# ELECTRONICS

# McGraw-Hill Electrical and Electronic Engineering Series

FREDERICK EMMONS TERMAN, *Consulting Editor*

W. W. HARMAN and J. G. TRUXAL, *Associate Consulting Editors*

*Bailey and Gault* · ALTERNATING-CURRENT MACHINERY
*Beranek* · ACOUSTICS
*Bruns and Saunders* · ANALYSIS OF FEEDBACK CONTROL SYSTEMS
*Cage* · THEORY AND APPLICATION OF INDUSTRIAL ELECTRONICS
*Cauer* · SYNTHESIS OF LINEAR COMMUNICATION NETWORKS, VOLS. I AND II
*Cuccia* · HARMONICS, SIDEBANDS, AND TRANSIENTS IN COMMUNICATION
  ENGINEERING
*Cunningham* · INTRODUCTION TO NONLINEAR ANALYSIS
*Eastman* · FUNDAMENTALS OF VACUUM TUBES
*Evans* · CONTROL-SYSTEM DYNAMICS
*Feinstein* · FOUNDATIONS OF INFORMATION THEORY
*Fitzgerald and Higginbotham* · BASIC ELECTRICAL ENGINEERING
*Fitzgerald and Kingsley* · ELECTRIC MACHINERY
*Geppert* · BASIC ELECTRON TUBES
*Glasford* · FUNDAMENTALS OF TELEVISION ENGINEERING
*Happell and Hesselberth* · ENGINEERING ELECTRONICS
*Harman* · FUNDAMENTALS OF ELECTRONIC MOTION
*Harrington* · INTRODUCTION TO ELECTROMAGNETIC ENGINEERING
*Hayt* · ENGINEERING ELECTROMAGNETICS
*Hessler and Carey* · FUNDAMENTALS OF ELECTRICAL ENGINEERING
*Hill* · ELECTRONICS IN ENGINEERING
*Johnson* · TRANSMISSION LINES AND NETWORKS
*Kraus* · ANTENNAS
*Kraus* · ELECTROMAGNETICS
*LePage* · ANALYSIS OF ALTERNATING-CURRENT CIRCUITS
*LePage and Seely* · GENERAL NETWORK ANALYSIS
*Millman and Seely* · ELECTRONICS
*Millman and Taub* · PULSE AND DIGITAL CIRCUITS
*Rodgers* · INTRODUCTION TO ELECTRIC FIELDS
*Rüdenberg* · TRANSIENT PERFORMANCE OF ELECTRIC POWER SYSTEMS
*Ryder* · ENGINEERING ELECTRONICS
*Seely* · ELECTRON-TUBE CIRCUITS
*Seely* · ELECTRONIC ENGINEERING
*Seely* · INTRODUCTION TO ELECTROMAGNETIC FIELDS
*Seely* · RADIO ELECTRONICS
*Siskind* · DIRECT-CURRENT MACHINERY
*Skilling* · ELECTRIC TRANSMISSION LINES
*Skilling* · TRANSIENT ELECTRIC CURRENTS
*Spangenberg* · FUNDAMENTALS OF ELECTRON DEVICES
*Spangenberg* · VACUUM TUBES
*Stevenson* · ELEMENTS OF POWER SYSTEM ANALYSIS
*Storer* · PASSIVE NETWORK SYNTHESIS
*Terman* · ELECTRONIC AND RADIO ENGINEERING
*Terman and Pettit* · ELECTRONIC MEASUREMENTS
*Thaler* · ELEMENTS OF SERVOMECHANISM THEORY
*Thaler and Brown* · SERVOMECHANISM ANALYSIS
*Thompson* · ALTERNATING-CURRENT AND TRANSIENT CIRCUIT ANALYSIS
*Truxal* · AUTOMATIC FEEDBACK CONTROL SYSTEM SYNTHESIS

# *ELECTRONICS*

JACOB MILLMAN, Ph.D.

*Professor of Electrical Engineering*
*Columbia University*

SAMUEL S̶E̶E̶LY, Ph.D.

*Professor and Chairman, Department of Electrical*
*Engineering, Syracuse University*

SECOND EDITION

INTERNATIONAL STUDENT EDITION

# ELECTRONICS

*INTERNATIONAL STUDENT EDITION*

Exclusive rights by Kōgakusha Co., Ltd. for manufacture and export from Japan. This book cannot be re-exported from the country to which it is consigned by Kōgakusha Co., Ltd. or by McGraw-Hill Book Company, Inc. or any of its subsidiaries.

III

# PREFACE TO THE SECOND EDITION

The second edition of *Electronics* has been extensively rewritten, although the book retains the same philosophy of approach as the first edition. The text is intended for a fundamental course in electronics in which physical concepts and engineering applications are integrated. It is designed to lay the groundwork for specialized courses in communications, electron-tube circuits, industrial electronics, etc. The revisions incorporated in this edition are the result of the experience gained in the use of the book over the many years since its original publication.

A detailed enumeration of the significant changes and additions is included below. In brief, the changes seek to (*a*) modernize the notation and the system of units (the rationalized mks system is used), (*b*) expand the discussion of a number of topics for added clarity, (*c*) include some information on new subjects which have become important during the past decade, (*d*) rearrange some of the material to make it more adaptable to teaching needs, (*e*) add new illustrative examples and problems, (*f*) add new line drawings and photographs, particularly of modern tubes, and (*g*) summarize important material, often in the form of a table of data.

The more important revisions are the following:

*a*. The material on cathode-ray tubes has been completely modernized.

*b*. The material on the electron theory of matter has been rearranged into two chapters. Chapter 4 now contains a summary of the subject, and Chapter 5 supplies many details and proofs that are omitted from the former. Chapter 5 may be omitted in an introductory course in engineering electronics without disturbing the continuity of the text.

*c*. The material on the kinetic theory of gases has also been rewritten in two parts, as outlined under *b*. The second half of Chapter 8 may be omitted without breaking the continuity of the text.

*d*. The chapter on electrical discharges in gases has been extensively rewritten.

*e*. The material on rectifiers has been rearranged. Single-phase rectifiers, both full-cycle and controlled, are contained in Chapter 12. The details of polyphase rectifiers are in Chapter 14. This latter chapter may be omitted from a first course, and reserved for part of a course in industrial electronics, if desired.

*f*. The chapter on filters for rectifiers has been expanded somewhat.

*g*. The material on triodes as circuit elements has been simplified and clarified.

v

The most important added material is the following: (*a*) a detailed study of the effects of initial velocity on the motion of ions in crossed electric and magnetic fields, (*b*) the dynamical considerations of electron bunching, as in the klystron, (*c*) synchrotron and f-m cyclotron, (*d*) betatron, (*e*) semiconductors, (*f*) metallic rectifiers, (*g*) streamer theory of high-pressure spark discharges, (*h*) equipotential plots and potential profiles in triodes, tetrodes, and pentodes, (*i*) triodes at ultrahigh frequencies, (*j*) excitrons and multigrid ignitrons.

It is with considerable regret that it must be reported that Chapters 19 and 20 on audio amplifiers, which had been extensively revised and enlarged, have fallen the victims of the rising costs of book manufacture.

It is a pleasure to acknowledge the many helpful suggestions which have reached us from many sources. We are particularly indebted to Professors A. G. Hill and G. G. Harvey of the Massachusetts Institute of Technology for their notes, "An Introduction to Physical Electronics"; to Dr. R. A. Hutner and the Philips Laboratories, Inc., for Technical Report No. 25, "Concerning Some Aspects of Solid State Theory"; to Dr. J. P. Molnar of the Bell Telephone Laboratories for his detailed suggestions and comments on Chapter 10; and to Dr. E. Blade of the City College for numerous suggestions and for assistance in reading proof. We also thank E. Brenner at The City College and H. Hellerman at Syracuse University for their assistance in reading proof.

The continued assistance of Miss S. Silverstein, the secretary of the Electrical Engineering Department, and J. B. O'Farrell and R. H. Whitford of the Technology Library of The City College of New York is gratefully acknowledged.

<div style="text-align: right">

JACOB MILLMAN
SAMUEL SEELY

</div>

NEW YORK, N.Y.
SYRACUSE, N.Y.
    *April*, 1951

# PREFACE TO THE FIRST EDITION

This book is the outgrowth of a course in electronics that the authors have been giving for more than five years to electrical engineering students. The scope of the material is so broad as to make the textbook suitable for courses in which the emphasis is to be placed on fundamental theory and the physics of electronic devices as well as those courses in which the principal objective is to study the technical and engineering applications of these devices. The authors feel that both aspects of electronics are of utmost importance. Hence, either the point of view of the physicist or that of the engineer has been adopted, depending upon which is the more appropriate for the particular topic under survey. In a great many cases the dividing line between these two viewpoints is very indistinct. It is one of the objectives of the text to bring these two viewpoints even closer together. Thus both the theoretical and practical aspects of electronics are included side by side, thereby presenting a unified treatment which should be of interest to students of either physics or engineering.

It is realized, however, that because of time limitations it may not be possible to cover the entire text in one course. Although the material is presented as a coordinated whole, it is possible to omit many sections without disturbing the continuity of the development. This permits the omission of some of either the more theoretical or the more practical details, as the case may require, depending upon the emphasis desired.

Chapter 1 is intended to indicate in a slight measure the general scope of the course. It also serves to introduce terms and concepts that may arise in the development of certain topics before the complete discussion of them. The first chapter is therefore intentionally qualitative in nature.

The remainder of the text is specific and so, of necessity, is somewhat mathematical in character. In all cases, however, it is the physical problem that is kept in the foreground, and the mathematical solution is presented in such a way as to emphasize the physical aspects of the topic under consideration.

Based upon the authors' classroom experience, there have been included unique and original treatments of many of the topics that the students frequently find very difficult to grasp when presented in the more conventional manner. All chapters are profusely illustrated with diagrams to aid in an understanding of the concepts involved. Also included are many sketches or photographs of actual apparatus, and curves plotted from data taken in the laboratory to illustrate certain aspects of the theory. It has

also been found that a better understanding of tube characteristics and of the operation of tubes as circuit elements is obtained from a study of the many (*unretouched*) oscillograms included in the text. In addition, numerous examples are carried out in sufficient detail so that the student can see how the theory developed is applied to specific problems.

The problems (found at the end of each chapter, except the first) have been carefully chosen to test the student's grasp of the subject matter. Problems that simply require the substitution of numerical values in formulas have been avoided. Many problems are divided into several parts. It may be advisable to assign only parts of these at one time and to consider the remaining parts as separate problems on the same topic. The number of problems is larger than would normally be assigned in any one term, in order to provide instructors with a large choice of problems that may be assigned during successive terms without repetition. The problems are based upon practical data wherever possible in order to familiarize the student with the actual orders of magnitude involved.

It is assumed that the student is familiar with the calculus and with elementary differential equations. In order to understand the engineering applications, it is necessary that he also have a working knowledge of the basic principles of alternating-current theory, including elementary Fourier analysis. This makes possible a smooth transition from the study of circuits containing only inductances, capacitances, linear resistances, and voltage sources to the analyses of networks that contain electron tubes as well.

Because of the lack of space, the important subjects of radio-frequency amplifiers, modulators, detectors, and oscillators (except those of the relaxation type) have been omitted from the text. Most electrical engineering curricula include, following an electronics course, a course in communications, where these topics are treated.

