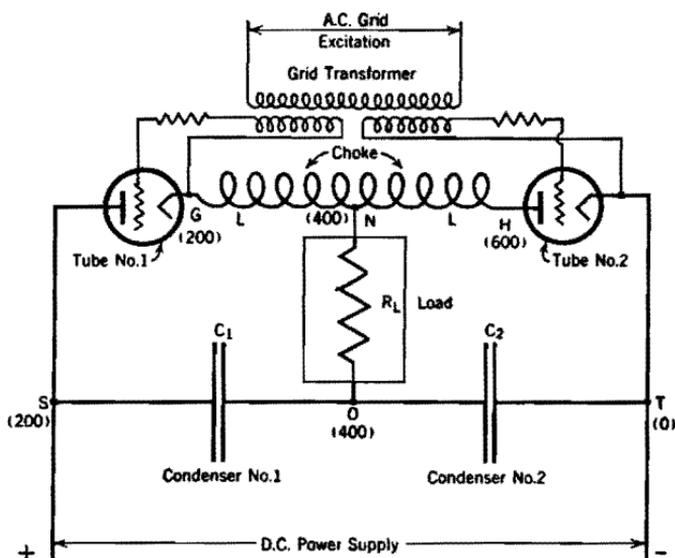
Fig. 199.   Typical series-type inverter employing thyratrons.

The potentials which various points on the circuit have just before the moment *B*, Fig. 200, are marked in parentheses on Fig. 199. Point *O* has a potential of $+400$ volts as moment *B* is approached,
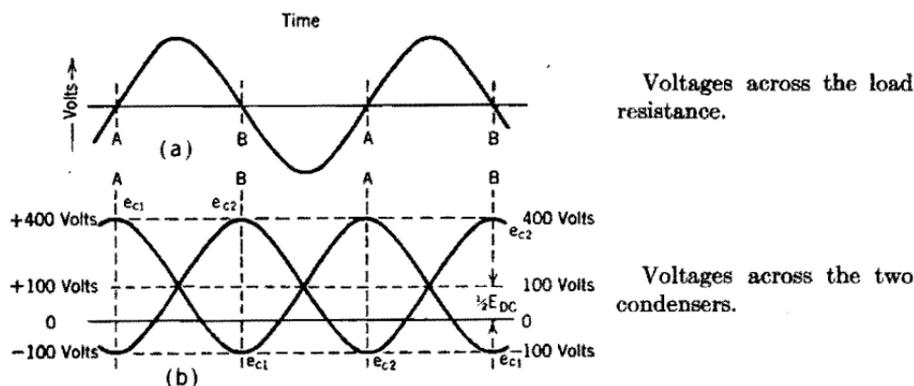
Fig. 200.   Voltage wave forms in the series inverter circuit of Fig. 199 when natural frequency of inverter and load is the same as the grid excitation frequency. At the moment *A*, tube No. 1 lights, tube No. 2 goes out.   At the moment *B*, the reverse happens.

because that is the condenser's charge (twice the applied charging potential in a completed, nearly undamped free oscillation); point *N* is at the same potential as *O* because at that moment the load current

is zero. Point G is at $+200$ volts because the tube's arc still ties it to the positive line terminal, as commutation has not yet occurred.

At moment B No. 2 tube switches the right-hand end H of the choke solidly to zero potential. The potential of point N cannot change instantaneously, because neither the condenser potential nor the zero value of the load current can abruptly acquire new values, the latter limitation existing because any load current must pass through the choke's inductance. Hence a potential difference of 400 volts appears abruptly across the right half of the choke. But this is on the same iron core as the left half, which must therefore also exhibit a 400-volt potential difference, and of the same polarity. Therefore point G is pushed up to $+800$ volts, 600 volts above No. 1 tube's plate; of course the arc in No. 1 tube is immediately extinguished. About a quarter of a cycle is subsequently available for de-ionization. Similar circuit behavior extinguishes the arc in No. 2 tube a half-cycle later.

A tendency toward failure of commutation at light loads exists, partly because of the fact that the load voltage at light loads may have a considerable instantaneous value at the end of each half-cycle. The primary reason for this is that the load resistance at light loads is likely to be large enough to reduce materially the natural oscillatory frequency [Equation (872)], so that each half-cycle of operation includes only a portion of the crest of the free oscillation.

To illustrate the effect of the existence of a load voltage at the end of a half-cycle, suppose that the potential across condenser No. 2 is $+225$ volts, and that across the load 150 volts, just prior to the moment B at which commutation is expected to occur. Then point N has the potential $225 + 150 = 375$ volts just prior to the moment B, and the potential $225 - 150 = 75$ volts just after the moment B. The reason for the reversal of sign of the voltage across the load is that reversal of load current occurs simultaneously with commutation, and the voltage across a resistance load reverses simultaneously with the load current.

The total commutating potential across the choke is twice the potential of point N after the moment B; in the example stated, 150 volts. But this is less than the direct-current line voltage. Therefore point G is not driven above the plate potential of No. 1 tube, so that the arc in No. 1 tube is not extinguished. Commutation has failed, and there is a short-circuit on the direct-current line through the two tubes and the choke.

If the potential across the load were to be initially 100 instead of 150 volts, and the potential across condenser No. 2 still $+225$ volts, commutation would not fail.

A mathematical expression of conditions necessary for commutation can be arrived at by generalizing the reasoning just pursued. Let

e_L symbolize the load voltage. Then point N comes up to moment B with a potential of

$$E_N = e_{C_2} + e_L \tag{879}$$

and starts away from it with a potential of

$$E_N = e_{C_1}^- - e_L \tag{880}$$

This latter quantity must satisfy the relation

$$2(e_{C_2} - e_D) > E_{DC} \tag{881}$$

if commutation is to take place. By the use of Equation (878) this same limitation can be expressed as follows:

$$(e_{C_2} - e_{C_1}) > 2e_L \tag{882}$$

This relation can easily be checked against the example given above, by noting that in that example $e_{C_1} = -25$ volts.

Equations (881) and (882) are perfectly general, holding equally well for resistive, inductive, and capacitative loads. Care must be taken in working with them to use for e_L the load voltage at N immediately *after*, rather than before, the moment of commutation.

PROBLEMS

CHAPTER XXII

1. Select the circuit constants, including transformer turn ratio and resistance of the choke, for a full-wave, choke input filter like that of Fig. 185, subject to the following requirements:

Full load current and voltage at the load 5.0 amperes and 500 volts.

$R_1 = R_2 = 1000$ ohms

220 volts
D.C.

R_1 R_2

No. 1 $C = 1$ mf.   No. 2

Thyratrons, tube drop
20 volts when conducting

FIG. G. Counting circuit using the inverter principle.

Maximum ripple to be 5 per cent, and to occur at 0.5 ampere.

Voltage change between point of maximum ripple and full load to be 25 volts.

Voltage change between no load and point of maximum ripple to be 25 volts.

2. For the circuit designed in Problem 1, find the per cent ripple at full load.

3. Refer to the circuit in Fig. G, containing two thyratrons.

(a) when the switch S is closed both grids are negative, so that no current flows. After a short time the operator makes grid 1 momentarily positive. State the currents in R_1 and R_2, and the potential across the condenser, after steady conditions exist. (b) Now the grid of tube 2 is made momentarily positive. State the potential, relative to the negative line wire, of the respective plates immediately after tube 2 becomes conducting. (c) How long after tube 2 becomes conducting will the potential across the condenser pass through zero?

TABLE I

(See Section 54)

DEPENDENCE OF β^2 ON r/r_c, IN EXPRESSIONS FOR SPACE-CHARGE-LIMITED CURRENTS FROM CYLINDRICAL CATHODES [61]

β^2 applies where $r > r_c$,
$-\beta^2$ applies where $r < r_c$

r/r_c or r_c/r	β^2	$-\beta^2$	r/r_c or r_c/r	β^2	$-\beta^2$
1.00	0.0000	0.0000	3.8	0.6420	5.3795
1.01	0.00010	0.00010	4.0	0.6671	6.0601
1.02	0.00039	0.00040	4.2	0.6902	6.7705
1.04	0.00149	0.00159	4.4	0.7115	7.5096
1.06	0.00324	0.00356	4.6	0.7313	8.2763
1.08	0.00557	0.00630	4.8	0.7496	9.0696
1.10	0.00842	0.00980	5.0	0.7666	9.887
1.15	0.01747	0.02186	5.2	0.7825	10.733
1.20	0.02815	0.03849	5.4	0.7973	11.601
1.30	0.05589	0.08504	5.6	0.8111	12.493
1.40	0.08672	0.14856	5.8	0.8241	13.407
1.50	0.11934	0.2282	6.0	0.8362	14.343
1.60	0.1525	0.3233	6.5	0.8635	16.777
1.70	0.1854	0.4332	7.0	0.8870	19.337
1.80	0.2177	0.5572	7.5	0.9074	22.015
1.90	0.2491	0.6947	8.0	0.9253	24.805
2.0	0.2793	0.8454	8.5	0.9410	27.701
2.1	0.3083	1.0086	9.0	0.9548	30.698
2.2	0.3361	1.1840	9.5	0.9672	33.791
2.3	0.3626	1.3712	10.0	0.9782	36.976
2.4	0.3879	1.5697	12.0	1.0122	50.559
2.5	0.4121	1.7792	16.0	1.0513	81.203
2.6	0.4351	1.9995	20.0	1.0715	115.64
2.7	0.4571	2.2301	40.0	1.0946	327.01
2.8	0.4780	2.4708	80.0	1.0845	867.11
2.9	0.4980	2.7214	100.0	1.0782	1174.9
3.0	0.5170	2.9814	200.0	1.0562	2946.1
3.2	0.5526	3.5293	500.0	1.0307	9502.2
3.4	0.5851	4.1126	∞	1.000	∞
3.6	0.6148	4.7298			

TABLE II

(See Section 56)

Values of $f\left[\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}\right]$ Which Determines Dependence of Total Triode

Plate Current on Voltage Drop Along the Filament [8]

$\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}$	$f\left[\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}\right]$	$\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}$	$f\left[\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}\right]$
0	0	5.0	16.51
0.25	0.031	6.0	24.1
0.50	0.177	8.0	32.5
0.75	0.414	10	74.0
1.00	1.000	15	138
1.50	2.57	20	211
2.00	4.65	40	622
3.00	7.13		
4.00	9.94	Higher	$\dfrac{5}{2}\left[\dfrac{E_b}{E_f\left(1+\dfrac{1}{\mu}\right)}\right]^{\frac{3}{2}}$

TABLE III

(See Section 73)

ELECTRON EMISSION CONSTANTS

Clean Metals*

Atomic Number	Metal	Thermionic Measurements		Photoelectric Measurements
		A_0	E_W (volts)	E_W (volts)
3	Li	—	—	(2.28)
11	Na	—	—	(2.46)
19	K	—	—	(2.24)
37	Rb	—	—	(2.16–2.19)
55	Cs	162	1.81	(1.87–1.96)
12	Mg	—	—	(2.42)
20	Ca†	60	2.24	—
56	Ba	60	2.11	—
40	Zr	330	4.12	—
72	Hf	14.5	3.53	—
90	Th	70	3.38	—
73	Ta	60	4.12	4.11
42	Mo	55	4.15	4.15
74	W	60	4.54	4.54
75	Re	200	5.1	—
26	Fe	—	4.77	4.77
27	Co	—	—	(4.12–4.28)
28	Ni	1380	5.03	5.01
45	Rh	—	4.58	(4.57)
46	Pd	60	4.99	4.97
76	Os	—	4.7	—
78	Pt	17000	6.27	(6.30)
47	Ag	—	—	4.74
79	Au	—	—	4.90
80	Hg	—	—	(4.52)
6	C†	5.93	3.93	—
50	Sn	—	—	(4.28–4.39)

SPECIAL THERMIONIC SURFACES‡

Emission from specially prepared surfaces varies so greatly with details of preparation, external field, etc., that the constants given here have very limited significance; they are given chiefly to emphasize the contrast with clean-metal constants.

Surface	A_0	E_W (volts)
CaO (Calcium oxide)	129–249	1.77
SrO (Strontium oxide)	4.1–258	1.27
BaO (Barium oxide)	2.9–272	0.99
CaO + SrO + BaO	.0083	1.24

* From REIMANN, *Thermionic Emission*, John Wiley & Sons, Inc., p. 99, except as noted for calcium and carbon.