JACOB MILLMAN
SAMUEL SEELY

NEW YORK, N.Y.
*August*, 1941

# CONTENTS

ix

# CHAPTER 1

# INTRODUCTION

THE purpose of this chapter is to give some indication of the scope of the text, to define certain terms and fundamental processes which will be discussed in detail later, and, in general, to arouse the scientific curiosity of the reader. The brief and qualitative discussions contained in this introductory chapter will bring a number of questions to mind, many of which will be answered by the detailed and quantitative analyses contained in the following chapters.

**1-1. Physical Electronics.** From the purely physical point of view, electronics might be defined as the "study of the motion of electrons in, or the interaction of electrons with, a field of force." This field of force might be a simple static electric field, such as that found in the region between two charged parallel plates, or a simple magnetic field, such as that produced at the center of a long solenoid which is carrying a d-c current. On the other hand, the fields might be the very complicated ones that exist in the interelectrode region of a multielectrode vacuum tube or the still more complex ones that exist within an atom.

Since both static and dynamic electric and magnetic fields produce forces on electrons, a large variety of possible field configurations and field sources must, theoretically at least, come under survey. Since stationary charges produce electrostatic fields and since charges moving with constant velocity (a steady current) produce steady magnetic fields, then an electron may interact with other free electrons or with the bound electrons of matter. The matter may exist in any of its states of aggregation, solid, liquid, or gaseous.

In so far as the dynamic fields are concerned, they may result, among other ways, from accelerated charges, from the application of a periodically varying potential to a pair of parallel plates, or from the presence of a periodically varying current in a solenoid. These might be slowly varying fields produced by power circuits (60 cycles per second), or the higher communication frequencies, such as those in the audio, radio, or television range. These dynamic fields might also be the combination of electric and magnetic fields that are known as *radiation fields*, whether the radiation fields are of low frequency (power frequencies), intermediate frequencies (communication frequencies), or the high frequencies of light, either visible

or invisible.  To summarize briefly, the study of physical electronics involves considerations of

1. The behavior of electrons in a space void of matter but permeated with electric and magnetic fields of force.

2. The interaction of electrons with matter, for example, other electrons and atoms.

3. The interaction of electrons with radiation.

**1-2. Engineering Electronics.**  It is the task of the engineer to develop useful devices based on the principles of physical electronics.  That he has succeeded in this endeavor is evidenced by the wide variety of fields (in some cases, industries) that depend upon electronics, *viz.*, communications, industrial control, rectification, lighting, instrumentation, etc.

The widest application of electronic devices is in the field of communications.  This includes audio, radio, television, and radar engineering.

The field of industrial electronics, which has expanded very materially during the past decade, includes a wide variety of applications.  The more important ones are rectification, for the conversion of a-c power to d-c power; inversion, from d-c power to a-c power; frequency changing, for example, from 60 to 25 cps, or vice versa; automatic controllers, for controlling the speed of motors, automatic voltage control of a-c and d-c generators; arcless "on" and "off" switching; electric-welding timing and control; lighting control; dielectric heating; induction heating; servomechanisms.

The interaction between radiation and electrons has made possible the phototube and also the gaseous luminous-discharge tube, for example, neon tubes, sodium-vapor lamps, mercury-vapor tubes, and the fluorescent lamps.  Electronics has established itself in a firm position in the field of illumination engineering both in the types of lighting elements and the devices designed for their control.  The development of control devices has proved of extreme importance in the field of theatrical and industrial lighting.

The fields of physical and electrical measurements have been revolutionized through the development of electronic instruments for use in making these measurements.  These include devices for the direct measurement of electrical quantities, such as current, voltage, charge, frequency, resistance, and also the measurement of nonelectrical quantities, such as time, speed, distance, light intensity, and color.  Electronic computing machines are of invaluable aid in solving complex mathematical problems.

**1-3. Sources of Electrons.**  According to modern theory, all matter is electrical in nature.  The atom, one of the fundamental units that enters into the structure of all matter, consists of a nucleus or central core of positive charge which contains nearly all the mass of the atom.  Surrounding this nucleus are negatively charged electrons.  The atom as a whole is

electrically neutral, so that the charge on the nucleus must be balanced by the charge of the electrons that surround it. Since all chemical substances consist of groups of these atoms bound more or less closely to one another, then all matter, whether in the solid, the liquid, or the gaseous state, is a potential source of electrons. All three states of matter are actually employed in commercial devices as the source of electrons.

The metallic state is an extremely important and common source of electrons. Metals are most generally employed in the form of a wire or ribbon filament. If such a filament contains electrons and if these electrons are relatively "free" to move about in the metal (this is true, since even a very small potential applied to the two ends of the filament will direct a stream of these electrons, a current, along the filament), then why should not these charges "leak" out of the metal of their own accord and so form a natural well of electrons? A simple answer to the question raised is possible on the basis of elementary electrostatic theory.

Suppose that an electron does tend to leave the metal. As soon as it leaves, the metal must become positively charged, because it was originally uncharged. Classically, it is said that a positive charge is induced on the metal. Since, however, opposite charges are attracted to each other, the electron will be drawn back into the metal by this force of attraction unless some external agent is applied to prevent this. The external agent must impart sufficient energy to the electron to overcome the attractive effect of the induced charge, if the electron is to be free of the metal.

Several methods exist by which sufficient energy may be supplied to the electrons in a metal in order to produce electron emission. They are as follows:

1. *Thermal.* The current that results when a difference of potential is applied across the ends of the filament will heat it. When the temperature of the metal is sufficiently high, some electrons will flow out of the surface into the surrounding space. A metal when so used is called a *thermionic cathode*, the liberation of the electrons being accomplished by the process that is known as *thermionic emission*. The process of thermionic emission may be considered to be quite analogous to the process of evaporation of molecules from the surface of a liquid. In both instances, work must be done. In the case of the liquid, this work is the latent heat of vaporization and represents the amount of energy required to change 1 g of the substance from the liquid to the gaseous state. In the case of thermionic emission, the "work function" represents the minimum amount of energy necessary to "evaporate" an electron from the metal. Practically all tubes in the communication industry depend upon thermionic emission for the liberation of electrons.

2. *Photoelectric.* Subject to certain restrictions discussed later with regard to the nature of the surface and the frequency of the impinging

light, illumination of a surface with radiation releases electrons. The energy contained in the radiant beam is used to supply the necessary energy to the electrons in order to enable them to escape from the surface of the metal. A metal when used in this manner is called a *photoelectric cathode*, the process of the liberation of the electrons being known as *photoelectric emission*.

3. *Secondary Emission.* When an electron strikes a metallic surface, it may transfer all or a part of its kinetic energy to one or more of the internal electrons, thus enabling them to escape. This process is called *secondary emission* and frequently plays an important role in electronic devices, sometimes of a deleterious nature. It is also possible to obtain electron emission from a metal by bombarding its surface with positive ions. Although this process is much less efficient than that of electron bombardment, nevertheless it plays an important role in certain gas discharges.

4. *High-field Emission.* If a very strong electric field exists at the surface of a metal, electrons may be pulled out of it. This is called *high-field* or *autoelectronic* emission.

Mercury is the only liquid that is extensively used as a source of electrons. It is used exclusively in pool-cathode rectifiers. Such mercury-pool cathodes provide sources of electrons of very high current density probably by the method of high-field emission. Rectifier assemblies which utilize such pool cathodes have power capacities in excess of 3,000 kw.

Electrons may be released from a gas in several different ways. The process in which an atom loses an electron is called *ionization*. The atom that has lost the electron is called a *positive ion*. The process of ionization may result from the following:

1. *Electron Bombardment.* A free electron in a gaseous tube may acquire enough energy from the applied potential field so that when it collides with a gas particle it may cause the removal of an electron. This process may become cumulative since it is now possible for both the emitted electron plus the original electron to collide with other gas particles and so release additional electrons.

Ionization by electron bombardment is responsible for the successful operation of gas- and vapor-filled rectifiers and gas-filled phototubes.

2. *Photoelectric Emission.* If the gas is exposed to either visible or invisible radiation, then the energy of this incident radiation may be enough, if absorbed by the atom, to dislodge some electrons. Furthermore, a photoelectric effect from the radiant energy of an excited gas may exist. These mechanisms play important roles in certain discharges.

3. *Positive-ion Bombardment.* It is possible for a positive ion to collide with a neutral gas molecule and thereby release electrons, exactly as occurred in the case of electron bombardment. However, it is found experimentally that such ionization by positive ions is negligibly small.

4. *Thermal Emission.* If the temperature of the gas is high enough, some electrons may acquire sufficient energy to enable them to become dislodged from their parent atoms. This is analogous to thermionic emission from a metal. This type of emission takes place to an appreciable extent only in a very high-pressure high-current arc.

**1-4. Utilizing the Electrons.** The control of an electron stream is accomplished with either electric fields or magnetic fields, or both. These fields perform one or both of the following functions: (1) control of *the number of electrons* that leave the region near the emitter (this is the more common method); (2) control of *the path of the electrons* after they leave the emitter. These methods of control and a few typical applications will now be discussed qualitatively.

**1-5. Diodes.** Consider a heated, isolated thermionic source situated in a vacuum. This cathode will emit electrons, most of which have very little energy upon emergence from the surface of the metal. Those electrons which first escape from the cathode will diffuse throughout the space within the container. The cathode becomes more positively charged with each escaping electron. The attraction between the positively charged cathode and the electron, forces back toward the cathode those electrons which subsequently attempt to escape. In terms of field theory, this "shutoff" action is explained by saying that, as a consequence of those electrons which are first emitted, a field is produced at the cathode in such a direction as to repel the other electrons which attempt to enter the field. The consequence of this action is that an isolated thermionic source would not emit a steady stream of electrons but would establish instead an equilibrium state in which the cathode is surrounded by a "cloud" of electrons, this being a "space charge" that will limit any further emission. A diagram illustrating this condition is given in Fig. 1-1.
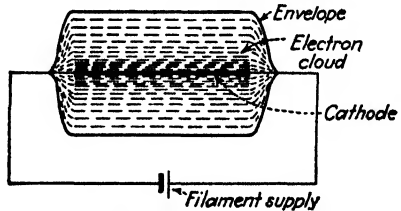


FIG. 1-1. The space-charge cloud surrounding an isolated thermionic source.

This diagram is supposed to show that the electron cloud is most dense in the neighborhood of the cathode and falls off rather rapidly with the distance from the cathode.

In order to employ such a thermionic cathode in a useful device, the repelling field at the cathode must be counterbalanced. Also, some provision must be made for controlling and collecting the electrons. The simplest way of accomplishing this is to exert an attractive force on those electrons which are first emitted from the cathode and so remove them from the region near the cathode. In practice, this is accomplished by surrounding the cathode with a second electrode, called an *anode* or *plate*,

to which a potential, positive with respect to the cathode, is applied. Such a unit is known as a *diode* and is illustrated in Fig. 1-2.

Experiment and theory show that a definite plate current $I_b$ will exist between the cathode and the plate for a given value of plate potential $E_b$. This is subject to the condition that the temperature of the cathode is sufficiently high so that the cathode may be considered as an unlimited source of electrons. Should the value of plate voltage be altered, then a new value of plate current will result. Thus the magnitude of the current may be varied by varying the plate potential. This is the first illustration of the control of electron flow by means of an electric field.
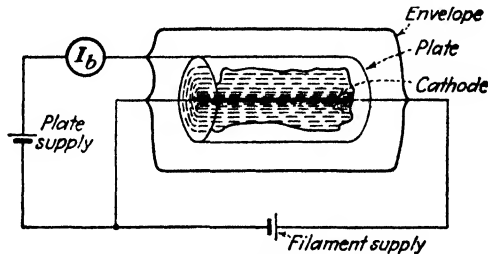


FIG. 1-2.   A thermionic diode.

Suppose now that the polarity of the plate-to-cathode supply shown in Fig. 1-2 is reversed. The direction of the field resulting from this applied potential will be such as to *repel* electrons, and so the field prevents the electrons from moving to the plate. Consequently, the plate current will fall to zero. Physically, of course, this arises from the fact that the electrons can flow only from the source to the collector within the tube.

**1-6. Diode Rectifiers.** If, instead of applying the d-c source between the cathode and the plate of the diode of Fig. 1-2 an a-c source is impressed, then, in accordance with the foregoing discussion, electrons will flow only during those half cycles when the plate is positive with respect to the cathode, and the current will be zero for the other, or "inverse," half cycle. This is one of the most important applications of diodes and is known as *rectification.*

**1-7. Phototubes.** If the thermionic cathode of the diode of Fig. 1-2 is replaced by a photoelectric cathode, *i.e.*, one that depends for the liberation of electrons on the interaction of light with the material of the cathode, then the diode becomes a phototube.

In the operation of this device the anode plays the same role that it did before; it again supplies the field that is necessary to direct the electrons away from the cathode. However, since the number of photoelectrons emitted from the cathode is generally small, a repelling electron cloud of the type discussed above will not exist. The sole function of the anode is

to collect the electrons that are emitted from the cathode. The magnitude of this current is dependent solely upon the intensity of the light falling on the cathode and is, in fact, directly proportional to it. This characteristic is important in many applications.

**1-8. Triode Amplifiers.** The introduction of a third element between the cathode and the plate of the diode by DeForest in 1907 opened up a wide field of new applications for vacuum tubes. This new electrode, called the *control grid*, consists of a wire mesh or screen which surrounds the cathode and is situated close to it. A tube of this type is known as a *triode*, and is represented schematically in Fig. 1-3. The potential applied to the grid in such a tube is usually a few volts negative with respect to
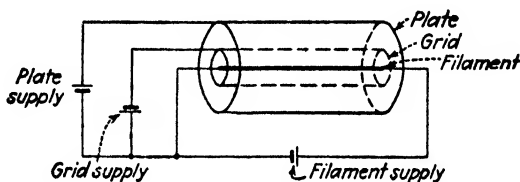


FIG. 1-3. A triode.

the cathode, whereas the plate is usually maintained several hundred volts positive with respect to the cathode. Hence, the electric field resulting from the potential of the grid is such as to help maintain a large space-charge cloud, whereas the field of the plate potential acts oppositely. An equilibrium condition will exist between the fields of the applied voltages and the resultant current.

Since the grid is a fine wire mesh that completely surrounds the cathode, and since it is usually very close to it, it furnishes partial electrostatic shielding between the electron source and the field due to the plate potential. Hence, a small change in the grid voltage will result in a relatively large change in the plate current—certainly a much larger change than would result from an equal variation in the plate voltage. Herein lies the amplifying feature of the triode. A small voltage is fed to the grid of an amplifier, and the resultant amplified variations are produced in the plate circuit.

**1-9. Gas Diodes.** A diode into which some inert gas (argon, neon, etc.) or some mercury vapor has been introduced possesses characteristics that are markedly different from those of the vacuum diode. Consider an electron that has been released from a thermionic or mercury-pool cathode and has gained enough energy from the applied anode potential field to ionize a gas molecule when it collides with it. As a result of the collision, two electrons will now be present, the original one and the electron that has been liberated by the process of ionization, and a positive ion. These two

electrons can then move in the applied field and acquire sufficient energy to cause additional ionizing collisions. This process is cumulative and will continue until "breakdown" occurs. Breakdown becomes manifest by the appearance of a luminous discharge in the tube and by the current in the external circuit. A discharge of this type, which is characterized by a low anode-to-cathode potential and a high plate current, is called an *arc*. This, however, is but one type of gas discharge phenomenon to be considered later.

**1-10. Electron-deflecting Devices.** Each of the devices already considered depends for its utility upon controlling the *effective number* of electrons leaving the cathode, by means of electric fields. A second group of devices depend for their usefulness upon the ability to control the *paths* of given electron streams by means of carefully selected electric and magnetic fields. Such devices include the cathode-ray tube, the magnetron, the cyclotron, the secondary-emission multiplier, the mass spectrograph, the betatron, and the klystron. The next two chapters will be devoted to a quantitative study of the motion of charged particles in various configurations of electric and magnetic fields, with particular reference to these devices.

### General References

HENNEY, K.: "Electron Tubes in Industry," McGraw-Hill Book Company, Inc., New York, 1937.

HORTON, J. W.: *Elec. Eng.*, **54**, 93, 1935.

Westinghouse Staff: "Industrial Electronics Reference Book," John Wiley & Sons, Inc., New York, 1948.

# CHAPTER 2

# MOTION OF CHARGED PARTICLES IN ELECTRIC AND MAGNETIC FIELDS

THE motion of a charged particle in electric and magnetic fields will be discussed, starting with simple paths and proceeding to more complex motions. First a uniform electric field will be considered, and then the analysis will be given for motions in a uniform magnetic field. This discussion will be followed in turn by the motion in parallel electric and magnetic fields, in perpendicular electric and magnetic fields, and in electric and magnetic fields making an arbitrary angle with each other. Some discussion is included of nonuniform fields.

**2-1. Charged Particles.** The charge or quantity of negative electricity of the electron has been found by numerous experiments to be $1.602 \times 10^{-19}$ coulomb. The values of many important physical constants are given in Appendix I. Some idea of the number of electrons per second that represent currents of the usual order of magnitude is readily possible. For example, since the charge per electron is $1.602 \times 10^{-19}$ coulomb, the number of electrons per coulomb is the reciprocal of this number, or approximately $6 \times 10^{18}$. Further, since a current of 1 amp is the flow of 1 coulomb/sec, then a current of only 1 $\mu\mu a$ (1 micromicroamp or $10^{-12}$ amp) represents the motion of approximately 6 million electrons per second. Yet a current of 1 $\mu\mu a$ is so small that considerable difficulty is experienced in attempting to measure it.

In addition to its charge, the electron possesses a definite mass. A direct measurement of the weight of an electron cannot be made, but the ratio of the charge to the mass $e/m$ has been determined by a number of experimenters using independent methods. The most probable present-day (1951) value for this ratio is $1.759 \times 10^{11}$ coulombs/kg. From this value of $e/m$ and the value of $e$, the charge on the electron, the mass of the electron is calculated to be $9.107 \times 10^{-31}$ kg.

The charge of a positive ion will be an integral multiple of the charge of the electron, although it is of opposite sign. For the case of singly ionized particles, the charge is equal to that of the electron. For the case of doubly ionized particles, the ionic charge is twice that of the electron.

The mass of an atom is expressed as a number that is based on the choice of the atomic weight of oxygen equal to 16. The mass of a hypothetical

atom of atomic weight unity is, by this definition, one-sixteenth that of the mass of monatomic oxygen. This has been calculated to be 1.660 $\times 10^{-27}$ kg. Hence, *in order to calculate the mass in kilograms of any atom, it is necessary only to multiply the atomic weight of the atom by* 1.660 $\times$ $10^{-27}$ *kg.* A table of atomic weights is given in Appendix III.

The radius of the electron has been estimated as $10^{-15}$ m and that of an atom as $10^{-10}$ m. These are so small that all charges will be considered as mass points in the following sections.

**2-2. The Force on Charged Particles in an Electric Field.** *The force on a unit positive charge at any point in an electric field is,* by definition, *the electric-field intensity* $\mathcal{E}$ *at that point.* Consequently, the force on a positive charge $q$ in an electric field of intensity $\mathcal{E}$ is given by $q\mathcal{E}$, the resulting force being in the direction of the electric field. Thus,

$$\mathbf{f}_q = q\mathcal{E} \qquad \text{newtons} \qquad (2\text{-}1)$$

where $q$ is in coulombs and $\mathcal{E}$ is in volts per meter. Boldface type will be employed wherever vector quantities (those having both magnitude and direction) are encountered.

The mks (meter-kilogram-second) rationalized system of units will be found most convenient for the subsequent studies. Therefore, unless otherwise stated, this system of units will be employed. Appendix IV lists the names of the most common quantities in the mks system. Conversions from the electrostatic (esu) and the electromagnetic (emu) systems of units to the mks system of units are listed in Appendix II.

In order to calculate the path of a charged particle in an electric field, the force, given by Eq. (2-1), must be related to the mass and the acceleration of the particle by Newton's second law of motion. Hence,

$$\mathbf{f}_q = q\mathcal{E} = m\frac{d\mathbf{v}}{dt} \qquad \text{newtons} \qquad (2\text{-}2)$$

where $m$ is in kilograms and $d\mathbf{v}/dt$ is in meters per second$^2$. The solution of this equation, subject to appropriate initial conditions, will give the path of the particle resulting from the action of the electric forces. If the magnitude of the charge on the electron is $e$, then the force on an electron in the field is

$$\mathbf{f} = -e\mathcal{E} \qquad \text{newtons} \qquad (2\text{-}3)$$

The minus sign denotes that the force is in the direction opposite to the field.

In investigating the motion of charged particles which are moving in externally applied force fields of electric and magnetic origin, it will be implicitly assumed that the number of particles is so small that their presence will not alter the field distribution. Furthermore, it will be as-

sumed that the vacuum in the region in which the particles move is sufficiently high so that no collisions occur with gas particles. That is, the fields alone will control the motion of the electrons.

For purposes of clarity, problems will be considered in order of increasing complexity. Where possible, use will be made of known behavior in order to simplify the analysis. This requires, of course, that the analysis of the motions in simple fields be mastered before proceeding to more complicated problems.



FIG. 2-1. Cartesian coordinate axes.

Since the electronic motion will take place, in general, in space, the motion will be specified mathematically with respect to the customary three mutually perpendicular Cartesian axes, illustrated in Fig. 2-1. The system of notation to be employed throughout the text is the following:

$x$ = position of the particle along the $X$ axis, m
$y$ = position of the particle along the $Y$ axis, m
$z$ = position of the particle along the $Z$ axis, m
$v$ = magnitude of the velocity of the particle, m/sec
$v_x, v_y, v_z$ = velocity components along the $X$, $Y$, and $Z$ axes, respectively, m/sec
$a$ = magnitude of the acceleration of the particle, m/sec$^2$
$a_x, a_y, a_z$ = components of the acceleration along the $X$, $Y$, and $Z$ axes, respectively, m/sec$^2$

Similar subscript notation will be used for other vector components. For example,

$f_y$ = component of the force along the $Y$ direction, newtons
$\mathcal{E}_z$ = component of the electric-field intensity along the $Z$ axis, volts/m

It should be kept in mind that any or all of the foregoing quantities may be functions of time.

The exact motion of a particle in a given force field cannot be determined unless the initial values of velocity and displacement are known. The term "initial" represents the value of the specified quantity at the time $t = 0$. The subscript 0 will be used to designate such initial values. For example,

$x_0$ = initial displacement of the particle along the $X$ axis
$v_{0z}$ = initial component of velocity in the $Z$ direction

In every case the path will be determined from an analysis of Eq. (2-2) subject to appropriate specified initial conditions.