† DUSHMAN, *Rev. Mod. Phys.*, **2**, 394, 1930.

‡ From *International Critical Tables*, McGraw-Hill.

TABLE IV

(See Section 76)

RATIO OF HOT TO COLD RESISTANCES OF FILAMENT MATERIALS

T, °K	$\dfrac{\text{Hot Resistance}}{\text{Cold Resistance}}$		
	*Combined-Type Oxide-Coated Cathode (W.E.Co.)	†*Konel Metal	‡Tungsten
300	1.00	1.00	1.00
700	1.80	1.11	2.85
800	1.96	1.13	3.36
900	2.12	1.16	3.88
1000	2.26	1.19	4.41
1100	2.40	1.21	4.95
1200	2.53	1.24	5.48
1300	2.65	1.27	6.03
1400	2.75	1.29	6.57
1500	2.85	1.32	7.13
1600			7.70
1800			8.99
2000			10.03
2200			11.22
2400			12.45
2600			13.70
2800			14.96
3000			16.30

* DUSHMAN, *Rev. Modern Phys.*, **2**, 381 (1930).

† LOWRY, *Phys. Review*, **35**, 1367 (1930).

‡ Calculated from data given by Chaffee, *Theory of Thermionic Vacuum Tubes*, p. 100.

TABLE V

Atomic Number	Symbol	Element	Atomic Weight	Ionizing Potential	Isotopes	Shell Distribution of Electrons						
						1	2	3	4	5	6	7
1	H	Hydrogen	1.0078	13.53	1, 2	1						
2	He	Helium	4.002	24.46	4	2						
3	Li	Lithium	6.940	5.37	7, 6	2	1					
4	Be	Beryllium	9.02	9.28	9	2	2					
5	B	Boron	10.82	8.28	11, 10	2	3					
6	C	Carbon	12.006	11.22	12, 13	2	4					
7	N	Nitrogen	14.008	14.48	14, 15	2	5					
8	O	Oxygen	16.000	13.55	16, 18, 17	2	6					
9	F	Fluorine	19.00	18.6	19	2	7					
10	Ne	Neon	20.18	21.47	20, 22, 21	2	8					
11	Na	Sodium	22.997	5.12	23	2	8	1				
12	Mg	Magnesium	24.32	7.61	24, 24, 26	2	8	2				
13	Al	Aluminum	26.97	5.96	27	2	8	3				
14	Si	Silicon	28.06	8.12	28, 29, 30	2	8	4				
15	P	Phosphorus	31.02		31	2	8	5				
16	S	Sulphur	32.065	10.3	32, 34, 33	2	8	6				
17	Cl	Chlorine	35.457	12.96	35, 37	2	8	7				
18	A	Argon	39.94	15.69	40, 36, 38	2	8	8				
19	K	Potassium	39.10	4.32	39, 41, 40	2	8	8	1			
20	Ca	Calcium	40.08	6.09	40, 44, 42, 43	2	8	8	2			
21	Sc	Scandium	45.10	6.7	45	2	8	9	2			
22	Ti	Titanium	47.9	6.81	48, 46, 47, 50, 49	2	8	10	2			
23	V	Vanadium	50.95	6.76	51	2	8	11	2			
24	Cr	Chromium	52.01	6.74	52, 53, 50, 54	2	8	13	1			
25	Mn	Manganese	54.93	7.41	55	2	8	13	2			
26	Fe	Iron	55.84	7.83	56, 54, 57, 58	2	8	14	2			
27	Co	Cobalt	58.94	8.5	59	2	8	15	2			

TABLE V (*Continued*)

Atomic Number	Symbol	Element	Atomic Weight	Ionizing Potential	Isotopes	Shell Distribution of Electrons						
						1	2	3	4	5	6	7
28	Ni	Nickel	58.69	7.61	58, 60, 62, 64	2	8	16	2			
29	Cu	Copper	63.57	7.68	63, 65	2	8	18	1			
30	Zn	Zinc	65.38	9.36	64, 66, 68, 67, 70	2	8	18	2			
31	Ga	Gallium	69.72	5.97	69, 71	2	8	18	3			
32	Ge	Germanium	72.60	8.09	74, 72, 70, 73, 76	2	8	18	4			
33	As	Arsenic	74.93	10	75	2	8	18	5			
34	Se	Selenium	78.96	9.5	80, 78, 76, 82, 77, 74	2	8	18	6			
35	Br	Bromine	79.916	11.80	79, 81	2	8	18	7			
36	Kr	Krypton	83.7	13.94	84, 86, 82, 83, 80, 78,	2	8	18	8			
37	Rb	Rubidium	85.44	4.16	85, 87	2	8	18	8	1		
38	Sr	Strontium	87.63	5.67	88, 86, 87, 84	2	8	18	8	2		
39	Y	Yttrium	88.92	6.5	89	2	8	18	9	2		
40	Zr	Zirconium	91.22	6.92	90, 92, 94, 91, 96	2	8	18	10	2		
41	{Nb / Cb	Niobium / Columbium }	93.3		93	2	8	18	12	1		
42	Mo	Molybdenum	96.0	7.35	{98, 96, 95, 94, 92 / 100, 97	2	8	18	13	1		
43	Ma	Masurium				2	8	18	14	1		
44	Ru	Ruthenium	101.7		102, 101, 104, 100	2	8	18	15	1		
45	Rh	Rhodium	102.91	7.7	103	2	8	18	16	1		
46	Pa	Palladium	106.7	8.3	102–110	2	8	18	18			
47	Ag	Silver	107.880	7.54	107, 109	2	8	18	18	1		
48	Cd	Cadmium	112.41	8.96	114, 112, 116, 110, 111, 113, 106, 108	2	8	18	18	2		
49	In	Indium	114.8	5.76	115, 113	2	8	18	18	3		
50	Sn	Tin	118.70	7.30	120, 118, 116, 119, 124, 117, 122, 121, 112, 114, 115	2	8	18	18	4		

TABLE V (*Continued*)

Atomic Number	Symbol	Element	Atomic Weight	Ionizing Potential	Isotopes	Shell Distribution of Electrons						
						1	2	3	4	5	6	7
51	Sb	Antimony	121.76	8.5	121, 123	2	8	18	18	5		
52	Te	Tellurium	127.6		120–130	2	8	18	18	6		
53	I	Iodine	126.92	10	127	2	8	18	18	7		
54	Xe	Xenon	131.3	12.08	132, 129, 131, 134, 136, 128, 130, 126, 124	2	8	18	18	8		
55	Cs	Caesium	132.81	3.87	133	2	8	18	18	8	1	
56	Ba	Barium	137.36	5.19	138, 130–137	2	8	18	18	8	2	
57	La	Lanthanum	138.92		139	2	8	18	18	9	2	
58	Ce	Cerium	140.13		140, 142	2	8	18	19	9	2	
59	Pr	Praseodimium	140.92		141	2	8	18	20	9	2	
60	Nd	Neodymium	144.27		142, 144, 146, 145	2	8	18	21	9	2	
61	Il	Illinium				2	8	18	22	9	2	
62	{Sm}{Sa}	Samarium	150.43		144–154	2	8	18	23	9	2	
63	Eu	Europium	152.0		153, 151	2	8	18	24	9	2	
64	Gd	Gadolinium	157.3		155–160	2	8	18	25	9	2	
65	Tb	Terbium	159.2		159	2	8	18	26	9	2	
66	{Ds}{Dy}	Dysprosium	162.46		164, 163, 162, 161	2	8	18	27	9	2	
67	Ho	Holmium	163.5		165	2	8	18	28	9	2	
68	Er	Erbium	167.64		166, 168, 167, 170	2	8	18	29	9	2	
69	{Tm}{Tu}	Thulium	169.4		169	2	8	18	30	9	2	
70	Yb	Ytterbium	173.5		174, 172, 173, 176, 171	2	8	18	31	9	2	
71	Lu	Lutecium	175.0		175	2	8	18	32	9	2	

TABLE V (*Concluded*)

Atomic Number	Symbol	Element	Atomic Weight	Ionizing Potential	Isotopes	Shell Distribution of Electrons						
						1	2	3	4	5	6	7
72	Hf	Hafnium	178.6		180, 178, 177, 179	2	8	18	32	10	2	
73	Ta	Tantalum	181.4		181	2	8	18	32	11	2	
74	W	Tungsten	184.0		184, 186, 182, 183	2	8	18	32	12	2	
75	Re	Rhenium	186.31		185, 187	2	8	18	32	13, 14	2, 1	
76	Os	Osmium	190.8		192, 190, 189, 188	2	8	18	32	14, 15	2, 1	
77	Ir	Iridium	193.1		191, 193	2	8	18	32	15, 16	2, 1	
78	Pt	Platinum	195.23	8.9	192–198	2	8	18	32	16, 17	2, 1	
79	Au	Gold	197.2	9.2	197	2	8	18	32	18	1	
80	Hg	Mercury	200.61	10.38	202, 200, 199, 198, 201, 204, 196	2	8	18	32	18	2	
81	Tl	Thallium	204.4	6.07	205, 203	2	8	18	32	18	3	
82	Pb	Lead	207.22	7.38	208, 206, 207, 204, 203, 205	2	8	18	32	18	4	
83	Bi	Bismuth	209.00	8.0	209	2	8	18	32	18	5	
84	Po	Polonium				2	8	18	32	18	6	
85	?	?				2	8	18	32	18	7	
86	Rn	Radon	222	10.69		2	8	18	32	18	8	
87	?	?				2	8	18	32	18	8	1
88	Ra	Radium	225.97	10.2		2	8	18	32	18	8	2
89	Ac	Actinium				2	8	18	32	18	9	2
90	Th	Thorium	232.12		232	2	8	18	32	18	10	2
91	Pa	Protoactinium				2	8	18	32	18	11	2
92	U	Uranium	238.14		238, 235	2	8	18	32	18	12	2

Ionizing Potentials chiefly from *Atomic Energy States*, by Bacher and Goudsmit, McGraw-Hill, 1932. They are for atoms, not molecules.

See also pages 703 and 710 in *Introduction to Modern Physics*, by Richtmeyer, Second Edition, McGraw-Hill, 1934, and pages 154 to 160 in *Elements of Nuclear Physics*, by Rasetti, Prentice-Hall, 1936.

TABLE VI

(See Section 87)

QUANTUM-NUMBER COMBINATIONS FOR RECTANGULAR QUANTIZATION

n_u	n_v	n_w	n^2 (proportional to energy)	n	Number of Possible Arrangements	Number of Electrons
0	0	0	0	0	1	2
1 0 0	0 1 0	0 0 1	1	1	3	6
1 1 0	1 0 1	0 1 1	2	1.414	3	6
1	1	1	3	1.732	1	2
2 0 0	0 2 0	0 0 2	4	2.000	3	6
2 2 0 1 1 0	1 0 2 2 0 1	0 1 1 0 2 2	5	2.236	6	12
2 1 1	1 2 1	1 1 2	6	2.449	3	6

and so on

TABLE VII

(See Section 95 and Fig. 76)

GROSS WORK FUNCTIONS OF THE ALKALI METALS

	Lithium	Sodium	Potassium	Rubidium	Caesium
Specific Gravity......... (Handbook of Chemistry and Physics)	0.534	0.971	0.86	1.53	1.87
Atoms per Cubic Centimeters..............	4.67×10^{22}	2.56×10^{22}	1.334×10^{22}	1.086×10^{22}	0.854×10^{22}
Normal Maximum Energy, E_N.......... (Electron Volts)	4.68	3.14	2.04	1.78	1.51
Net Work Function E_W, Electron Volts........	2.28	2.46	2.24	2.17	1.91
Gross Work Function E_G, Electron Volts........	6.96	5.60	4.28	3.95	3.41
Lattice-Constant b, Centimeters.............. (Two atoms per cube, body-centered crystallization)	3.50×10^{-8}	4.27×10^{-8}	5.31×10^{-8}	5.68×10^{-8}	6.15×10^{-8}

TABLE VIII
(See Chapter X)