This chapter will present the fundamental physical and mathematical theory of the motion of particles in electric and magnetic fields of force.

The succeeding chapter will give many important applications based upon these analyses.

**2-3. Constant Electric Field.** Suppose that an electron is situated
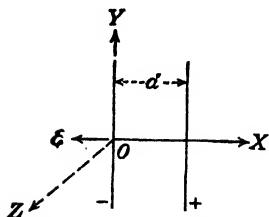


FIG. 2-2. The one-dimensional electric field between the plates of a parallel-plate capacitor.

between the two plates of a parallel-plate capacitor which are contained in an evacuated envelope, as illustrated in Fig. 2-2. A difference of potential is applied between the two plates, the direction of the electric field in the region between the two plates being as shown. If the distance between the plates is small compared with the dimensions of the plates, the electric field may be considered to be uniform, the lines of force pointing along the negative $X$ direction. That is, the only field that is present is $\mathcal{E}$ along the $-X$ axis. It is desired to investigate the characteristics of the motion, subject to the initial conditions

$$\left.\begin{array}{c} v_x = v_{0x} \\ x = x_0 \end{array}\right\} \quad \text{when } t = 0 \qquad (2\text{-}4)$$

This means that the initial velocity is chosen along $\mathcal{E}$, the lines of force.

Since there is no force along the $Y$ or $Z$ directions, Newton's law states that the acceleration along these axes must be zero. However, zero acceleration means constant velocity; and since the velocity is initially zero along these axes, the particle will not move along these directions. That is, the only possible motion is one-dimensional, and the electron moves along the $X$ axis.

Newton's law applied to the $X$ direction yields

$$e\mathcal{E} = ma_x$$

or

$$a_x = \frac{e\mathcal{E}}{m} = \text{const} \qquad (2\text{-}5)$$

This analysis indicates that the electron will move with a constant acceleration in a uniform electric field. Consequently, the problem is analogous to that of a freely falling body in the uniform gravitational field of the earth. The solution of this problem is given by the well-known expressions for the velocity and displacement, *viz.*,

$$v_x = v_{0x} + a_x t \qquad x = x_0 + v_{0x}t + \tfrac{1}{2}a_x t^2 \qquad (2\text{-}6)$$

*provided that $a_x = $ const, independent of the time.*

It is to be emphasized that, if the acceleration of the particle is not a constant but depends upon the time, then Eqs. (2-6) are no longer valid.

Under these circumstances the motion is determined by integrating the equations

$$\frac{dv_x}{dt} = a_x \quad \text{and} \quad \frac{dx}{dt} = v_x \tag{2-7}$$

These are simply the definitions of the acceleration and the velocity, respectively. Equations (2-6) follow directly from Eqs. (2-7) by integrating the latter equations subject to the conditions of a constant acceleration.

*Example.* A sinusoidal potential having a frequency of 1 million cycles per second (1 megacycle) and whose maximum value is 10 volts is applied to the plates of a parallel-plate capacitor which are 2 cm apart. If an electron is released from one plate at an instant when the applied potential is zero, find the position of the electron at any subsequent time $t$. Assume that the initial velocity of the electron is $10^6$ m/sec along the lines of force.

. *Solution.* Assume that the plates are oriented with respect to a Cartesian system of axes as illustrated in Fig. 2-2. The electric-field intensity is

$$\varepsilon = \frac{10}{0.02} \sin 2\pi ft = 500 \sin (6.28 \times 10^6 t) \quad \text{volts/m}$$

whence

$$a_x = \frac{dv_x}{dt} = \frac{f_x}{m} = \frac{e\varepsilon}{m} = 1.76 \times 10^{11} \times 500 \sin (6.28 \times 10^6 t)$$

$$= 8.80 \times 10^{13} \sin (6.28 \times 10^6 t) \quad \text{m/sec}^2$$

This becomes, upon integration,

$$v_x = -1.40 \times 10^7 \cos (6.28 \times 10^6 t) + A$$

where $A$ = constant of integration. $A$ is determined from the initial condition that

$$v_x = 10^6 \text{ m/sec} \quad \text{when } t = 0$$

Thus,

$$A = 1.50 \times 10^7 \text{ m/sec}$$

so that the velocity is given by

$$v_x = 1.50 \times 10^7 - 1.40 \times 10^7 \cos (6.28 \times 10^6 t) \quad \text{m/sec}$$

Integration with respect to $t$, subject to the condition that $x = 0$ when $t = 0$, yields

$$x = 1.50 \times 10^7 t - 2.23 \sin (6.28 \times 10^6 t) \quad \text{m}$$

**2-4. Potential.** The discussion to follow need not be restricted to uniform fields, but $\varepsilon_x$ may be a function of distance. However, it will be assumed that $\varepsilon_x$ is *not a function of time.* Then, from Newton's second law

$$-\frac{e\varepsilon_x}{m} = \frac{dv_x}{dt}$$

Multiply this equation by $dx = v_x \, dt$, and integrate. This leads to

$$-\frac{e}{m} \int_{x_0}^{x} \varepsilon_x \, dx = \int_{v_{0x}}^{v_x} v_x \, dv_x \tag{2-8}$$

The definite integral

$$\int_{x_0}^{x} \mathcal{E}_x \, dx$$

is an expression for the work done by the field in carrying a unit positive charge from the point $x_0$ to the point $x$.

By definition, *the potential V of point x with respect to point $x_0$ is the work done* **against** *the field in taking a unit positive charge from $x_0$ to $x$.* Thus,*

$$V \equiv -\int_{x_0}^{x} \mathcal{E}_x \, dx \qquad \text{volts} \tag{2-9}$$

By virtue of Eq. (2-9), Eq. (2-8) integrates to

$$eV = \tfrac{1}{2}m(v_x{}^2 - v_{0x}{}^2) \qquad \text{joules} \tag{2-10}$$

This shows that an electron that has "fallen" through a certain difference of potential $V$ in going from point $x_0$ to point $x$ has acquired a specific value of kinetic energy and velocity, independent of the form of the variation of the field distribution between these points and depending only upon the magnitude of the potential difference $V$.

Although this derivation supposes that the field has only one component, namely, $\mathcal{E}_x$ along the $X$ axis, the final result given by Eq. (2-10) is simply a statement of the law of conservation of energy. This is known to be valid even if the field is multidimensional. This result is extremely important in electronic devices. Stated in its most general form, Eq. (2-10) becomes

$$qV_{BA} = \tfrac{1}{2}mv_A{}^2 - \tfrac{1}{2}mv_B{}^2 \qquad \text{joules} \tag{2-11}$$

where $q$ is the charge in coulombs, $V_{BA}$ is the potential in volts of any point $B$ with respect to any other point $A$ in an arbitrary electrostatic field, and $v_A$ and $v_B$ are the corresponding initial and final speeds in meters per second at the points $A$ and $B$, respectively. By definition, *the potential energy between two points equals the potential multiplied by the charge in question.* Thus, the left-hand side of Eq. (2-11) is the *rise* in *potential energy* from $A$ to $B$. The right-hand side represents the *drop* in *kinetic energy* from $A$ to $B$. Thus, Eq. (2-11) states that the rise in potential energy equals the drop in kinetic energy, which is equivalent to the statement that the total energy remains unchanged.

It must be emphasized that *Eq. (2-11) is not valid if the field varies with time.*

If the particle is an electron, then $-e$ must be substituted for $q$. If the electron starts at rest, then its final speed $v$, as given by Eq. (2-11) with

---

* The symbol $\equiv$ is used to designate "equal to by definition."

$v_A = 0$, $v_B = v$, and $V_{BA} = V$, is

$$v = \left(\frac{2eV}{m}\right)^{\frac{1}{2}} \quad \text{m/sec} \tag{2-12}$$

or

$$v = 5.93 \times 10^5 V^{\frac{1}{2}} \quad \text{m/sec} \tag{2-13}$$

Thus if an electron "falls" through a difference of only 1 volt, its final speed is $5.93 \times 10^5$ m/sec, or approximately 370 miles/sec. Despite this tremendous speed, the electron possesses very little kinetic energy because of its minute mass.

It must be emphasized that *Eq. (2-13) is valid only for an electron starting at rest.* If the electron does not have zero initial velocity or if the particle involved is not an electron, then the more general formula [Eq. (2-11)] must be used.

**2-5. The ev Unit of Energy.** The joule is the unit of energy in the mks system. In some engineering power problems this unit is very small, and a factor of $10^3$ or $10^6$ is introduced to convert from watts (joules per second) to kilowatts or megawatts, respectively. However, in other problems, the joule is too large a unit, and a factor of $10^{-7}$ is introduced to convert from joules to ergs. For a discussion of the energies involved in electronic devices even the erg is much too large a unit. This is not to be construed to mean that only minute amounts of energy can be obtained from electron tubes. It is true that each electron possesses a tiny amount of energy, but as previously pointed out (see Sec. 2-1), an enormous number of electrons is involved even in a small current so that considerable power may be represented.

The new unit of work or energy, called the *electron volt* (ev), is defined as follows:

$$1 \text{ ev} \equiv 1.60 \times 10^{-19} \text{ joule}$$

Of course, any type of energy, whether it be electrical, mechanical, thermal, etc., may be expressed in electron volts.

The name "electron volt" arises from the fact that if an electron falls through a potential of one volt its kinetic energy will increase by the decrease in potential energy or by

$$eV = (1.60 \times 10^{-19} \text{ coulomb})(1 \text{ volt}) = 1.60 \times 10^{-19} \text{ joule} \equiv 1 \text{ ev}$$

However, as mentioned above, the electron-volt unit may be used for any type of energy and is not restricted to problems involving electrons.

The abbreviations Mev and Bev are used to designate 1 million and 1 billion electron volts, respectively.

It is common practice to designate energies in terms of "volts" although it must be clearly understood that the terms "volt" and "electron volt"

are being used synonymously. For example, the phrase "a 5-volt mercury ion" means simply that the ion has a kinetic energy of $5 \times 1.60 \times 10^{-19}$ joule. Similarly, the phrase "0.1-ev thermal energy" means an amount of thermal energy of $0.1 \times 1.60 \times 10^{-19}$ joule.

**2-6. Relationship between Field Intensity and Potential.** The definition of potential is expressed mathematically by Eq. (2-9). If the electric field is uniform, the integral may be evaluated to the form

$$-\int_{x_0}^{x} \mathcal{E}_x dx = -\mathcal{E}_x(x - x_0) = V \qquad \text{volts}$$

which shows that the electric-field intensity resulting from an applied potential difference $V$ between the two plates of the capacitor illustrated in Fig. 2-2 is given by

$$\mathcal{E}_x = \frac{-V}{x - x_0} = -\frac{V}{d} \qquad \text{volts/m} \qquad (2\text{-}14)$$

where $d$ is the distance between plates, in meters.

In the general case where the field may vary with the distance, this equation is no longer true and the correct result is obtained by differentiating Eq. (2-9). It is

$$\mathcal{E}_x = -\frac{dV}{dx} \qquad (2\text{-}15)$$

The minus sign shows that the electric field is directed from the region of higher potential to the region of lower potential.

**2-7. Two-dimensional Motion.** Suppose that an electron enters the region between the two parallel plates of a parallel-plate capacitor which



Fig. 2-3. Two-dimensional motion in a uniform electric field.

are oriented as shown in Fig. 2-3, with an initial velocity in the $+X$ direction. It will again be assumed that the electric field between the plates is uniform. Then, as chosen, the electric field $\mathcal{E}$ is in the direction of the $-Y$ axis, no other fields existing in this region.
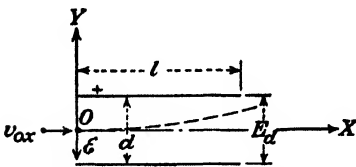
The motion of the particle is to be investigated, subject to the initial conditions

$$\left.\begin{array}{ll} v_x = v_{0x} & x = 0 \\ v_y = 0 & y = 0 \\ v_z = 0 & z = 0 \end{array}\right\} \qquad \text{when } t = 0 \qquad (2\text{-}16)$$

Since there is no force in the $Z$ direction, the acceleration in that direction is zero. Hence, the component of velocity in the $Z$ direction remains con-

stant. Since the initial velocity in this direction is assumed to be zero, the motion must take place entirely in one plane, the plane of the paper.

For a similar reason, the velocity along the $X$ axis remains constant and equal to $v_{0x}$. That is,

$$v_x = v_{0x}$$

from which it follows that $\hspace{4cm}$ (2-17)

$$x = v_{0x}t$$

On the other hand, a constant acceleration exists along the $Y$ direction, and the motion is given by Eqs. (2-6), with the variable $x$ replaced by $y$.

$$v_y = a_y t \qquad y = \tfrac{1}{2}a_y t^2 \tag{2-18}$$

where

$$a_y = -\frac{e\mathcal{E}_y}{m} = \frac{eE_d}{md} \tag{2-19}$$

and where the potential across the plates is $V = E_d$. These equations indicate that in the region between the plates the electron is accelerated upward, the velocity component $v_y$ varying from point to point, whereas the velocity component $v_x$ remains unchanged in the passage of the electron between the plates.

The path of the particle with respect to the point $O$ is readily determined by combining Eqs. (2-17) and (2-18), the variable $t$ being eliminated. This leads to the expression

$$y = \left(\frac{1}{2}\frac{a_y}{v_{0x}{}^2}\right)x^2 \tag{2-20}$$

which shows that the particle moves in a parabolic path in the region between the plates.

*Example.* A source introduces 500-volt potassium ions (atomic weight 40) into the region between two parallel plates oriented as shown in Fig. 2-4. The ions leave point $A$ at an angle of 30 deg with the horizontal. There is a fixed voltage $E_d$ across the plates, and the top plate is positive. Discuss the resultant motion. In particular, answer the following:

*a.* How long will it take an ion to reach point $B$ on the negative plate at a distance of 10 cm away from $A$?

*b.* What must be the value of $E_d$, if the ion is to pass through $B$?

*c.* What is the highest point of ascent of the ion?
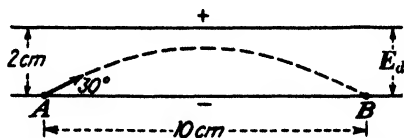
*Solution.* From the above theory, it



FIG. 2-4. The path of a charged particle in a uniform electric field is a parabola, provided that there exists a component of velocity normal to the field.

follows that the path is a parabola as shown by the dashed curve in Fig. 2-4. This problem is analogous to the firing of a gun in the earth's gravitational field. The bullet will travel in a parabolic path, first rising because of the muzzle velocity of the gun and then falling because of the downward attractive force of the earth. Because of this

analogy, the study of the motion of charged particles in a field of force is called *electron ballistics*. The source of the charged particles is called an "electron gun" or an "ion gun."

The initial components of velocity will now be found. The mass of the potassium ion is $40 \times 1.66 \times 10^{-27} = 6.64 \times 10^{-26}$ kg (see Sec. 2-1). The initial energy of the ion is $500 \times 1.60 \times 10^{-19} = 8.00 \times 10^{-17}$ joule. Hence the initial velocity is given by

$$\tfrac{1}{2}mv_0{}^2 = 8.00 \times 10^{-17} \text{ joule}$$

$$v_0{}^2 = \frac{2 \times 8.00 \times 10^{-17}}{6.64 \times 10^{-26}} = 2.41 \times 10^9 \text{ m}^2/\text{sec}^2$$

$$v_0 = 4.91 \times 10^4 \text{ m/sec}$$

$$v_{0x} = v_0 \cos 30° = 4.91 \times 10^4 \times 0.866 = 4.26 \times 10^4 \text{ m/sec}$$

$$v_{0y} = v_0 \sin 30° = 4.91 \times 10^4 \times 0.500 = 2.46 \times 10^4 \text{ m/sec}$$

*a.* Since the horizontal velocity is constant, the time $t_1$ to travel a distance $x = 10$ cm $= 0.10$ m is

$$t_1 = \frac{x}{v_{0x}} = \frac{0.1}{4.26 \times 10^4} = 2.35 \times 10^{-6} \text{ sec}$$

*b.* If the ion passes through point $B$, then its vertical displacement is zero. The vertical motion is one of constant acceleration of *magnitude* $a_y$ downward with an initial velocity $v_{0y}$ upward. Hence

$$v_y = v_{0y} - a_y t$$

and

$$y = v_{0y}t - \tfrac{1}{2}a_y t^2$$

If

$$y = 0 \quad \text{for } t = t_1$$

then

$$v_{0y}t_1 - \tfrac{1}{2}a_y t_1{}^2 = 0$$

or

$$a_y = \frac{2v_{0y}}{t_1} = \frac{2 \times 2.46 \times 10^4}{2.35 \times 10^{-6}} = 2.09 \times 10^{10} \text{ m/sec}^2$$

But

$$a_y = \frac{q\mathcal{E}_y}{m} = \frac{qE_d}{md}$$

where $q$ is the charge on the ion (which is equal to the electronic charge for a singly ionized atom) and $d$ is the separation of the plates.
Hence

$$E_d = \frac{mda_y}{q} = \frac{6.64 \times 10^{-26} \times 0.02 \times 2.09 \times 10^{10}}{1.60 \times 10^{-19}}$$

$$= 173.5 \text{ volts}$$

*c.* At the highest point of ascent, the ion is traveling neither up nor down, whence $v_y = 0$. The time $t_2$ at which this takes place is given by

$$v_y = 0 = v_{0y} - a_y t_2$$

or

$$t_2 = \frac{v_{0y}}{a_y} = \frac{2.46 \times 10^4}{2.09 \times 10^{10}} = 1.18 \times 10^{-6} \text{ sec}$$

Note that $t_2 = t_1/2$. This means that the time of ascent is half the total time, or that the ion takes the same time traveling upward as it does traveling downward.

The maximum vertical displacement $y_m$ is given by

$$y_m = v_{0y}t_2 - \tfrac{1}{2}a_y t_2^2$$
$$= (2.46 \times 10^4)(1.18 \times 10^{-6}) - \tfrac{1}{2} \times 2.09 \times 10^{10} \times (1.18 \times 10^{-6})^2$$
$$= 2.90 \times 10^{-2} - 1.45 \times 10^{-2} = 1.45 \times 10^{-2} \text{ m} = 1.45 \text{ cm}$$

The same result may be obtained by using the theorem of the conservation of energy. If the potential energy is taken as zero at point $A$, it is $y_m \mathcal{E}_y q$ at the highest point, or

$$y_m \frac{E_d}{d} q = \frac{173.5 \times 1.60 \times 10^{-19}}{0.02} y_m = 1.39 \times 10^{-15} y_m$$

This must equal the change in kinetic energy. Since the horizontal component of velocity is unchanged and since there is no vertical component of velocity at the highest point, then the change in kinetic energy just equals

$$\tfrac{1}{2}mv_{0y}^2 = \tfrac{1}{2}(6.64 \times 10^{-26})(2.46 \times 10^4)^2 = 2.02 \times 10^{-17}$$
$$y_m = \frac{2.02 \times 10^{-17}}{1.39 \times 10^{-15}} = 1.45 \times 10^{-2} \text{ m} = 1.45 \text{ cm}$$

as above.

## 2-8. Relativistic Variation of Mass with Velocity.

The theory of relativity postulates an equivalence of mass and energy according to the relationship

$$W = mc^2 \qquad (2\text{-}21)$$

where $W$ is the energy in joules, $m$ is the mass in kilograms, and $c$ is the velocity of light in vacuum, in meters per second. According to this theory, the mass of a particle will increase with its energy and hence with its speed. This relationship between mass and speed is now to be derived.

If an electron starts at the point $A$ with zero velocity and reaches the point $B$ with a velocity $v$, then the increase in energy of the particle must be given by the expression $eV$ joules, where $V$ is the difference of potential between the points $A$ and $B$. Its total energy must then be, according to Eq. (2-21),

$$mc^2 = m_0c^2 + eV \qquad \text{joules} \qquad (2\text{-}22)$$

where $m_0c^2$ is the energy possessed at the point $A$. The quantity $m_0$ is known as the *rest mass* or the *electrostatic mass* of the particle and is a constant independent of the velocity; $m$ is the total mass of the particle, which will be shown to be a function of the velocity.

By differentiating this expression,

$$c^2 \, dm = e \, dV = e \frac{dV}{dx} dx = -e\mathcal{E} \, dx$$

or

$$c^2 \, dm = f \, dx \qquad (2\text{-}23)$$

according to Eqs. (2-15) and (2-3).   Newton's second law in its most general form states that the force is the rate of change of momentum, and not simply the product of the mass and the acceleration.   That is,

$$f = \frac{d}{dt}(mv)$$

But remembering that $v = dx/dt$, Eq. (2-23) now becomes

$$c^2\, dm = \frac{d}{dt}(mv)\, dx = v\, d(mv)$$

Upon performing the indicated differentiation, there results

$$c^2\, dm = mv\, dv + v^2\, dm$$

or

$$(c^2 - v^2)\, dm = mv\, dv$$

Upon separating variables, it follows that

$$\frac{dm}{m} = \frac{v\, dv}{c^2 - v^2} \tag{2-24}$$

By integrating this expression, subject to the condition that $m = m_0$ when $v = 0$, there results finally

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \tag{2-25}$$

This result, which was originally derived by Lorentz and then by Einstein as a consequence of the theory of special relativity, predicts an increasing mass with an increasing velocity, the mass approaching an infinite value as the velocity of the particle approaches the velocity of light.   The variation of mass with velocity predicted by this expression has been verified experimentally by several independent methods.[1]   From Eqs. (2-22) and (2-25), the decrease in potential energy, or, equivalently, the increase in kinetic energy, is

$$eV = mc^2 - m_0c^2 \tag{2-26}$$

or

$$eV = m_0c^2 \left( \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right) \tag{2-27}$$

This expression enables one to find the velocity of an electron after it has fallen through any potential difference $V$.   By defining the quantity $v_N$ as the velocity that would result if the relativistic variation in mass were neglected, i.e.,

$$v_N \equiv \sqrt{\frac{2eV}{m_0}} \tag{2-28}$$

then Eq. (2-27) can be solved for $v$, the true velocity of the particle. The result is

$$v = c \left[ 1 - \frac{1}{(1 + v_N{}^2/2c^2)^2} \right]^{\frac{1}{2}} \tag{2-29}$$

This expression looks imposing at first glance. It should, of course, reduce to $v = v_N$ for small velocities. That it does so is seen by applying the binomial expansion to Eq. (2-29). The result becomes

$$v = v_N \left( 1 - \frac{3}{8} \frac{v_N{}^2}{c^2} + \cdots \right) \tag{2-30}$$

From this expression, it is seen that, if the speed of the particle is much less than the speed of light, the second and all subsequent terms in the expansion can be neglected, and then $v = v_N$, as it should. This equation also serves as a criterion to determine whether the simple classical expression or the more formidable relativistic one must be used in any particular case. For example, if the speed of the electron is one-tenth of the speed of light, Eq. (2-30) shows that an error of only three-eighths of 1 per cent will result if the speed is taken as $v_N$ instead of $v$. Consequently, the relativistic expression will be used only for speeds exceeding one-tenth that of light.