INTEGRALS CONTAINING ϵ^{-r^2}

A. The various integrals given below can in general be derived from the simple forms first given by a process of integration by parts, as follows:

$$\int_r^\infty \epsilon^{-r^2} r^2\, dr = \int_r^\infty \frac{r}{2}\, (2r\epsilon^{-r^2})dr = -\int_r^\infty \frac{r}{2}\, d\epsilon^{-r^2}$$

$$= -\frac{r\epsilon^{-r^2}}{2}\Big]_r^\infty + \int_r^\infty \frac{\epsilon^{-r^2}}{2}\, dr$$

$$= +\frac{r\epsilon^{-r^2}}{2} + \frac{1}{2}\int_r^\infty \epsilon^{-r^2}\, dr$$

B. Definite integrals of the form $\int_0^\infty \epsilon^{-r^2} r^n\, dr$

n even	n odd
$\dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^\infty \epsilon^{-r^2}\, dr = 1$	$2\displaystyle\int_0^\infty \epsilon^{-r^2} r\, dr = 1$
$\dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^\infty \epsilon^{-r^2} r^2\, dr = \dfrac{1}{2}$	$2\displaystyle\int_0^\infty \epsilon^{-r^2} r^3\, dr = 1$
$\dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^\infty \epsilon^{-r^2} r^4\, dr = \dfrac{3}{4}$	$2\displaystyle\int_0^\infty \epsilon^{-r^2} r^5\, dr = 2$
$\dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^\infty \epsilon^{-r^2} r^n\, dr = \dfrac{1\cdot 3\cdot 5\cdots(n-1)}{2^{n/2}}$	$2\displaystyle\int_0^\infty \epsilon^{-r^2} r^n\, dr = 1\cdot 2\cdot 3\cdot 4\cdots\dfrac{n-1}{2}$

C. Indefinite integrals of the form $\int \epsilon^{-r^2} r^n\, dr$, n being *odd*.

$$2\int \epsilon^{-r^2} r\, dr = -\epsilon^{-r^2}$$

$$2\int \epsilon^{-r^2} r^3\, dr = -\epsilon^{-r^2}(r^2 + 1)$$

$$2\int \epsilon^{-r^2} r^5\, dr = -\epsilon^{-r^2}(r^4 + 2r^2 + 2)$$

$$2\int \epsilon^{-r^2} r^n\, dr = -\epsilon^{-r^2} r^{n-1} + (n-1)\int \epsilon^{-r^2} r^{n-2}\, dr$$

D. Indefinite integrals of the form $\int_r^\infty \epsilon^{-r^2} r^n\, dr$, n being *even*.

Definition:

$$\text{erf } r = \frac{2}{\sqrt{\pi}} \int_0^r \epsilon^{-r^2} r^2\, dr \qquad \text{(this is called the } error\ function\text{)}$$

TABLE VIII (*Continued*)

Therefore, also using the definite integral in case $n = 0$, Part B, above

$$1 - \operatorname{erf} r = \frac{2}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} dr, \quad \text{so that}$$

$$\frac{2}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} dr = 1 - \operatorname{erf} r$$

$$\frac{2}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} r^2 dr = \frac{r\epsilon^{-r^2}}{\sqrt{\pi}} + \frac{1}{2}(1 - \operatorname{erf} r)$$

$$\frac{2}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} r^4 dr = \frac{r\epsilon^{-r^2}}{\sqrt{\pi}} \left(r^2 + \frac{3}{2} \right) + \frac{3}{4}(1 - \operatorname{erf} r)$$

$$\frac{2}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} r^n dr = \frac{\epsilon^{-r^2}}{\sqrt{\pi}} r^{n-1} + \frac{(n-1)}{\sqrt{\pi}} \int_r^\infty \epsilon^{-r^2} r^{n-2} dr$$

E. Miscellaneous

$$\int_0^\infty \frac{r\,dr}{B\epsilon^{r^2} + 1} = \frac{1}{2} \log \left(\frac{1}{B} + 1 \right)$$

(This is obtained by expansion into an infinite series, in two steps, one for the range $B\epsilon^{r^2} < 1$, one for the range $B\epsilon^{r^2} > 1$, with subsequent term-by-term integration.)[82]

TABLE IX
(See Fig. 95, Section 124)
POTENTIAL DISTRIBUTION FUNCTION OUTSIDE A PLANE ELECTRON-EMITTING CATHODE*

η	ξ negative, for points between true cathode and virtual cathode; $x < x_m$	ξ positive, for points beyond virtual cathode; $x > x_m$
0.00	0.0000	0.0000
0.10	0.5941	0.6693
0.20	0.8170	0.9674
0.40	1.1081	1.4092
0.60	1.3120	1.7636
1.00	1.5996	2.3522
1.40	1.8009	2.8539
1.80	1.9515	3.3040
2.20	2.0681	3.7187
2.60	2.1602	4.1071
3.00	2.2338	4.4750
4.00	2.3615	5.3274
5.00	2.4376	6.1098
6.00	2.4834	6.8416
8.00	2.5280	8.1963
10	2.5444	9.4465
15	2.5531	12.2747
20	2.5538	14.8260
30	2.5539	19.4253
40	2.5539	23.5939
50		27.4740
60		31.141
70		34.642
80		38.007
90		41.258
100		44.412
200		72.479
400		119.185
600		159.885
800		197.146
1000	2.5539	232.054

*From I. LANGMUIR, *The Effect of Space Charge and Initial Velocities on the Potential Distribution and Thermionic Current between Parallel-Plane Cathodes.* *Phys. Rev.*, Series 2, **21**, 419 (1923). See also reference No. 74 in Bibliography.

TABLE X
POSSIBLE COMBINATIONS OF POLAR QUANTUM NUMBERS; ALSO, GROUPS AND SUBGROUPS OF ELECTRONS IN SHELLS
(See Sections 182, 183, and 188, also Table V.)

n	n_r	l =	m_l	+	n_θ	m_s (spin)	Number of Electrons in Subgroup	Subgroup Symbol	Number of Electrons in Shell $(2n^2)$	Represented in Configuration Code as
1	1	0 =	0	+	0	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	2	s	2	$1s^2$
2	2	0 =	0	+	0	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	2	s	8	$2s^2 2p^6$
	1	1 =	$\{$ 1 clockwise, 0, 1 counter $\}$	+	$\{$ 0, 1, 0 $\}$ each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	6	p		
3	3	0 =	0	+	0	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	2	s	18	$3s^2 3p^6 3d^{10}$
	2	1 =	$\{$ 1 clockwise, 0, 1 counter $\}$	+	$\{$ 0, 1, 0 $\}$ each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	6	p		
	1	2 =	$\{$ 2 clockwise, 1 clockwise, 0, 1 counter, 2 counter $\}$	+	$\{$ 0, 1, 2, 1, 0 $\}$ each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	10	d		
4	4	0 =	0	+	0	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	2	s	32	$4s^2 4p^6 4d^{10} 4f^{14}$
	3	1 =	[three possible combinations, as above]		each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	6	p		
	2	2 =	[five possible combinations, as above]		each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	10	d		
	1	3 =	$\{$ 3 clockwise, 2 clockwise, 1 clockwise, 0, 1 counter, 2 counter, 3 counter $\}$	+	$\{$ 0, 1, 2, 3, 2, 1, 0 $\}$ each	$\{$ ½ clockwise $\}$ $\{$ ½ counter $\}$	14	f		

TABLE XI
TERM VALUES FOR THE NaI SPECTRUM*
(Arc Spectrum of Sodium)

The configuration of sodium's 11 electrons is as follows in the normal state:
$$1s^2\, 2s^2\, 2p^6\, 3s \quad {}^2S_{\frac{1}{2}}$$
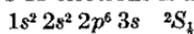
Configuration	Symbol	J	Term Value (cm^{-1})
$3s$	${}^2S$	$\frac{1}{2}$	41449.0
$3p$	${}^2P^\circ$	$\frac{1}{2}$	*24492.83*
		$1\frac{1}{2}$	*24475.65*
$4s$	${}^2S$	$\frac{1}{2}$	15709.50
$3d$	${}^2D$	$1\frac{1}{2}, 2\frac{1}{2}$	12276.18
$4p$	${}^2P^\circ$	$\frac{1}{2}$	*11181.63*
		$1\frac{1}{2}$	*11176.14*
$5s$	${}^2S$	$\frac{1}{2}$	8248.28
$4d$	${}^2D$	$1\frac{1}{2}, 2\frac{1}{2}$	6900.35
$4f$	${}^2F^\circ$	$2\frac{1}{2}, 3\frac{1}{2}$	6860.37

SERIES OF LEVELS

	mp	
m	${}^2P_{\frac{1}{2}}{}^\circ$	${}^2P_{1\frac{1}{2}}{}^\circ$
3	*24492.83*	*24475.65*
4	*11181.63*	*11176.14*
5	*6408.83*	*6406.34*
6	*4152.80*	*4151.30*
7	*2908.93*	*2907.46*
8	*2150.69*	*2149.80*
9	*1655.31*	*1654.08*
10	*1312.28*	*1312.28*
11	*1065.86*	*1065.86*
12	*883.40*	*883.40*

	ms
m	${}^2S_{\frac{1}{2}}$
3	41449.00
4	15709.50
5	8248.28
6	5077.31
7	3437.28
8	2480.65
9	1874.49
10	1466.0
11	1175.5
12	966.1

	md
m	${}^2D_{1\frac{1}{2}, 2\frac{1}{2}}$
3	12276.18
4	6900.35
5	4412.47
6	3061.92
7	2248.56
8	1720.88

	mf
m	${}^2F_{2\frac{1}{2}, 3\frac{1}{2}}{}^\circ$
4	*6860.37*
5	*4390.37*

* This table is reproduced, by permission of the publishers, from page 307 of the book *Atomic Energy States*, by R. F. Bacher and S. Goudsmit, published by the McGraw-Hill Book Company, Inc. 1932.

TABLE XII

TERM VALUES FOR THE HgI SPECTRUM*
(Arc Spectrum of Mercury)

The configuration of mercury's 80 electrons is as follows in the normal state of the atom:

$$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 4f^{14}\ 5s^2\ 5p^6\ 5d^{10}\ 6s^2\ {}^1S_0$$
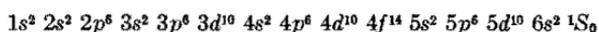
Configuration	Symbol	J	Term Value (cm⁻¹)
$6s^2$	^1S	0	84178.5
$6s\ 6p$	$^3P^\circ$	0	*46536.2*
		1	*44768.9*
		2	*40138.3*
$6s\ 6p$	$^1P^\circ$	1	*30112.8*
$6s\ 7s$	^3S	1	21830.8
$6s\ 7s$	^1S	0	20253.1
$6s\ 7p$	$^3P^\circ$	0	*14664.6*
		1	*14519.1*
		2	*12973.5*
$6s\ 6d$	^1D	2	12848.3
$6s\ 6d$	^3D	1	12845.1
		2	12785.0
		3	12749.0
$6s\ 7p$	$^1P^\circ$	1	*12886.1*

* This table is reproduced, by permission of the publishers, from page 227 of the book *Atomic Energy States*, by R. F. Bacher and S. Goudsmit, published by the McGraw-Hill Book Company, Inc., 1932.

TABLE XIII

TERM VALUES FOR THE NeI SPECTRUM*
(Arc Spectrum of Neon)

The configuration of neon's 10 electrons is as follows in the normal state of the atom:

$$1s^2 \ 2s^2 \ 2p^6 \ {}^1S_0$$

Old Notation	Configuration	Symbol	J	Term Value (cm^{-1})
$1P_0$	$2p^6$	^1S	0	173930
$1s_5$	$2p^5 \ (^2P_{1\frac{1}{2}}) \ 3s$	$1°$	2	39887.61
$1s_4$		$2°$	1	39470.16
$1s_3$	$2p^5 \ (^2P_{\frac{1}{2}}) \ 3s$	$3°$	0	39110.81
$1s_2$		$4°$	1	38040.73
$2p_{10}$	$2p^5 \ (^2P_{1\frac{1}{2}}) \ 3p$	1	1	25671.65
$2p_9$		2	3	24272.41
$2p_8$		3	2	24105.23
$2p_7$		4	1	23807.85
$2p_6$		5	2	23613.59
$2p_3$		6	0	23012.02
$2p_5$	$2p^5 \ (^2P_{\frac{1}{2}}) \ 3p$	7	1	23157.34
$2p_4$		8	2	23070.94
$2p_2$		9	1	22891.00
$2p_1$		10	0	20958.72
$3d_6$	$2p^5 \ (^2P_{1\frac{1}{2}}) \ 3d$	$1°$	0	12419.87
$3d_5$		$2°$	1	12405.23
$3d_4'$		$3°$	4	12339.15
$3d_4$		$4°$	3	12337.32
$3d_3$		$5°$	2	12322.26
$3d_2$		$6°$	1	12292.85
$3d_1''$		$7°$	2	12229.82
$3d_1'$		$8°$	3	12228.05
$3s_1''''$	$2p^5 \ (^2P_{\frac{1}{2}}) \ 3d$	$9°$	2	11520.82
$3s_1'''$		$10°$	3	11519.26
$3s_1''$		$11°$	2	11509.50
$3s_1'$		$12°$	1	11493.78
$2s_5$	$2p^5 \ (^2P_{1\frac{1}{2}}) \ 4s$	$1°$	2	15328.31
$2s_4$		$2°$	1	15133.47
$2s_3$	$2p^5 \ (^2P_{\frac{1}{2}}) \ 4s$	$3°$	0	14549.47
$2s_2$		$4°$	1	14395.75

* This table is reproduced, by permission of the publishers, from page 312 of the book *Atomic Energy States*, by R. F. Bacher and S. Goudsmit, published by the McGraw-Hill Book Company, Inc., 1932.

TABLE XIV

TERM VALUES FOR THE CuI SPECTRUM*

(Arc Spectrum of Copper)

The configuration of copper's 29 electrons is as follows in the normal state of the atom:

$$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s\ {}^2S_{\frac{1}{2}}$$

Configuration	Symbol	J	Term Values (cm⁻¹)
$3d^{10}\ 4s$	${}^2S$	$\frac{1}{2}$	62308.000
$3d^9\ 4s^2$	${}^2D$	$2\frac{1}{2}$	51105.435
		$1\frac{1}{2}$	49062.577
$3d^{10}\ 4p$	${}^2P^\circ$	$\frac{1}{2}$	*31772.698*
		$1\frac{1}{2}$	*31524.314*
$3d^9\ 4s\ 4p$	${}^4P^\circ$	$2\frac{1}{2}$	*23289.348*
		$1\frac{1}{2}$	*22194.01*
		$\frac{1}{2}$	*21364.27*
$3d^9\ 4s\ 4p$	${}^4F^\circ$	$4\frac{1}{2}$	*21398.862*
		$3\frac{1}{2}$	*21154.567*
		$2\frac{1}{2}$	*20745.105*
		$1\frac{1}{2}$	*20005.53*
$3d^{10}\ 5s$	${}^2S$	$\frac{1}{2}$	19170.791
$3d^9\ 4s\ 4p$	${}^4D^\circ$	$3\frac{1}{2}$	*18794.05*
		$2\frac{1}{2}$	*17901.732*
		$1\frac{1}{2}$	*17763.847*
		$\frac{1}{2}$	*17392.39*
$3d^9\ 4s\ 4p$	${}^2F^\circ$	$2\frac{1}{2}$	*18581.809*
		$3\frac{1}{2}$	*17344.777*
$3d^9\ 4s\ 4p$	${}^2P^\circ$	$\frac{1}{2}$	*16487.00*
		$1\frac{1}{2}$	*16428.689*
$3d^9\ 4s\ 4p$	${}^2D^\circ$	$1\frac{1}{2}$	*16135.158*
		$2\frac{1}{2}$	*15709.66*
$3d^{10}\ 5p$	${}^2P^\circ$	$\frac{1}{2}, 1\frac{1}{2}$	*12925.05*
$3d^{10}\ 4d$	${}^2D$	$1\frac{1}{2}$	12372.800
		$2\frac{1}{2}$	12365.943
$3d^{10}\ 6s$	${}^2S$	$\frac{1}{2}$	9459.251
$3d^{10}\ 6p$	${}^2P^\circ$	$1\frac{1}{2}$	*7523.94*
		$\frac{1}{2}$	*7280.26*
$3d^{10}\ 5d$	${}^2D$	$1\frac{1}{2}$	6920.332
		$2\frac{1}{2}$	6916.708
$3d^{10}\ 4f$	${}^2F^\circ$	$3\frac{1}{2}$	*6881.8*
		$2\frac{1}{2}$	*6878.2*
$3d^9\ 4s\ 4p$	${}^2F^\circ$	$3\frac{1}{2}$	*6278.05*
		$2\frac{1}{2}$	*4188.72*
$3d^9\ 4s\ 4p$	${}^2P^\circ$	$1\frac{1}{2}$	*5964.26*
		$\frac{1}{2}$	*3943.27*
$3d^9\ 4s\ 4p$	${}^2D^\circ$	$2\frac{1}{2}$	*5656.52*
		$1\frac{1}{2}$	*3617.14*

TABLE XV
Electron Mean Free Paths

In the derivations leading to Equations (486) and (487), for the mean free paths respectively of gas particles (or ions) and of electrons, it was assumed that gas particles are perfect spheres each having a definite radius b. Actually, gas particles are not spherical, nor are their boundaries sharply defined.

Various physical properties of gases depend primarily on atomic and molecular dimensions, and can therefore be used to evaluate atomic and molecular radii. It is customary in making such evaluations to assume that the gas particles are elastic spheres. Actually, however, the radial forces act in different ways, and are effective at different distances, in the interactions that give rise to the various properties. Therefore the different methods of evaluation lead to somewhat different values for the radii.

For electronic purposes the most satisfactory values of radii are those determined by measurements of the mean free paths of electrons of moderate velocities, that is, electrons having kinetic energies of a few electron volts. Brode has assembled the facts relative to the variations of electron mean free paths in various gases into a series of diagrams.[98] In each diagram a quantity P_c, which Brode calls the *probability of collision*, is plotted against electron energy. Values of P_c, read from Brode's curves, are tabulated below.

To quote Brode: "P_c can be considered as the effective area for collision of all the atoms in unit volume at unit pressure and 0° C temperature. This is a very convenient description of the probability of collision, but one is not justified in ascribing this area to any particular part of the atom." $\pi N b^2$ in Equation (487) is of course just such an area. Therefore, as Brode points out,

$$P_c = \frac{1}{l_0} \tag{883}$$

where l_0 is the electron mean free path at unit pressure, that is, a pressure of one millimeter of mercury and a temperature of 0° C.

The mean free path l at gas temperature T_g in degrees Kelvin, and at gas pressure p in millimeters of mercury is related to l_0 as follows:

$$\frac{l}{l_0} = \frac{1}{p}\left(\frac{T_g}{273}\right) \tag{884}$$

According to Equations (486) and (487) the electron mean free path should be $4\sqrt{2}$ longer than the ion mean free path, in any given gas. However, experiments have shown that electron mean free paths are usually longer than ion mean free paths by factors considerably greater than $4\sqrt{2}$.

The values of P_c that appear below have been read from Brode's curves. In regard to accuracy, Brode states that "the difference between the average curve and the measurements of individual observers is seldom as large as 10 per cent." For careful work direct use of Brode's curves is recommended, rather than use of the values of P_c tabulated below.

The use of Equation (883) and the Loschmidt number (see Table XVII) in Equation (487) shows that P_c and the atomic or molecular radius b are related as follows:

$$b = 0.3\sqrt{P_c} \times 10^{-8} \tag{885}$$

Loeb[V] gives a table of values of atomic and molecular radii determined by various methods, mostly nonelectrical. Such values are helpful in estimating *ionic* mean free paths.

TABLE XV (*Continued*)

P_c, also described as $1/l_0$, where l_0 is the electron mean free path, in centimeters, at 0° C and 1 millimeter of mercury pressure.[96]

Electron Velocity in Square Root Volts, $\sqrt{E}$		1	2	3	4	5	7	10	Maxima	Minima
Gas	*Symbol*									
Neon	N	6	8	10	11	12	11	8	12 at 5.5	
Argon	A	4	28	68	72	49	28	21	83 at 3.6	1 at 0.6
Krypton	Kr	4	48	96	93	64	36	26	99 at 3.5	4 at 0.8
Xenon	Xe	6	107	132	102	71			140 at 2.6	6 at 0.9
Helium	He	19	18	15	11	8	5	4		
Hydrogen	H_2	47	52	35	25	19	12	7	52 at 2	
Nitrous oxide	N_2O	21	44	47	57	60	50		80 at 1.4 / 60 at 5.0	21 at 0.9 / 41 at 2.2
Carbon dioxide	CO_2	23	49	40	53	58	48		50 at 2.1 / 58 at 4.8	21 at 1.3 / 31 at 2.4
Nitrogen	N_2	29	33	36	40	41	33	24	89 at 1.6 / 42 at 4.4	32 at 2.3
Oxygen	O_2	22	23	32	37	36	32		37 at 4.0	
Carbon monoxide	CO	45	67	39	43	43	34	26	122 at 1.6 / 45 at 4.5	39 at 3.2
Mercury	Hg	380	180	76	57	55	58	50	60 at 6	54 at 4.5
Cadmium	Cd	780	340	177	136	123	120	88	125 at 6.2	123 at 5.2
Zinc	Zn	{about 550}	300	142	112	88	69	63	72 at 8.3	70 at 6.6
Sodium	Na	980	640	470	340	300	170	120	1280 at 1.3	870 at 1.1
Potassium	K	1630	1000	690	500	390	200	120	1710 at 1.1	1360 at 0.7
Rubidium	Rb	1300	860	670	550	430	230	130	1500 at 1.1	
Caesium	Cs	930	1100	900	680	510	310	150	2000 at 1.5	930 at 1.0
Methane	CH_4	10	60	82	67	54	35		86 at 2.7	8 at 0.8
	C_2H_6	25	76	114	91	74	56		116 at 2.7	
	C_3H_8	35	100	143	120	100	74		150 at 2.6	

TABLE XVI

VAPOR PRESSURE OF MERCURY VAPOR AT VARIOUS TEMPERATURES.[UU] MERCURY-VAPOR AND CONDENSED-MERCURY TEMPERATURES THE SAME. (FIG. 178).

T_{Hg}, °C	Vapor Pressure P in Millimeters of Mercury	T_{Hg}, °C	Vapor Pressure P in Millimeters of Mercury
−30	0.00000478	+200	17.287
−20	0.0000181	210	23.723
−10	0.0000606	220	32.133
0	0.000185	230	42.989
+10	0.000490	240	56.855
+20	0.001201	250	74.375
+30	0.002777	260	96.296
40	0.006079	270	123.47
50	0.01267	280	156.87
60	0.02524	290	197.57
70	0.04825	300	246.80
80	0.08880	310	305.89
90	0.1582	320	376.33
100	0.2729	330	459.74
110	0.4572	340	557.90
120	0.7457	350	672.69
130	1.186	360	806.23
140	1.845	370	960.66
150	2.807	380	1138.4
160	4.189	390	1341.9
170	6.128	400	1574.1
180	8.796		
190	12.423	From 400° C to 1300° C, use	
200	17.287		

$$\log_{10} P = \frac{-3065}{T_{Hg}} + 7.75$$

TABLE XVII

PROBABLE VALUES OF THE FUNDAMENTAL GENERAL PHYSICAL CONSTANTS.* (esu), (emu) AND (p) SIGNIFY VALUES EXPRESSED IN TERMS OF UNITS OF THE ELECTROSTATIC, ELECTROMAGNETIC, AND PRACTICAL SYSTEMS RESPECTIVELY.