Since the voltage is usually known, and it is the velocity that is to be deduced therefrom, the foregoing criterion may be expressed in terms of potential. For an electron, the potential difference through which the particle must fall in order to attain a velocity of $0.1c$ is readily found to be 2,560 volts. Thus, if an electron falls through a potential in excess of about 3 kv, the relativistic corrections must be applied. If the particle under question is not an electron, the value of the nonrelativistic velocity is first calculated. If this is greater than $0.1c$, then the calculated value of $v_N$ must be substituted in Eq. (2-29) and the true value of $v$ then calculated. In cases where the speed is not too great, the simplified expression (2-30) may be used.

The accelerating potential in high-voltage cathode-ray tubes is sufficiently high to require that relativistic corrections be made in order to calculate the velocity and mass of the particle. Other devices employing potentials that are high enough to require these corrections are X-ray tubes, the cyclotron, the betatron, and other particle-accelerating machines. Unless specifically stated otherwise, nonrelativistic conditions will be assumed in what follows.

**2-9. Force in a Magnetic Field.** To investigate the force on a moving charge in a magnetic field, the well-known "motor law" is recalled. It has been verified by experiment that, if a conductor of length $L$ meters, carry-

ing a current of $I$ amperes, is situated in a magnetic field of intensity $B$ webers per square meter,* the force acting on this conductor is $BIL$ newtons. This assumes that the directions of $I$ and $B$ are perpendicular to each other. The direction of this force is perpendicular to the plane of $I$ and $B$ and has
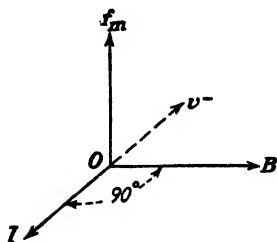


the direction of advance of a right-handed screw which is placed at $O$ and is rotated from $I$ to $B$ through 90 deg, as illustrated in Fig. 2-5. *If $I$ and $B$ are not perpendicular to each other, then only the component of $I$ perpendicular to $B$ contributes to the force.* This is a simple way of finding the direction of the force, which is equivalent to the well-known left-hand rule of Fleming.

Fig. 2-5. Pertaining to the determination of the direction of the force $f_m$ on a charged particle in a magnetic field.

Some caution must be exercised with regard to the meaning of Fig. 2-5. If the particle under consideration is a positive ion, then $I$ is to be taken along the direction of its motion. This is so because the conventional direction of the current is taken in the direction of flow of positive charge. Since the current is generally due to the flow of electrons, the direction of $I$ is to be taken as opposite to the direction of the motion of the electrons. If, therefore, a negative charge moving with a velocity $v^-$ is under consideration, one must first draw $I$ antiparallel to $v^-$ as shown and then apply the "direction rule."



Fig. 2-6. Pertaining to the determination of the magnitude of the force on a charged particle in a magnetic field.

If $N$ electrons are contained in a length $L$ of conductor (see Fig. 2-6) and if it takes an electron a time $T$ seconds to travel a distance of $L$ meters in the conductor, then the total number of electrons passing through any cross section of wire in unit time is $N/T$. Thus the total charge per second passing any point, which is the current, by definition, is

$$I = \frac{Ne}{T} \quad \text{amp} \tag{2-31}$$

The force on a length $L$ meters (or the force on the $N$ conduction charges contained therein) is

$$BIL = \frac{BNeL}{T} \quad \text{newtons} \tag{2-32}$$

* One weber per square meter equals $10^4$ gauss. A unit of more practical size in most applications is the milliweber per square meter, which equals 10 gauss. Other conversion factors are given in Appendix II.

Furthermore, since $L/T$ is the average, or *drift*, speed $v$ meters per second of the electrons, then the force per electron is

$$f_m = evB \qquad \text{newtons} \tag{2-33}$$

The subscript $m$ indicates that the force is of magnetic origin. To summarize: *The force on a negative charge e (coulombs) moving with a component of velocity $v^-$ (meters per second) normal to a field B (webers per square meter) is given by $ev^-B$ (newtons) and is in a direction perpendicular to the plane of $v^-$ and B, as noted in Fig.* 2-5.

**2-10. Current Density.** Before proceeding with the discussion of possible motions of charged particles in a magnetic field, it is convenient to introduce the concept of current density. This concept will be very useful in many later applications. By definition, the current density, denoted by the symbol $J$, is the current per unit area of the conducting medium. That is, assuming a uniform current distribution,

$$J \equiv \frac{I}{A} \qquad \text{amp/m}^2 \tag{2-34}$$

where $A$ is the cross-sectional area of the conductor. This becomes by Eq. (2-31),

$$J = \frac{Ne}{TA}$$

But it has already been pointed out that $T = L/v$. Then

$$J = \frac{Nev}{LA} \tag{2-35}$$

From Fig. 2-6 it is evident that $LA$ is simply the volume containing the $N$ electrons, and so the total charge $Ne$. Hence $Ne/LA$ is the electron charge per unit volume, which is denoted by the symbol $\rho$. Thus

$$\rho \equiv \frac{Ne}{LA} \qquad \text{coulombs/m}^3 \tag{2-36}$$

and Eq. (2-35) reduces to

$$J = \rho v \qquad \text{amp/m}^2 \tag{2-37}$$

where $\rho$ is in coulombs per cubic meter and $v$ is in meters per second.

This derivation is independent of the form of the conducting medium. Consequently, Fig. 2-6 does not necessarily represent a wire conductor. It may represent equally well a portion of a gaseous-discharge tube or a volume element in the space-charge cloud of a vacuum tube. Furthermore, neither $\rho$ nor $v$ need be constant but may vary from point to point in space or may vary with time. Numerous occasions will arise later in the text when reference will be made to Eq. (2-37).

**2-11. Motion in a Magnetic Field.** The path of a charged particle that is moving in a magnetic field will be investigated. Consider an electron to be placed in the region of the magnetic field. If the particle is at rest, $f_m = 0$ and the particle remains at rest. If the initial velocity of the particle is along the line of the magnetic flux, then there is no force acting on the particle, in accordance with the rule associated with Eq. (2-33). Hence, *a particle whose initial velocity has no component normal to a uniform magnetic field will continue to move with constant speed along the lines of flux.*

Now consider an electron moving with a speed $v_0$ to enter a constant uniform magnetic field normally as shown in Fig. 2-7. Since the force $f_m$ is perpendicular to **v** and so to the motion at every instant, then *no work is done on the electron.* This means that its kinetic energy is not increased, and so its speed remains unchanged. Further, since **v** and **B** are each constant in magnitude, then $f_m$ is constant in magnitude and perpendicular to the direction of motion of the particle. This type of force results in motion in a circular path with constant speed. It is analogous to the problem of a mass tied to a rope and twirled around with constant speed. The force (which is the tension in the rope) remains constant in magnitude and is always directed toward the center of the circle, and so is normal to the motion.
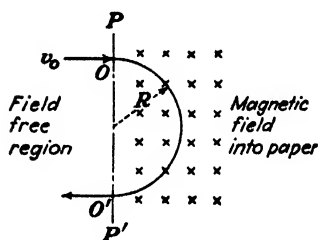
FIG. 2-7. Circular motion of electron in transverse magnetic field.

To find the radius of the circle, it is recalled that a particle moving in a circular path with a constant speed $v$ has an acceleration toward the center of the circle of magnitude $v^2/R$, where $R$ is the radius of the path. Then

$$\frac{mv^2}{R} = evB$$

from which

$$R = \frac{mv}{eB} \qquad \text{m} \tag{2-38}$$

The corresponding angular velocity is given by

$$\omega = \frac{v}{R} = \frac{eB}{m} \qquad \text{rad/sec} \tag{2-39}$$

The time for one complete revolution, called the "period," is

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{eB} \qquad \text{sec} \tag{2-40}$$

For an electron, this reduces to

$$T = \frac{3.57 \times 10^{-11}}{B} \quad \text{sec} \tag{2-41}$$

In these equations, $e/m$ is in coulombs per kilogram and $B$ in webers per square meter.

It is noticed that the radius of the path is directly proportional to the speed of the particle. Further, *the period and the angular velocity are independent of speed or radius.* This means, of course, that faster moving particles will traverse larger circles in the same time that a slower particle moves in its smaller circle. This very important result is the basis of operation of numerous devices, for example, the cyclotron and magnetic focusing apparatus.

*Example.* Calculate the deflection of a cathode-ray beam caused by the earth's magnetic field. Assume that the tube axis is so oriented that it is normal to the field,
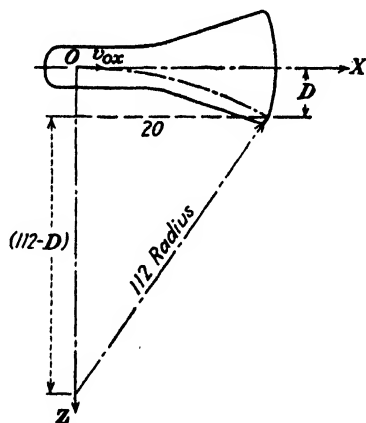


Fig. 2-8. The circular path of an electron in a cathode-ray tube resulting from the earth's transverse magnetic field. This figure is not drawn to scale.

the strength of which is 0.6 gauss. The anode potential is 400 volts; the anode-screen distance is 20 cm (see Fig. 2-8).

*Solution.* According to Eq. (2-13), the velocity of the electrons will be

$$v_{0x} = 5.93 \times 10^5 \sqrt{400} = 1.19 \times 10^7 \text{ m/sec}$$

Since 1 weber/m² $= 10^4$ gauss, then $B = 6 \times 10^{-5}$ weber/m². From Eq. (2-38) the radius of the circular path is

$$R = \frac{v_{0x}}{(e/m)B} = \frac{1.19 \times 10^7}{1.76 \times 10^{11} \times 6 \times 10^{-5}} = 1.12 \text{ m} = 112 \text{ cm}$$

Furthermore, it is evident from the geometry of Fig. 2-8 that (in centimeters)

$$112^2 = (112 - D)^2 + 20^2$$

from which it follows that

$$D^2 - 224D + 400 = 0$$

The evaluation of $D$ from this expression yields the value $D = 1.8$ cm.

This example indicates that the earth's magnetic field can have a large effect on the position of the cathode-beam spot in a low-voltage cathode-ray tube. If the anode voltage is higher than the value used in this example or if the tube is not oriented normal to the field, the deflection will be less than that calculated. In any event, this calculation indicates the advisability of carefully shielding a cathode-ray tube from stray magnetic fields.