Electronic charge (e)	4.767×10^{-10} statcoulombs
	1.590×10^{-20} abcoulombs
	1.590×10^{-19} coulombs
Ratio of *esu* to *emu* units	2.99796×10^{10} cm per sec
Velocity of light	2.99796×10^{10} cm per sec
Ratio of charge to mass of an electron (e/m_e)	5.274×10^{17} (esu)
	1.759×10^{7} (emu)
	1.759×10^{8} (p)
Velocity of an electron whose kinetic energy is one electron volt	5.932×10^{7} cm per sec (p)
Mass of an electron	9.038×10^{-28} grams
Mass of a proton	1.6608×10^{-24} grams
Mass of a hydrogen atom	1.6617×10^{-24} grams
Mass of an atom of unit atomic weight (no such atom exists)	1.649×10^{-24} grams
Ratio of the mass of a hydrogen atom to that of an electron	1839
Ratio of the mass of an atom of unit atomic weight to that of an electron	1824
Ratio of the mass of a proton to that of an electron	1838
Atomic weights: Oxygen	16.0000
Hydrogen	1.00777 ± 0.00002
Helium	4.0022 ± 0.0004

TABLE XVII (*Continued*)

Ice point (absolute scale)	$273.18°$ K
Boltzmann's gas constant, k	1.3708×10^{-16} ergs per degree K
Loschmidt number (the number of molecules per cubic centimeter of any gas at $0°$ C and 760 millimeters of mercury pressure)	2.705×10^{19} molecules per cc
Ratio of temperature in degrees Kelvin to voltage equivalent of temperature . . .	11600 degrees per electron volt (p)
Avogadro's number (the number of molecules per gram molecule of any gas) . . .	6.064×10^{23}
The Stefan-Boltzmann radiation constant	5.735×10^{-5} ergs per sq cm per sec per (degree)^4
Density of mercury at $0°$ C, 760 millimeters of mercury pressure	13.59509 grams per cc
Wave length of red cadmium line (at $15°$ C, atmospheric pressure)	6438.4696 angstrom units
Ionizing potential of a hydrogen atom	13.529 electron volts (p)
Planck's quantum of action	6.542×10^{-27} erg seconds
Wave length of a photon whose energy is 1 electron volt	12335 angstrom units (p)

* Some uncertainty exists as to the *exact* values of the electronic charge e, the electronic mass m_e, and Planck's constant h. For the sake of consistency, and to avoid confusion, the values given here for those quantities are as tabulated by Richtmeyer in 1934.[X 712] For discussions of the uncertainties, see various articles by R. T. Birge and others.[125] Values for quantities not directly related to nor derived from e, m_e, and h are mostly as given by Birge in the Reviews of Modern Physics[125] in 1929. Calculations in the text may in some cases be based on earlier values of the constants than appear here.

BIBLIOGRAPHY

The bibliography consists of, first, a list of the books referred to in the text by letter and Arabic or Roman numbers, second, a list of technical articles, referred to by number. Thus the superior reference $^{F\ VIII}$ refers the reader to Chapter VIII of the book listed below as item F; also, superior reference $^{F\ 364}$ refers the reader to page 364 in the same book. The superior reference ^{94} refers the reader to the article or articles listed below as item 94.

A. ATTWOOD, S. S., "Electric and Magnetic Fields." John Wiley & Sons, 1932.

B. JEANS, J. H., "The Mathematical Theory of Electricity and Magnetism," 5th edition. Cambridge University Press, 1925.

C. MOORE, A. D., "Fundamentals of Electrical Design." McGraw-Hill Book Co., 1927.

D. MAXWELL, CLERK, "Electricity and Magnetism," 3rd edition, Vol. 1. Oxford Press, 1904.

E. CHAFFEE, E. L., "Theory of Thermionic Vacuum Tubes." McGraw-Hill Book Co., 1933.

F. GLASGOW, R. S., "Principles of Radio Engineering." McGraw-Hill Book Co., 1936.

G. McILWAIN, KNOX, and BRAINERD, J. G., "High Frequency Alternating Currents." John Wiley & Sons, 1931.

H. EVERITT, W. L., "Communication Engineering," 2nd edition. McGraw-Hill Book Co., 1936.

I. TERMAN, F. E., "Radio Engineering." McGraw-Hill Book Co., 1932.

J. DWIGHT, H. B., "Tables of Integrals and Other Mathematical Data." Macmillan, 1934.

K. PIERCE, B. O., "A Short Table of Integrals," 3rd edition. Ginn & Co., 1929.

L. VON ARDENNE, M., Die Kathodenstrahlrohren. Julius Springer Co., Berlin, 1934.

M. WATT, R. A. WATSON, "The Cathode Ray Oscillograph in Radio Research." His Majesty's Stationery Office, London.

N. ALBERTI, E., "Braunsche Kathodenstrahlrohren und ihre Anwendung." Julius Springer Co., Berlin, 1934.

O. HUND, AUGUST, "High-Frequency Measurements." McGraw-Hill Book Co. 1933.

P. "Radio Instruments and Measurements." U. S. Bureau of Standards Bulletin 74.

Q. VANDER BIJL, H. J., "Thermionic Vacuum Tubes." McGraw-Hill Book Co., 1920.

R. McARTHUR, E. D., "Electronics and Electron Tubes." John Wiley & Sons, 1936.

S. KOLLER, L. R., "The Physics of Electron Tubes." 2nd ed. McGraw-Hill Book Co., 1937.

T. REIMANN, A. L., "Thermionic Emission." John Wiley & Sons, 1934.

U. SLATER, J. C., and FRANK, N. H., "Introduction to Theoretical Physics." McGraw-Hill Book Co., 1933.

V. LOEB, L. B., "The Kinetic Theory of Gases," 2nd ed. McGraw-Hill Book Co., 1934.

W. LOEB, L. B., "The Nature of a Gas." John Wiley & Sons, 1931.

X. RICHTMYER, F. K., "Introduction to Modern Physics," 2nd ed. McGraw-Hill Book Co., 1934.

Y. WYCKOFF, R. W. G., "The Structure of Crystals." Chemical Catalog Co., 1924.

Z. "Handbook of Chemistry and Physics." Chemical Rubber Publishing Co., Cleveland, Ohio.

AA. PETERS, L. J., "Thermionic Vacuum Tube Circuits." McGraw-Hill Book Co., 1927.

BB. MOYER and WOSTREL, "Radio Handbook. McGraw-Hill Book Co., 1931.

CC. HENNEY, KEITH, "Radio Engineering Handbook." McGraw-Hill Book Co., 1935.

DD. PAULING, LINUS, and GOUDSMIT, S., "The Structure of Line Spectra." McGraw-Hill Book Co., 1930.

EE. BACHER, R. F., and GOUDSMIT, S., "Atomic Energy States." McGraw-Hill Book Co., 1932.

FF. CAMPBELL, N. R., and RITCHIE, D., "Photoelectric Cells." Isaac Pittman & Sons, 1934.

GG. ZWORYKIN, V. K., and WILSON, E. D., "Photocells and Their Application." John Wiley & Sons, 1932.

HH. HUGHES, A. L. and DuBRIDGE, L. A., "Photoelectric Phenomena." McGraw-Hill Book Co., 1932.

II. WOOD, R. W., "Physical Optics," 3rd ed. Macmillan, 1934.

JJ. PIERCE, G. W., "Electric Oscillations and Electric Waves." McGraw-Hill Book Co., 1920.

KK. BEWLEY, L. V., "Traveling Waves on Transmission Systems." John Wiley & Sons, 1933.

LL. SLEPIAN, J., "Conduction of Electricity in Gases, Lectures On." Westinghouse Electric & Manufacturing Co., East Pittsburgh, Pennsylvania.

MM. THOMSON, J. J., and THOMSON, G. P., "Conduction of Electricity through Gases," 2 volumes. Cambridge University Press, 1928.

NN. DARROW, K. K., "Electrical Phenomena in Gases." Williams and Wilkins, Baltimore, 1932.

OO. TOWNSEND, J. S. E., "Electricity in Gases." Oxford Clarendon Press.

PP. PRINCE, D. C., and VOGDES, F. B., "Principles of Mercury Arc Rectifiers and Their Circuits." McGraw-Hill Book Co., 1927.

QQ. MARTI, O. K., and WINOGRAD, H., "Mercury Arc Power Rectifiers." McGraw-Hill Book Co., 1930.

RR. GUNTHERSCHULZE, I. A., and DeBRUYNE, N. A., "Electric Rectifiers and Valves." John Wiley & Sons, 1928.

SS. KURTZ, E. B., and CORCORAN, J. F., "Introduction to Electric Transients." John Wiley & Sons, 1935.

TT. HENNEY, KEITH, "Electron Tubes in Industry," 2nd ed. McGraw-Hill Book Co., 1937.

UU. "International Critical Tables." McGraw-Hill Book Co., 1929.

VV. RASSETTI, F., "Elements of Nuclear Physics." Prentice-Hall, New York, 1936.

WW. GULLIKSEN and VEDDER, "Industrial Electronics." John Wiley & Sons, 1935.

XX. MOTT, N. F., and MASSEY, H. S. W., "The Theory of Atomic Collisions." Oxford Clarendon Press, 1933.

YY. WALKER, MILES, "Conjugate Functions for Engineers." Oxford University Press, 1933.

ZZ. MILLER, S. C., "Neon Signs." McGraw-Hill Book Co., 1935.

1. VODGES, F. B., and ELDER, F. R., "Formulas for the Amplification Constant for Three-Element Tubes." *Phys. Rev.*, **24**, 683 (December, 1924).

2. KING, R. W., "Calculation of the Constants of a Thermionic Vacuum Tube." *Phys. Rev.*, **15**, 256 (April, 1920).

3. MILLER, *Proc. I.R.E.*, **8**, 64 (1920).

4. ABRAHAM, *Arch. Electrotechnik*, **8**, 42 (1919).

5. SCHOTTKY, *Arch. Electrotechnik*, **8**, 21 (1919).

6. VON LAUE, M., "Uber die Wirkungsweise der Verstarkerruhren." *Ann. der Physik*, **59**, 465 (August, 1919).

7. VON LAUE, *Jahrb. Draht. Tel. Tel.*, **14**, 243 (1919).

8. KUSUNOSE, QUZIRO, "Calculation of Characteristics and the Design of Triodes." *Proc. I.R.E.*, **17**, 1706 (October, 1929).

9. HULL, A. W., "The Effect of a Uniform Magnetic Field on the Motion of Electrons between Coaxial Cylinders." *Phys. Rev.*, **18**, 31 (July, 1921).
HULL, A. W., "The Paths of Electrons in the Magnetron." *Phys. Rev.*, **23**, 112 (January, 1924).

10. HULL, A. W., "The Axially Controlled Magnetron." *Trans. A.I.E.E.*, **42**, 915 (1923).

11. HULL, A. W., "The Motions of Electrons between Coaxial Cylinders under the Influence of Current along the Axis." *Phys. Rev.*, **25**, 645 (May, 1925).

12. HULL, A. W., "The Measurement of Magnetic Fields of Medium Strength by Means of the Magnetron." *Phys. Rev.*, **22**, 279 (September, 1923).

13. ELDER, F. R., "The Magnetron Amplifier and Power Oscillator." *Proc. I.R.E.*, **13**, 159 (April, 1925).

14. KILGORE, G. R., "Magnetron Oscillators for the Generation of Frequencies Between 300 and 600 Megacycles." *Proc. I.R.E.*, **24**, 1140 (August, 1936). This paper contains many references to recent periodical literature. *See also* references 128 and 129.

15. MALOFF, I. G. and EPSTEIN, D. W., "Theory of Electron Gun." *Proc. I.R.E.*, **22**, 1386 (December, 1934).

16. BUSCH, H., "Berechnung der Bahn von Kathodenstrahlen im axialsymmetrischen elektromagnetischen Felde." *Ann. der Physik*, **81**, 974 (1926).

17. GLASER, W., "Uber geometrish-optische Abbildung durch Elektronenstrahlen." *Zeits. f. Phys.*, **80**, 451 (February, 1933).
GLASER, W., "Zur geometrischen Electronoptic des axialsymmetrischen electromagnetischen Feldes. *Zeits. f. Phys.*, **81**, 647 (April, 1933).
GLASER, W., "Theorie des Elektronenmikroskopes." *Zeits. f. Phys.*, **83**, 104 (June, 1933).

18. PICHT, J., "Beitrage zur Theorie der geometrischen Electronenoptik." *Ann. der Physik*, **15**, 926 (1932).

19. EPSTEIN, D. W., "Electron Optical System of Two Cylinders as Applied to Cathode-Ray Tubes." *Proc. I.R.E.*, **24**, 1095 (August, 1936). *See also* reference 15.

20. JOHNSON, J. B., "A Low Voltage Cathode Ray Oscillograph." *Am. Opt. Soc. Jl.*, **6**, 701 (September, 1922) and *Phys. Rev.*, **17**, 420 (March, 1921).

21. DUFOUR, A., *Comptes Rendus*, 1244 and 1478 (1924).

22. MacGREGOR-MORRIS, J. T., and MINES, R., "Measurements in Electrical Engineering by Means of Cathode Rays." *I.E.E. Jl. London*, **63**, 1056 (November, 1925).