**2-12. Magnetic Focusing.** As another application of the theory developed in the previous section, one method of measuring $e/m$ will be discussed. The essentials of a cathode-ray tube needed for this purpose are shown in Fig. 2-9. The hot cathode $K$ emits electrons which are accelerated
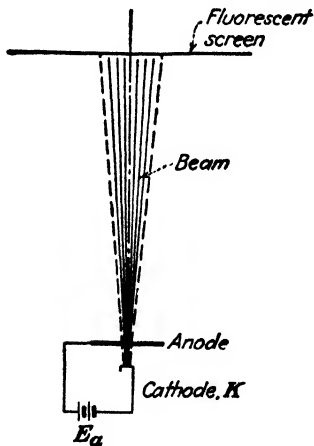


FIG. 2-9. The cathode, anode, and fluorescent screen of a cathode-ray tube. The unfocused electron beam is shown.
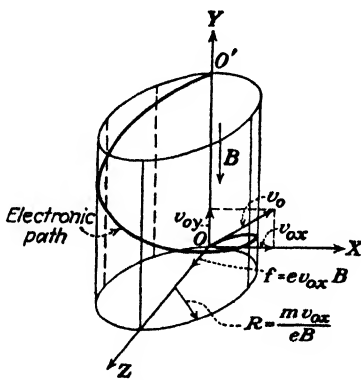
FIG. 2-10. The helical path of an electron introduced at an angle (not 90 deg) with a constant magnetic field.

toward the anode by the potential $E_a$. Those electrons which are not collected by the anode pass through the tiny anode hole and strike the end of the glass envelope. This has been coated with a material that fluoresces when bombarded by electrons. Thus the position where the electrons strike the screen are made visible to the eye. A more detailed discussion of the cathode-ray tube will be given in the next chapter.

Imagine that the cathode-ray tube is placed in a constant longitudinal magnetic field, the axis of the tube coinciding with the direction of the

magnetic field.  A magnetic field of the type here considered is obtained
through the use of a long solenoid, the tube being placed within the coil.
An inspection of Fig. 2-10 reveals the motion.  The $Y$ axis represents the
axis of the cathode-ray tube.  The origin $O$ is the point at which the elec-
trons emerge from the anode.  The velocity at the origin is $v_0$, the initial
transverse velocity due to the mutual repulsion of the electrons being $v_{0x}$.
It will now be shown that the resulting motion is a helix, as illustrated.

The electronic motion can most easily be analyzed by resolving the
velocity into two components $v_y$ and $v_\theta$ along and transverse to the mag-
netic field, respectively.  Since the force is perpendicular to $\mathbf{B}$, then there
is no acceleration in the $Y$ direction.  *Hence $v_y$ is constant and equal to $v_{0y}$.*
A force $ev_\theta B$ normal to the path will exist, resulting from the transverse
velocity.  *This force gives rise to circular motion*, the radius of the circle
being $mv_\theta/eB$, with $v_\theta$ a constant, and equal to $v_{0x}$.  *The resultant path is
a helix* whose axis is parallel to the $Y$ axis and displaced from it by a dis-
tance $R$ along the $Z$ axis, as illustrated.

The pitch of the helix, which is defined as the distance traveled along
the direction of the magnetic field in one revolution, is given by

$$p = v_{0y}T$$

where $T$ is the period, or the time for one revolution.  It follows from
Eq. (2-40) that

$$p = \frac{2\pi m}{eB} v_{0y} \qquad \text{m} \qquad (2\text{-}42)$$

If the electron beam is defocused, then a smudge is seen on the screen
when the applied magnetic field is zero.  This means that the various elec-
trons in the beam pass through the anode hole with different transverse
velocities $v_{0x}$ and so strike the screen at different points (see Fig. 2-9).  This
accounts for the appearance of a broad, faintly illuminated area instead
of a bright point on the screen.  As the magnetic field is increased from
zero, the electrons will move in helixes of different radii, since the velocity
$v_{0x}$ that controls the radius of the path will be different for different elec-
trons.  However, the period or the time to trace out the path is independ-
ent of $v_{0x}$, and so the period will be the same for all electrons.  If, then, the
distance from the anode to the screen is made equal to one pitch, all of the
electrons will be brought back to the $Y$ axis (the point $O'$ in Fig. 2-10)
since they all will have made just one revolution.  Under these conditions
an image of the anode hole will be observed on the screen.

As the field is increased from zero, the smudge on the screen resulting
from the defocused beam will contract and will become a tiny sharp spot
(the image of the anode hole) when a critical value of the field is reached.
This critical field is that which makes the pitch of the helical path just

equal to the anode-screen distance, as discussed above. By continuing to increase the strength of the field beyond this critical value, the pitch of the helix decreases, and the electrons travel through more than one complete revolution. The electrons then strike the screen at various points so that a defocused spot is again visible. A magnetic-field strength will ultimately be reached at which the electrons make two complete revolutions in their path from the anode to the screen, and once again the spot will be focused on the screen. This process may be continued, numerous foci being obtainable. In fact, the current rating of the solenoid is the factor that generally furnishes a practical limitation to the order of the focus.

The foregoing considerations may be generalized in the following way: If the screen is perpendicular to the $Y$ axis at a distance $L$ from the point of emergence of the electron beam from the anode, then, for an anode-cathode potential equal to $E_a$, the electron beam will come to a focus at the center of the screen provided that $L$ is an integral multiple of $p$. Under these conditions, Eq. (2-42) may be rearranged to read

$$\frac{e}{m} = \frac{8\pi^2 E_a n^2}{L^2 B^2} \qquad \text{coulombs/kg} \qquad (2\text{-}43)$$

where $n$ is an integer representing the order of the focus. It is assumed, in this development, that $eE_a = \frac{1}{2}mv_{0y}^2$ or that the only effect of the anode potential is to accelerate the electron along the tube axis. This implies that the transverse velocity $v_{0x}$, which is variable and unknown, is negligible in comparison with $v_{0y}$. This is a justifiable assumption.

This arrangement was suggested by Busch [2] and has been used [3] to measure the ratio $e/m$ for electrons very accurately.

**2-13. Parallel Electric and Magnetic Fields.** Consider the case where both electric and magnetic fields exist simultaneously, the fields being in the same or in opposite directions. If the initial velocity of the electron either is zero or is directed along the fields, then *the magnetic field exerts no force on the electron* and the resultant motion depends solely upon the electric-field intensity $\mathcal{E}$. If $\mathcal{E}$ is constant, the electron will move in a direction parallel to the fields with a constant acceleration. If the fields are chosen as in Fig. 2-11, the complete motion is specified by

$$v_y = v_{0y} - a_y t \qquad y = v_{0y}t - \tfrac{1}{2}a_y t^2 \qquad (2\text{-}44)$$

where $a_y = e\mathcal{E}/m$ is the magnitude of the acceleration. The negative sign results from the fact that the direction of the acceleration of an electron is opposite to the direction of the electric-field intensity $\mathcal{E}$.

If, initially, a component of velocity $v_{0x}$ perpendicular to the magnetic field exists, this component, together with the magnetic field, will give rise to circular motion, the radius of the circular path being independent of $\mathcal{E}$.

However, because of the electric field $\mathcal{E}$, the velocity along the field changes with time.  Consequently, the resulting path is helical with a pitch that changes with the time.  That is, the distance traveled along the $Y$ axis per revolution increases with each revolution.
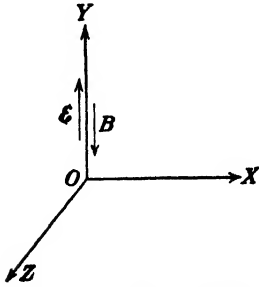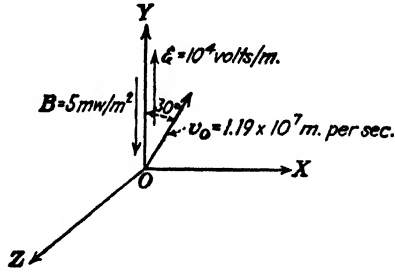


FIG. 2-11.  Parallel electric and magnetic fields.

FIG. 2-12.  A problem illustrating helical electronic motion.

*Example.*  Given a uniform electric field of $10^4$ volts/m parallel to and opposite in direction to a magnetic field of 5 milliwebers/m². An electron gun, directed at an angle of 30 deg with the direction of the electric field, introduces 400-volt electrons into the region of the fields (see Fig. 2-12).  Give a quantitative description of the electronic motion.

*Solution.*  As discussed above, the path is a helix of variable pitch.  The plane determined by $\mathcal{E}$ and $v_0$ is chosen as the $XY$ plane.  From Eq. (2-13),

$$v_0 = 5.93 \times 10^5\sqrt{400} = 1.19 \times 10^7 \text{ m/sec}$$

$$v_{0x} = v_0 \sin 30° = 5.93 \times 10^6 \text{ m/sec}$$

$$v_{0y} = v_0 \cos 30° = 1.03 \times 10^7 \text{ m/sec}$$

$$a = \frac{e\mathcal{E}}{m} = 1.76 \times 10^{15} \text{ m/sec}^2$$

along the $-Y$ direction.  Hence, from Eq. (2-38),

$$R = \frac{mv_{0x}}{eB} = \frac{5.93 \times 10^6}{1.76 \times 10^{11} \times 5 \times 10^{-3}} = 6.75 \times 10^{-3} \text{ m}$$

Thus, the projection of the path on the $XZ$ plane is a circle of radius 0.675 cm.  An application of the direction rule shows that the circle is being traversed in the counterclockwise direction (when looking along the $+Y$ direction).  The path is essentially that illustrated in Fig. 2-10 except that the velocity along the fields is not constant but is given by Eqs. (2-44), *viz.*,

$$v_y = 1.03 \times 10^7 - 1.76 \times 10^{15}t$$

Also

$$y = 1.03 \times 10^7 t - 0.88 \times 10^{15}t^2$$

Under the conditions chosen, the particle starts to move in the $+Y$ direction; but since the acceleration is along the $-Y$ direction, the velocity will shortly be reduced to

zero, and the particle will then reverse its $Y$-directed motion. This reversal will occur at the time $t'$ for which $u_y = 0$, namely,

$$t' = \frac{1.03 \times 10^7}{1.76 \times 10^{15}} = 5.86 \times 10^{-9} \text{ sec}$$

The distance traveled in the $+Y$ direction to the position at which the reversal occurs is

$$y' = 1.03 \times 10^7 \times 5.86 \times 10^{-9} - 0.88 \times 10^{15}(5.86 \times 10^{-9})^2$$

$$= (6.04 - 3.04) \times 10^{-2} \text{ m} = 3.00 \text{ cm}$$

It should be kept in mind that the term "reversal" refers only to the $Y$-directed motion, not to the direction in which the electron traverses the circular component of its path.
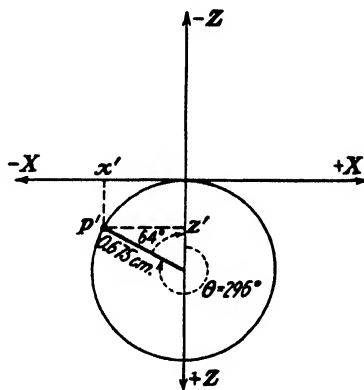


FIG. 2-13.   The projection of the path in the $XZ$ plane is a circle.

The rotation in this circular component is determined entirely by the quantities $B$ and $v_{0x}$. Furthermore, only one reversal is made, the $Y$-directed motion continuing in the $-Y$ direction with an ever-increasing linear velocity, even though the angular velocity remains constant and equal to

$$\omega = \frac{eB}{m} = 1.76 \times 10^{11} \times 5 \times 10^{-3} = 8.80 \times 10^8 \text{ rad/sec}$$

By using either the relationship $T = 2\pi/\omega$ or Eq. (2-41), there is obtained $T = 7.14 \times 10^{-9}$ sec for the period.