23. WOOD, A. B., "Cathode Ray Oscillograph." *I.E.E. Jl. London*, **63**, 1046 (November, 1925).

24. ROGOWSKI-GROSSER, *Archiv für Elektr.*, **15**, 377 (1925).

25. ROGOWSKI-FLEGLER, *Archiv für Elektr.*, **15**, 297 (1925).

26. NORINDER, H., "Teknisk Tidskrift," Issue 31, *Electroteknik* **8** (1925).

27. NORINDER, H., *Bulletin of the Royal Board of Waterfalls*, Upsala, Sweden, Series E., No. 14.

28. BUSCH, H., *Archiv für Elektr.*, **18**, 583 (1927). (Magnetic cathode-ray focusing.)

29. GABOR, D., *Forschungshefte der Studiensellschaft für Hochtspannungsanlagen*, September, 1927.

30. TOBLER, F., and BERGER, K., *Bulletin S.E.V. Zurich*, **18**, 652.

31. BERGER, K., *Bulletin S.E.V. Zurich*, **19**, 688.

32. ROGOWSKI-FLEGLER-TAMM, *Archiv für Elektr.*, **18**, 513 (1927).

33. LEE, E. S., "Cathode Ray Oscillographs and Their Uses." *Gen. Elec. Rev.* (August, 1928).

34. GEORGE, R. H., "A New Type of Hot Cathode Oscillograph." *Trans. A.I.E.E.*, **48**, 884 (July, 1929).

35. THIBAUD, JEAN, *Jl. de Physique*, **10**, 161 (1929). (Magnetic cathode-ray focusing.)

36. BEDELL, F., and KUHN, J., "Linear Correction for Cathode Ray Oscillograph." *Phys. Rev.*, **36**, 993 (September, 1930).

37. von ARDENNE, M., "A Braun Tube for Direct Photographic Recording." *Experimental Wireless*, London, **7**, 66 (1930).

38. ACKERMANN, O., "A Cathode Ray Oscillograph with Norinder Relay." *Trans. A.I.E.E.*, **49**, 467 (April, 1930).

39. OSBON, W. O., "A New Cathode Ray Oscilloscope." *Elec. Jl.*, **28**, 322, (May, 1931).

40. JOHNSON, J. B., "The Cathode Ray Oscillograph." *Journal of the Franklin Institute*, p. 687, December, 1931; also *Bell System Technical Journal*, **11**, 1 (January, 1932).
   BEDFORD, L. H., "The Comparative Properties of Soft and Hard Cathode-Ray Tubes." *Jl. Sci. Inst.*, **13**, 177 (June, 1936).

41. ZWORYKIN, V. K., "Improvements in Cathode-Ray Tube Design." *Electronics*, **3**, 188 (November, 1931).

42. BRUCHE, E., "Die Geometrie des Beschleunigungsfeldes in ihrer Bedutung für den gaskonzentrierten Elektronenstrahl." *Zeits. f. Phys.*, **78**, 26 (September, 1932).

43. BRUCHE, E., "Biegsame Elektronenstrahlen. *Zeits. f. Phys.*, **78**, 177 (September, 1932).

44. von ARDENNE, M., *Hochfrequenz-technic und elektroakustic*, **18**, 1932. (Gas cathode-ray focusing).

45. KNOLL, M., and RUSKA, E., "Das Electronen-mikroskop." *Zeits. f. Phys.*, **78**, 318 (October, 1932).

46. METCALF, G. F., "A New Cathode-Ray Oscillograph Tube." *Electronics*, **4**, 158 (May, 1932).

47. RUSKA, ERNST, "Zur Fokussierbarkeit von Kathodenstrahlbundeln grosser Ausgangsguerschnitte." *Zeits. f. Phys.*, **83**, 684 (July, 1933).

48. TAYLOR, C. W., HEADRICK, L. B., and ORTH, R. T., "Cathode Ray Tubes for Oscillographic Purposes." *Electronics*, **6**, 332 (December, 1933).

49. ZWORYKIN, V. K., "On Electron Optics." *Journal of the Franklin Institute*, **215**, 535 (May, 1933).

50. VON ARDENNE, M., "A New Method of Removing Distortions Due to the Space Charge in Gas-Filled Cathode Ray Oscillograph Tubes." *Proc. I.R.E.*, **22**, 423 (April, 1934).

51. STINCHFIELD, J. M., "Cathode Ray Tubes and Their Application." *Trans. A.I.E.E.*, **53**, 1608 (December, 1934).

52. MALOFF, I. G., "Problems of Cathode Ray Television." *Electronics*, p. 10, January, 1934,
    VON ARDENNE, M., "Contribution on the Construction of a High Vacuum Braun Tube for Television and Measurements." *Zeitschr. f. Hochfrequenztechnik*, **44**, 168 (1934).

53. KUEHNI, H. P., "A Low Voltage Cathode Ray Oscillograph with Fluorescent Viewing Screen." *Gen. Elec. Rev*, **37**, 429 (September, 1934).
    "RCA Cathode-Ray Tubes and Allied Types," Technical Series TS-2, RCA Manufacturing Co., 1935.

54. GEORGE, R. H., HEIM, H. J., MAYER, A. F., and ROYS, C. S., "A Cathode Ray Oscillograph for Observing Two Waves." *Trans. A.I.E.E.*, **54**, 1095 (October, 1935).
    TARCEAUX, "Duplex Cathode-Ray Oscillograph." *Rev. Sci. Inst.*, **6**, 171 (1935).
    HUGHES, "Fluxation Selector for Double-Trace Cathode-Ray Oscillograph." *Rev. Sci. Inst.*, **7**, 89 (February, 1936).
    KNOLL, M., "Multiple Cathode-Ray Oscillographs." *Elekt. Zeits.*, **53**, 1101 (November, 1932).
    VON ARDENNE, M., "A Double-Beam Cathode-Ray Tube." *Electronics* (October, 1936).

55. DOWELL, J. C., and FOUST, C. M., "Laboratory Measurements of Impulse Voltages." *Trans. A.I.E.E.*, **52**, 537 (June, 1933). *See also* associated discussion.

56. BELLASCHI, P. L., "The Measurement of High-Surge Voltages." *Trans. A.I.E.E.*, **52**, 544 (June, 1933). *See also* associated bibliography and discussion.

57. CHILD, "Discharge from Hot CaO." *Phys. Rev.*, **32**, 498 (May, 1911).

58. LANGMUIR, I. L., "The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacua." *Phys. Rev.*, **2**, 450 (December, 1913).

59. KING, R. W., "Calculation of the Constants of a Thermionic Vacuum Tube." *Phys. Rev.*, **15**, 256 (April, 1920).

60. LANGMUIR, I. L., "Thermionenstrome im hohen Vakuum." *Phys. Zeits.*, **15**, 348 (April, 1914).

61. LANGMUIR, I. L., and BLODGETT, KATHARINE, "Currents Limited by Space Charge between Coaxial Cylinders." *Phys. Rev.*, **22**, 347 (October, 1923). Also, "Currents Limited by Space Charge Between Concentric Spheres." *Phys. Rev.*, **24**, 49 (July, 1924).

62. LANGMUIR, I. L., "The Effect of Space Charge and Initial Velocities on the Potential Distribution and Thermionic Current between Parallel Plane Electrodes." *Phys. Rev.*, **21**, 419 (April, 1923).

63. DREYER, J. F., "The Beam Power Output Tube." *Electronics*, p. 18 (April, 1936).

64. HARRIES, J. H. O., "Critical Distance Tubes." *Electronics*, p. 33 (May, 1936).

    HARRIES, J. H. O., "A New Power Output Valve." *Wireless World*, **37**, 105 (August 2, 1935).

    THOMPSON, H. C., "Electron Beams and Their Applications in Low Voltage Devices." *Proc. I.R.E.*, **24**, 1276 (October, 1936).

65. HERSHBERGER, W. D., "Modes of Oscillation in Barkhausen-Kurtz Tubes." *Proc. I.R.E.*, **24**, 964 (July 1936).

    PFETCHER and MULLER, *Hochfrequenz. und Elektroakustic*, **45**, 1 (1935).

    THOMPSON, B. J., and ZOTTU, P. D., "An Electron Oscillator with Plane Electrodes." *Proc. I.R.E.*, **22**, 1374 (December, 1934).

    HERSHBERGER, W. D., et al., *Physics*, **4**, 291 (1933).

    CLAVIER, *Elec. Communication*, **12**, 3 (1933).

    KOZANOWSKI, H. N., "A New Circuit for the Production of Ultra-Short-Wave Oscillations." *Proc. I.R.E.*, **20**, 957 (June, 1932).

    DYTRT, L. F., "Barkhausen-Kurtz Operation with Positive Plate Potentials." *Proc. I.R.E.*, **23**, 241 (March, 1935).

    MEGAW, E. C. S., "Electronic Oscillations." *Jl. I.E.E., London*, **72**, 313 (1933). (Contains comprehensive review of and bibliography relative to prior work.)

66. DUSHMAN, SAUL, "Thermionic Emission." *Rev. Mod. Phys.*, **2**, 381 (October, 1930).

    DUSHMAN, SAUL, "Electron Emission." *Trans. A.I.E.E.*, **53**, 1054 (July, 1934).

    DUSHMAN, SAUL, "Theory of Electron Emission." *Trans. Am. Electrochem Soc.*, **44**, 101 (1923).

67. KING, R. W., "Thermionic Vacuum Tubes and Their Applications." *Bell System Technical Journal*, p. 31, (October, 1923). *See also* page 82 in Vander Biji's book.[q]

68. HERZFELD, KARL F., "The Present Theory of Electric Conduction." *Trans. A.I.E.E.*, **53**, 523 (April, 1934). Contains comprehensive list of references to recent literature.

69. KARAPETOFF, VLADIMIR, "The Fermi-Dirac Statistical Theory of Gas Degeneration." *Mech. Eng.*, **55**, 237, 290, 358 (April, May, June, 1933).

70. SLATER, J. C., "The Electronic Structure of Metals." *Rev. Mod. Phys.*, **6**, 209 (October, 1934). (Extensive bibliography included.)

    BIELER, E. S., "The Fermi-Dirac Hypothesis of Gas Degeneration and Its Applications." *Jl. Franklin Inst.*, **206**, 65 (1928).

    LANGMUIR, I., "The Nature of Absorbed Films of Caesium on Tungsten. Part I. The Space Charge Sheath and the Image Force." *Phys. Rev.*, **43**, 224 (February 15, 1933).

    PETERSON, E. L., and NORDHEIM, L. W., "Resistance of Monovalent Metals." *Phys. Rev.*, **51**, 355 (March 1, 1937).

71. DARROW, KARL K., "Statistical Theories of Matter, Radiation, and Electricity." *Bell System Technical Jl.*, **8**, 700 (October, 1929); also *Rev. of Mod. Phys.*, **1**, 90 (July, 1929).

72. LANGMUIR, I. L., *Trans. Am. Electrochem Soc.*, **29**, 157 (1916).

73. COMPTON, K. T., and LANGMUIR, I., "Electrical Discharges in Gases, Part I. Survey of Fundamental Processes." *Rev. Mod. Phys.*, **2**, 123 (April, 1930).

74. LANGMUIR, I., and COMPTON, K. T., "Electrical Discharges in Gases, Part II. Fundamental Phenomena in Electrical Discharges." *Rev. Mod. Phys.* **2**, 191 (April, 1931).

75. SCHOTTKY, W., "Elektronemission der Metalle." *Physik. Zeitschr.*, **15**, 872 (November, 1914).

76. DE BRUYNE, N. A., "The Action of Strong Electric Fields on the Current from a Thermionic Cathode." *Proc. Roy. Soc. (A)*, **120**, 423 (September, 1928).

77. PFORTE, W. S., "Über die Vergrosserung des Sattigungsstromes von Gluhka-thoden durch starke elektrische Felder." *Zeits. f. Phys.* **49**, 46 (June, 1928).

78. RICHARDSON and COMPTON, *Phil. Mag.*, **24**, 575 (1912).

79. RICHARDSON, O. W., and ROBERTSON, F. S., "Contact Difference of Potential and Thermionic Emission." *London Phil. Mag.*, **43**, 557 (March, 1922).

80. VAN VOORHIS, C. C., "Heats of Condensation of Electrons and Positive Ions on Molybdenum in Gas Discharges." *Phys. Rev.*, **30**, 318 (September, 1927).

81. MILLIKAN, R. A., "The Distinction between Intrinsic and Spurious Contact E.M.F.'s and the Question of Absorption of Radiation by Metals in Quanta." *Phys. Rev.*, **18**, 236 (September, 1921).
    BOWIE, R. M., "This Matter of Contact Potential." *Proc. I.R.E.*, **24**, 1501 (November, 1936).

82. BARTLETT, R. S., and WATERMAN, A. T., "Space Charge vs. Image Force in Thermionic Emission." *Phys. Rev.*, **37**, 279 (February, 1931).
    BARTLETT, R. S., "Fermi-Dirac Statistics Applied to the Problem of Space Charge in Thermionic Emission." *Phys. Rev.*, **37**, 959 (April, 1931).
    WATERMAN, A. T., "The Equilibrium Distribution of Potential and of Electrons outside the Surface of a Conductor." *Phys. Rev.*, **38**, 1497 (October, 1931).

83. PEARSON, G. L., "Shot Effect and Thermal Agitation in an Electron Current Limited by Space Charge." *Physics*, **6**, 6 (January, 1935).
    LLEWELLYN, F. B., "A Study of Noise in Vacuum Tube and Attached Circuits." *Proc. I.R.E.*, **18**, 243 (February, 1930).
    See also references given by Pearson, Llewellyn, and Chaffee.[E 134]

84. CARSON, J. R., "A Theoretical Study of the Three-Element Vacuum Tube." *Proc. I.R.E.*, **7**, 187 (1919).

85. THOMPSON, B. J., "Graphical Determination of Performance of Push-Pull Audio Amplifiers." *Proc. I.R.E.*, **21**, 591 (April, 1933).
    EVERITT, W. L., "Optimum Operating Conditions for Class B Radio-Frequency Amplifiers." *Proc. I.R.E.*, **24**, 305 (February, 1936).
    FAY, C. E., "The Operation of Vacuum Tubes as Class B and Class C Amplifiers." *Bell System Tech. Jl.*, **11**, 28 (January, 1932).

85.5 GAGER, F. M., and RUSSELL, J. B., "A Quantitative Study of the Dynatron," **23**, 1536 (1935).
    GAGER, F. M., "The Grid-Coupled Dynatron." *Proc. I.R.E.*, **23**, 1048 (1935).
    TATUO HAYASI, "The Inner-Grid Dynatron and the Duodynatron." *Proc. I.R.E.*, **22**, 751 (1934).
    VANDER POL, B., "The Nonlinear Theory of Electric Oscillations." *Proc. I.R.E.*, **22**, 1051 (1934). (Extensive bibliography.)

86. DUSHMAN, SAUL, "Production of Light from Discharges in Gases." *Gen. Elec. Rev.*, **37**, 260 (June, 1934).

DUSHMAN, SAUL, "Low Pressure Gaseous Discharge Lamps." *Trans. A.I.E.E.*, **53**, 1204, 1283 (August and September, 1934).

FONDA, G. R., and YOUNG, A. H., "The A-C. Sodium Vapor Lamp." *Gen. Elec. Rev.*, **37**, 331 (July, 1934).

GORDON, N. T., "Operating Characteristics of Sodium Vapor Lamps." *Gen. Elec. Rev.*, **37**, 338 (July, 1934).

87. PRESCOTT, C. H., JR., and KELLY, M. J., "The Caesium-Oxygen-Silver Photo-electric Cell." *Bell System Tech. Jl.*, **11**, 334 (July, 1932).

IVES, H. E., and BRIGGS, H. B., "The Depth of Origin of Photoelectrons." *Phys. Rev.*, **40**, 802 (June 1, 1932).

IVES, H. E., and BRIGGS, H. B., "The Photoelectric Effect of Thin Films of Alkali Metal on Silver." *Phys. Rev.*, **38**, 1477 (Oct. 15, 1931).

FRY, T. C., "Plane Waves of Light." *Jl. Optical Soc. America*, **15**, 137 (1927); **16**, 1 (1928); **22**, 307 (June, 1932).

IVES, H. E., and KINGSBURY, E. F., "The Applicability of Photoelectric Cells to Colorimetry." *Jl. Opt. Soc. America*, **21**, 541 (September, 1931).

88. LINFORD, L. B., "The External Photoelectric Effect." *Rev. Mod. Phys.*, **5**, 34 (January, 1933).

FOWLER, R. H., "The Analysis of Photoelectric Sensitivity Curves for Clean Metals at Various Temperatures." *Phys. Rev.*, **38**, 45 (July, 1931).

89. OLPIN, A. R., "The Selective Photoelectric Effect from Two-Component Cathodes." *Phys. Rev.*, **38**, 1745 (Nov. 1, 1931).

FOWLER, R. H., "A Possible Explanation of the Selective Photoelectric Effect." *Proc. Royal Soc.*, **128**, 123 (1930).

90. SLEPIAN, J., and LUDWIG, L. R., Unpublished information relative to experiments on igniter rectifiers, Westinghouse Research Laboratories.

91. LANGMUIR, I., "Positive Ion Currents in the Positive Column of the Mercury Arc." *Gen. Elec. Rev.* **26**, 731 (November, 1923). *See* also *Jl. Franklin Inst.*, **196**, 751 (1923) and *Science*, **58**, 290 (1923).

92. LANGMUIR, I., and MOTT-SMITH, H., "Studies of Electric Discharges in Gases at Low Pressure." *Gen. Elec. Rev.*, **27**, 449, 538, 616, and 762 (July, August September, and November, 1924). *See also* under reference 94 below.

GREEVES, F. D., and JOHNSTON, J. E., McF., "The Glow Discharge through Oxygen." *London Phil. Mag.*, **21**, 659 (March, 1936). Contains numerous references to recent periodical literature.

MASON, R. C., "Probe Measurements on High Pressure Arcs." *Phys. Rev.*, **51**, 28 (January 1, 1937).

93. TONKS, L., and LANGMUIR, I., "A General Theory of the Plasma of an Arc." *Phys. Rev.*, **34**, 876 (September, 1929).

94. TONKS, L., "Plasma-Electron Resonance, Plasma Resonance, and Plasma Shape." *Phys. Rev.* **38**, 1219 (September 15, 1931).

TONKS, L., "The High Frequency Behavior of a Plasma." *Phys. Rev.*, **37**, 1458 (June 1, 1931). Contains many references.

THOMSON, J. J. "Oscillations in Discharge Tubes and Allied Phenomena." *London Phil. Mag.*, **11**, 697 (March, 1931).

TONKS, L., and LANGMUIR, I., "Oscillations in Ionized Gases." *Phys. Rev.* **33**, 195 (February, 1929).

LANGMUIR, I., "Oscillations in Ionized Gases." *Proc. Nat. Acad. Sciences*, **14**, 627 (August, 1928).

95. MILNE, A. E., "Statistical Equilibrium in Relation to Photoelectric Effect [and Recombination] and Its Application to the Determination of Absorption Coefficients." *London Phil. Mag.*, **47**, 209 (January, 1924).

MOHLER, F. L., "Recombination Spectra of Atomic Ions and Electrons." *Phys. Rev.*, **31**, 187 (February, 1928).

MOHLER, F. L., "Recombination and Photo-Ionization." *Rev. Mod. Phys.*, **1**, 216 (October, 1929). *See also* references 96.

96. CRAVATH, A. M., "The Rate of Formation of Negative Ions by Electron Attachment." *Phys. Rev.*, **33**, 605 (April, 1929).

CRAVATH, A. M., "The Rate at Which Ions Lose Energy in Elastic Collisions." *Phys. Rev.*, **36**, 248 (July 15, 1930).

COMPTON, K. T., "On the Motions of Electrons in Gases." *Phys. Rev.* **22**, 333 (October, 1923).

COMPTON, K. T., "Mobilities of Electrons in Gases." *Phys. Rev.*, **22**, 432 (November, 1923).

SUITS, C. G., "The Determination of Arc Temperature from Sound Velocity Measurements." *Physics*, **6**, 190 (June, 1935).

SUITS, C. G., "Studies of Arc Temperature by an Optical Method." *Physics*, **6**, 315 (October, 1935).

97. SLEPIAN, J., "Extinction of an A-C Arc." *Trans. A.I.E.E.*, **47**, 1398 (October, 1928).

SLEPIAN, J., "Extinction of Long A-C Arcs." *Trans. A.I.E.E.*, **49**, 421 (April, 1930).

SLEPIAN, J., and STROM, A. D., "Arcs in Low Voltage A-C Networks." *Trans. A.I.E.E.* **50**, 847 (September, 1931).

SLEPIAN, J., "Flames from Arcs." *Trans. A.I.E.E.*, **49**, 56 (January, 1930).

SLEPIAN, J., and DENAULT, C. L., "The Expulsion Fuse." *Trans. A.I.E.E.*, **51**, 157 (March, 1932).

See also referenes 74 and 93 above.

98. BRODE, R. B., "Quantitative Study of the Collisions of Electrons with Atoms." *Rev. Mod. Phys.*, **5**, 257 (October, 1933). Contains many references to periodical literature and a summary of results to date.

WEBB, G. M., "The Elastic Scattering of Electrons in Argon and Krypton." *Phys. Rev.*, **47**, 379 (March 1, 1935).

WEBB, G. M., "The Elastic Scattering of Electrons in Molecular Hydrogen." *Phys. Rev.*, 47, 384 (March 1, 1935).

McMILLEN, J. H., "Elastic Electron Scattering in Potassium." *Phys. Rev.*, **46**, 983 (Dec. 1, 1934).

See also the book "The Theory of Atomic Collisions," by Mott and Massey.[xx]

99. COMPTON, K. T., "The Electric Arc." *Trans. A.I.E.E.*, **46**, 868 (June, 1927).

100. MYER, J. L., "New Studies of the Arc Discharge." *Trans. A.I.E.E.*, **52**, 250 (March, 1933). Contains bibliography of earlier literature.

101. SLEPIAN, J., "The Electric Arc in Circuit Interrupters." *Franklin Inst. Jl*, **214**, 413 (October, 1932). *See also* Figs. 159, 160, 161 in Seeliger, "Physik der Gasentladungen," and Figs. 419 *ff* in "Handbuch der experimental Physik," XIV, 3 theil, p. 687.

102. ATTWOOD, S. S., Dow, W. G., and KRAUSNICK, W., "Reignition of Metallic A-C Arcs in Air." *Trans. A.I.E.E.*, **50**, 854 (September, 1931).

102. PARK, R. H., and SKEATS, W. F., Circuit Breaker Recovery Voltages."
Trans. A.I.E.E., **50**, 204 (March, 1931).

DOW, W. G., ATTWOOD, S. S., and TIMOSHENKO, G. S., "Probe Measurements and Potential Distribution in Copper A-C Arcs." *Trans. A.I.E.E.*, **52**, 926 (September, 1933).

BOEHNE, E. W., "The Determination of Circuit Recovery Rates." *Trans. A.I.E.E.*, **54**, 530 (May, 1935).

VAN SICKLE, R. C., and BERKEY, W. E., "Arc Extinction Phenomena in High Voltage Circuit Breakers." *Trans. A.I.E.E.*, **52**, 850 (September, 1933).

103. SLEPIAN, J., "Theory of the Deion Circuit Breaker." *Trans. A.I.E.E.*, **48**, 523 (April, 1929).

DICKINSON, R. C., and BAKER, B. P., "The Structural Development of the Deion Circuit Breaker." *Trans. A.I.E.E.*, **48**, 528 (April, 1929).

104. EMELEUS, K. G., and DUFFENDACK, O. S., "Spectral and Impact Phenomena In the Faraday Dark Space." *Phys. Rev.*, **47**, 460 (March, 1935). Contains many references to recent periodical literature.

EMELEUS, K. W., BROWN, W. L., and COWAN, H. McN., "Negative Sections of the Cold-Cathode Glow Discharge in Helium." *London Phil. Mag.* **17**, 146 (January, 1934). Contains an extensive list of references to recent periodical literature.

BREWER, A. K., and MILLER, R. R., "The Positive Ion Current at the Cathode in the Glow Discharge. *Phys. Rev.*, 42, 786 (Dec. 15, 1932).
See also the book "Electrical Phenomena in Gases," by Darrow.[NN]

105. JOHNSON, E. A., and HARRIS, L., "Disintegration of Sputtered Deposits." *Phys. Rev.*, **45**, 630 (May 1, 1934).

JOHNSON, E. A., and HARRIS, L., *Rev. Sci. Inst.*, **5**, 153 (1934).

KINGDON, R. H., and LANGMUIR, I., "The Removal of Thorium from the Surface of a Thoriated Tungsten Filament by Positive Ion Bombardment." *Phys. Rev.*, **22**, 148 (August, 1923). Suggests a theory of sputtering mechanism.

COMPTON, K. T., and LAMAR, E. S., "A Test of the Classical Momentum Transfer Theory of Accommodation Coefficients of Ions at Cathodes." *Phys. Rev.*, **44**, 338 (September 1, 1933).

See also reports on recent cathode sputtering investigations carried on in the Electrical Engineering and Physics Departments, Massachusetts Institute of Technology.

106. See the book "Neon Signs," by Miller.[ZZ]

107. MASON, R. C., "The Cathode Fall of an Arc." *Phys. Rev.*, **38**, 427 (August 1, 1931).

TONKS, L., "The Pressure of Plasma Electrons and the Force on the Cathode of an Arc." *Phys. Rev.*, **46**, 278 (August 15, 1934).

FOUNDS, C. G., "Fundamental Phenomena in Sodium-Vapor Lamps." *Gen. Elec. Rev.*, **37**, 269 (June, 1934).

108. See reference 91 for Langmuir's original suggestions relative to the mechanism of electron emergence at the cathode spot of an arc. Other papers developing this suggestion are listed below. Mason's paper, included in reference 107 preceding, points out weaknesses in it.

MACKEOWN, S. S., "The Cathode Drop in an Electric Arc." *Phys. Rev.* **34**, 611 (August 15, 1929).

COMPTON, K. T., and VAN VOORHIS, *Proc. Nat. Acad. Sciences*, **13**, 336 (1927). *See also* reference 99.

109. See papers by Langmuir and Mott-Smith, reference 92 above, also reference 93.

110. See the articles by Langmuir and Mott-Smith listed in reference 92, also others included in references 92 and 104, and papers referred to therein.

111. HULL, A. W., "Hot-Cathode Thyratrons. Part I. Characteristics." *Gen. Elec. Rev.*, **23**, 213 (April, 1929).
    HULL, A. W., "Hot-Cathode Thyratrons. Part II. Applications." *Gen. Elec. Rev.*, **32**, 390 (July, 1929).
    See also McArthur's book [R] for discussion of magnetic control of thyratrons.

112. GULLIKSEN, F. H., and STODDARD, R. N., "Industrial Electronic Control Applications." *Trans. A.I.E.E.*, **54**, 40 (January, 1935).
    CHAMBERS, D. E., "Applications of Electron Tubes in Industry." *Trans. A.I.E.E.*, **54**, 82 (January, 1935).
    See also the book by Gulliksen and Vedder,[WW] and one by Henney.[TT]
    HORTON, J. W., " Use of Vacuum Tubes in Measurements." *Trans. A.I.E.E.*, **54**, 93 (January, 1935). Contains comprehensive bibliography of recent periodical literature.
    MORACK, M. M., "Voltage Impulses for Thyratron Grid Control." *Gen. Elec. Rev.*, **37**, 288 (June, 1934).
    VEDDER, E. H., and EVANS, M. S., "Photoelectric Control of Resistance Type Metal Heaters." *Trans. A.I.E.E.*, **54**, 645 (June, 1935).
    GULLIKSEN, F. H., "An Electronic Regulator for D-C Generators." *Trans. A.I.E.E.*, **55**, 873 (August, 1936).
    REICH, H. J., "Electronic Transient Visualizers." *Trans. A.I.E.E.*, **55**, 1314 (December, 1936). *See also* references 124 and various other papers in *Trans. A.I.E.E.* and in *Electronics* for recent years.
    DAWSON, J. W., "New Developments in Ignitron Welding Control." *Trans. A.I.E.E.*, **55**, 1371 (December, 1936).
    PACKARD, D., and HUTCHINGS, J. H., "Sealed-Off Ignitrons for Resistance Welding Control." *Gen. Elec. Rev.*, **40**, 93 (February, 1937).

113. JOURNEAUX, DIDIER, "Voltage Control of Vapor Rectifiers." *Trans. A.I.E.E.*, **53**, 976 (June, 1934). *See also* references at end of Journeaux's article.

114. SLEPIAN, J., and LUDWIG, L. R., "A New Method for Initiating the Cathode of an Arc." *Trans. A.I.E.E.*, **52**, 693 (June, 1933).
    CAGE, J. M., "Theory of the Immersion Mercury-Arc Ignitor." *Gen. Elec. Rev.*, **38**, 464 (October, 1935).

115. SLEPIAN, J., and LUDWIG, L. R., "Backfires in Mercury Arc Rectifiers." *Trans. A.I.E.E.*, **51**, 92 (March, 1932).

116. HULL, A. W., and BURGER, E. E., "The Disappearance of Gas in Discharge Tubes." *Phys. Rev.*, **40**, 1044 (June 15, 1932). *See also* "Neon Signs" by Miller.[ZZ]

117. LIVINGSTON, O. W., and MASER, H. T., "Shield Grid Thyratrons." *Electronics*, April, 1934.

118. DOW, W. G., and POWERS, W. H., "Firing Time of an Igniter Type of Tube." *Trans. A.I.E.E.*, **54**, 942 (September, 1935).
    See also the article by Slepian and Ludwig, in reference 114.

119. STOUT, M. B., "Analysis of Rectifier Filter Circuits." *Trans. A.I.E.E.*, **54**, 977 (September, 1935).
    HUSS, P. O., "An Analysis of Copper-Oxide Rectifier Circuits." *Trans. A.I.E.E.*, **56**, 354 (March, 1937).

120. REICH, H. J., "The Relaxation Inverter and D-C Transformer." *Rev. Sci. Inst.*, **4**, 147 (March, 1933).

121. WAGNER, C. F., "Parallel Inverter with Resistance Load." *Trans. A.I.E.E.*, **54**, 1227 (November, 1935).

122. WAGNER, C. F., "Parallel Inverter with Inductive Load." *Trans. A.I.E.E.*, **55**, 970 (September, 1936).

123. SABBAH, C. A., "Series-Parallel Type Static Converters." Parts I, II, III. *Gen. Elec. Rev.*, **34**, 288, 580, 738 (May, October, December, 1931).

124. WILLIS, C. H., BEDFORD, B. D., and ELDER, F. R., "Constant-Current D-C Transmission." *Trans. A.I.E.E.*, **54**, 103 (January, 1935).

    ALEXANDERSON, E. F. W., and MITTAG, A. H., "The 'Thyratron' Motor." *Trans. A.I.E.E.*, **53**, 1517 (November, 1934).

    SCHMIDT, A., JR., and GRIFFITH, R. C., "A Static Thermionic Tube Frequency Changer." *Trans. A.I.E.E.*, **54**, 407 (October, 1935).

    WAGNER, C. F., and LUDWIG, L. R., "The 'Ignitron' Type of Inverter." *Trans. A.I.E.E.*, **53**, 1384 (October, 1934).

    HERSKIND, C. C., "Grid Controlled Rectifiers and Inverters." *Trans. A.I.E.E.*, **53**, 926 (June, 1934).

125. BIRGE, R. T., "Values of the General Physical Constants." *Rev. Mod. Phys.*, **1**, 1 (July, 1929).

    BIRGE, R. T., *Phys. Rev.*, **40**, 228 (1932) and reference 1 therein.

    ROBINSON, H. R., *Phil. Mag.*, **18**, 1086 (1934).

    BOND, W. N., *Nature*, **133**, 327 (1934).

    KELLSTROM, G., *Nature*, **136**, 682 (1935).

    BIRGE, R. T., *Phys. Rev.* (December 15, 1935).

    BIRGE, R. T., *Nature*, **137**, 187 (1936).

    BIRGE, R. T., *Science*, **79**, 483 (1934).

    BIRGE, R. T., *Nature*, **134**, 771 (1934).

    BEARDEN, J. A., *Phys. Rev.*, **48**, 385 (1935).

126. GRONDAHL, L. O. "The Copper-Cuprous-Oxide Rectifier and Photocell." *Rev. Mod. Phys.*, **5**, 141 (April, 1933).

127. LOEB, L. B., "The Problem of the Mechanism of Static Spark Discharge." *Rev. Mod. Phys.*, **8**, 267 (July, 1936).

    DRUYVESTEYN, M. J., "Calculation of Townsend's α for Neon." *Physica*, **3**, 65 (February, 1936).

128. OKABE, KINJIRO, "On the Short-Wave Limit of Magnetron Oscillations." *Proc. I.R.E.*, **17**, 652 (April, 1929).

129. HOAG, J. B., "A Note on the Theory of the Magnetron Oscillator." *Proc. I.R.E.*, **21**, 1132 (August, 1933).

130. MORSE, P. M., ALLIS, and LAMAR. "Velocity Distributions for Elastically Colliding Electrons." *Phys. Rev.*, **48**, 412 (September, 1935).

131. LLEWELLYN, F. B., "Vacuum-Tube Electronics at Ultra-High Frequencies." *Bell System Tech. Jl.*, **13**, 59 (January, 1934).

    LLEWELLYN, F. B., "Note on Vacuum Tube Electronics at Ultra-High Frequencies." *Proc. I.R.E.*, **23**, 112 (February, 1935).

132. BROWN, G. H., "A Critical Study of the Characteristics of Broadcast Antennas as Affected by Antenna Current Distribution." *Proc. I.R.E.*, **24**, 48 (1936).

    BARROW, W. L., "On the Impedance of a Vertical Half-Wave Antenna above an Earth of Finite Conductivity." *Proc. I.R.E.*, **23**, 150 (1935).

133. PIKE, O. W., and ULREY, DAYTON, "Ratings of Industrial Electronic Tubes." *Trans. A.I.E.E.*, **53**, 1577 (December, 1934).

INDEX

Amplifier
plate current
a-c and d-c components, 270
expansion as series, 298
plate dissipation, 287
in Class B amplifiers, 305
plate resistance, choice of, 283
plate voltage components, 271
power, 318
Class B, 303
decibel gain, 314 ff
load resistance selection, 283 ff
maximum undistorted output, 284
plate circuit efficiency, 286 ff, 305
plate dissipation, 287, 305
reactive load
circuit diagram, 272
current-voltage locus, 272
equivalent circuit, 272
harmonic distortion, 282, 297
phase relations, 274
vector diagram, 278
resistance coupling; *see* Resistance
coupling
resistance load
circuit diagram, 266
equivalent circuit, 266
harmonic distortion, 282, 296 ff
phase relations, 270 ff
power output, 285
voltage gain, 276, 281, 283, 303
transformer coupling; *see* Trans-
former coupling
tuned
parallel, 330
series, 336
vector diagrams; *see* Vector diagram
voltage, 318
voltage gain
decibel measure of, 315
definition, 1, 267
parallel feed, 291
reactive load, 278
resistance load, 276, 281, 283, 303
resistance coupling, 318
tuned amplifier, 330
transformer coupling, 321, 324
Angstrom units
cadmium red, standard, 557
definition, 344

Angstrom units
measure atomic spacing, 388
photon's electron volts, relation to
344, 557
Angular momenta, orbital
combination principles, 359, 362, 365,
372
dimensions same as action, 358
neon's ion, 369
Russell-Saunders coupling, 362
total angular momentum, 358, 363 ff
vector combination, 362, 372
Anode
arc-back to, 488
fall of potential, 428, 466
fall space, 429, 466
formation on grid, 471
spot, 426
Antenna, radiation from, 379, 381, 382
Arc, electric
alternating current, 453
anode fall, 466
cathode fall, 463
cathode spot, 463
current dependence, 426
description, 426
dimensions, 434
high-frequency characteristic, 452
high pressure
cathode spot, 465
collisions in, 436
energy scattering in plasma, 433
sheaths around, 466
initiation, 427
arc-back, 484, 488, 501
at sheath, 489
grid control of, 494
igniter rod, 481, 482, 500 ff
mercury pool, 481
thyratron, mechanism, 497
keep-alive, 481, 483
light from, 341, 426
loop characteristic, 453
low pressure
cathode spot, 465
energy scattering in plasma, 433
energy transfer, 446
mercury; *see* Mercury vapor rectifier
oscillating, 451
plasma; *see* Plasma