Since $t' < T$, then less than one revolution is made before the reversal.

The point $P'$ in space at which the reversal takes place is obtained by considering the projection of the path in the $XZ$ plane (since the $Y$ coordinate $y'$ is already known). Refer to Fig. 2-13. The angle $\theta$ through which the electron has rotated about the $Y$ axis is

$$\theta = \omega t' = 8.80 \times 10^8 \times 5.86 \times 10^{-9} = 5.17 \text{ rad} = 296°$$

From the figure it is clear that

$$x' = -0.675 \sin 64° = -0.608 \text{ cm}$$

and

$$z' = 0.675 - 0.675 \cos 64° = 0.317 \text{ cm}$$

The time $t''$ that it takes the electron to return to the $XZ$ plane is obtained by setting $y$ equal to zero. Thus,

$$t'' = \frac{1.03 \times 10^7}{0.88 \times 10^{15}} = 1.17 \times 10^{-8} \text{ sec}$$

Since $t'' = 2t'$, it takes the electron just as long to travel from the $XZ$ plane to the point of reversal as it does to return to this plane. The point to which the electron returns in the $XZ$ plane may be obtained by using for $\theta$ the value $\omega t'' = 2\omega t' = 592°$ in Fig. 2-13. The result is

$$x'' = -0.532 \text{ cm} \qquad \text{and} \qquad z'' = 1.09 \text{ cm}$$

It should be noted that the electron does *not* return to the origin.

**2-14. Perpendicular Electric and Magnetic Fields.** The directions of the fields are shown in Fig. 2-14. The magnetic field is directed along the $-Y$ axis, and the electric field is directed along the $-X$ axis. The force on an electron due to the electric field is directed along the $+X$ axis. Any force due to the magnetic field is always normal to **B** and hence lies in a plane parallel to the $XZ$ plane. Thus there is no component of force along the $Y$ direction, and the $Y$ component of acceleration is zero. Hence the motion along $Y$ is given by

$$f_y = 0$$

$$v_y = v_{0y} \qquad \text{and} \qquad y = v_{0y}t \qquad (2\text{-}15)$$

assuming that the electron starts at the origin.

*If the initial velocity component parallel to* **B** *is zero, then the path lies entirely in a plane perpendicular to* **B**.

It is desired to investigate the path of an electron *starting at rest* at the origin. The initial magnetic force is zero, since the velocity is zero. The electric force is directed along the $+X$ axis, and the electron will be accelerated in this direction. As soon as the electron is in motion, the magnetic force will no longer be zero. There will then be a component of this force which will be proportional to the $X$ component of velocity and will be directed along the $+Z$ axis. The path will thus bend away from the $+X$ direction toward the $+Z$ direction. Clearly, the electric and magnetic forces interact with one another. In fact, the analysis cannot be carried along further profitably in this qualitative fashion. The arguments given above do, however, indicate the manner in which the electron starts on its path. This path will be shown to be a cycloid.

To determine the path of the electron quantitatively, the force equations must be set up. The force due to the electric field $\mathcal{E}$ is $e\mathcal{E}$ along the $+X$ direction. The force due to the magnetic field is found as follows: At any instant, the velocity is determined by the three components $v_x$, $v_y$, and $v_z$ along the three coordinate axes. Since **B** is in the $Y$ direction, no force
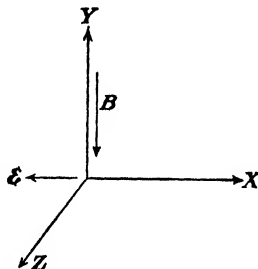
Fig. 2-14. Perpendicular electric and magnetic fields.

will be exerted on the electron due to $v_y$. Because of $v_x$, the force is $ev_xB$ in the $+Z$ direction, as can be verified by the direction rule of Sec. 2-9. Similarly, the force due to $v_z$ is $ev_zB$ in the $-X$ direction. Hence Newton's law, when expressed in terms of the three components, yields

$$f_x = m\frac{dv_x}{dt} = e\mathcal{E} - ev_zB \qquad f_z = m\frac{dv_z}{dt} = ev_xB \qquad (2\text{-}46)$$

By writing for convenience

$$\omega \equiv \frac{eB}{m} \qquad u \equiv \frac{\mathcal{E}}{B} \qquad (2\text{-}47)$$

then the foregoing equations may be written in the form

$$\frac{dv_x}{dt} = \omega u - \omega v_z \qquad \frac{dv_z}{dt} = +\omega v_x \qquad (2\text{-}48)$$

A straightforward procedure is involved in the solution of these equations. In order to evaluate the velocity components, the first equation of (2-48) is differentiated and combined with the second, *viz.*,

$$\frac{d^2v_x}{dt^2} = -\omega\frac{dv_z}{dt} = -\omega^2 v_x \qquad (2\text{-}49)$$

The solution of this differential equation is

$$v_x = A\cos\omega t + C\sin\omega t \qquad (2\text{-}50)$$

where $A$ and $C$ are arbitrary constants. In order to evaluate these constants, the initial conditions, $v_{0x} = v_{0z} = 0$ when $t = 0$, are imposed on this equation. Thus,

and from Eq. (2-48)
$$\left. \begin{array}{c} v_x = 0 \\ \dfrac{dv_x}{dt} = \omega u \end{array} \right\} \quad \text{when } t = 0$$

From these, it is found that

$$A = 0 \qquad \text{and} \qquad C = u \qquad (2\text{-}51)$$

so that

$$v_x = u\sin\omega t$$

The first equation of (2-48) and the solution for $v_z$ give

$$v_z = u - \frac{1}{\omega}\frac{dv_x}{dt} = u - u\cos\omega t$$

Thus, the solutions of Eqs. (2-48) are

$$v_x = u\sin\omega t \qquad v_z = u - u\cos\omega t \qquad (2\text{-}52)$$

In order to find the coordinates $x$ and $z$ from these expressions, each equation must be integrated. Thus

$$x = \int v_x \, dt = \int u \sin \omega t \, dt$$

or

$$x = -\frac{u}{\omega} \cos \omega t + D$$

Since $x = 0$ when $t = 0$, then

$$D = \frac{u}{\omega}$$

In a similar way, from Eq. (2-52)

$$z = \int v_z \, dt = ut - \frac{u}{\omega} \sin \omega t$$

the constant of integration being zero since $z = 0$ at $t = 0$.

The complete solution of this problem is, therefore,

$$x = \frac{u}{\omega} (1 - \cos \omega t) \qquad z = ut - \frac{u}{\omega} \sin \omega t \qquad (2\text{-}53)$$

If, for convenience,

$$\theta \equiv \omega t \qquad \text{and} \qquad Q \equiv \frac{u}{\omega} \qquad (2\text{-}54)$$

then

$$x = Q(1 - \cos \theta) \qquad z = Q(\theta - \sin \theta) \qquad (2\text{-}55)$$

where $u$ and $\omega$ are as defined in Eq. (2-47).

It will now be shown that these are the parametric equations of a *common cycloid*, defined as *the path generated by a point on the circumference of a circle of radius Q which rolls along a straight line*, the Z axis. This is illustrated in Fig. 2-15. The point $P$ whose coordinates are $x$ and $z$ $(y = 0)$ represents the position of the electron at any time. The dark curve is the locus of the point $P$. The reference line $CC'$ is drawn through the center of the generating circle parallel to the $X$ axis. Since the circle rolls on the $Z$ axis, then $OC'$ represents the length of the circumference that has already come in contact with the $Z$ axis. This length is evidently equal to the arc $PC'$ (and equals $Q\theta$). The angle $\theta$ gives the number of radians through which the circle has rotated. From the diagram, it readily follows that
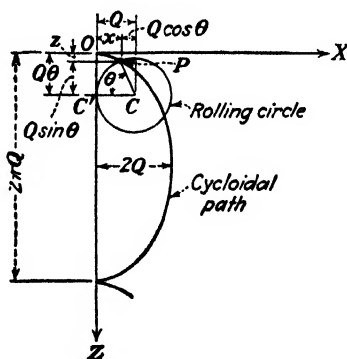


Fig. 2-15. The cycloidal path of an electron in perpendicular electric and magnetic fields when the initial velocity is zero.

$$x = Q - Q \cos \theta \qquad z = Q\theta - Q \sin \theta \qquad (2\text{-}56)$$

which are identical with Eq. (2-55), thus proving that the path is cycloidal as predicted.

The physical interpretation of the symbols introduced above merely as abbreviations is as follows:

$\omega$ *represents the angular velocity of rotation of the rolling circle.*

$\theta$ *represents the number of radians through which the circle has rotated.*

$Q$ *represents the radius of the rolling circle.*

Since $u = \omega Q$, then $u$ *represents the velocity of translation of the center of the rolling circle.*

From these interpretations and from Fig. 2-15 it is clear that the maximum displacement of the electron along the $X$ axis is equal to the diameter of the rolling circle, or $2Q$. Also, the distance along the $Z$ axis between cusps is equal to the circumference of the rolling circle, or $2\pi Q$. At each cusp the speed of the electron is zero, since at this point the velocity is reversing its direction (see Fig. 2-15). This is also seen from the fact that each cusp is along the $Z$ axis and hence at the same potential. Therefore the electron has gained no energy from the electric field, and its speed must again be zero.

If an initial velocity exists that is directed parallel to the magnetic field, then the projection of the path on the $XZ$ plane will still be a cycloid but the particle will now have a constant velocity normal to the plane. This might be called a "cycloidal helical motion." The motion is described by Eqs. (2-55) with the addition of Eq. (2-45).

There is one other special case of importance. Suppose that the electron is released perpendicular to both the electric and magnetic fields so that $v_{0x} = v_{0y} = 0$ and $v_{0z} \neq 0$. The electric force is $e\mathcal{E}$ along the $+X$ direction (see Fig. 2-14), and the magnetic force is $ev_{0z}B$ along the $-X$ direction. If the net force on the electron is zero, then it will continue to move along the $Z$ axis with the constant speed $v_{0z}$. This condition is realized when

$$e\mathcal{E} = eBv_{0z}$$

or

$$v_{0z} = \frac{\mathcal{E}}{B} = u \tag{2-57}$$

from Eq. (2-47).

This gives another interpretation to $u$. It represents that velocity with which an electron may be injected into perpendicular electric and magnetic fields and suffer no deflection, *the net force being zero*. Note that this velocity $u$ is independent of the charge or mass of the ions. Such a system of perpendicular fields will act as a *velocity filter* and allow only those particles whose velocity is given by the ratio $\mathcal{E}/B$ to be selected.

*Example.* A magnetic field of 0.01 weber/m² is applied along the axis of a cathode-ray tube. A field of $10^4$ volts/m is applied to the deflecting plates. If an electron

leaves the anode with a velocity of $10^6$ m/sec along the axis, how far from the axis will it be when it emerges from the region between the plates? The length $l$ of the deflecting plates along the tube axis is 2.0 cm.

*Solution.* Choose the system of coordinate axes illustrated in Fig. 2-14. Then,

$$v_{0x} = v_{0z} = 0; \qquad v_{0y} = 10^6 \text{ m/sec}$$

As shown above, the projection of the path is a cycloid in the $XZ$ plane, and the electron travels with constant velocity along the $Y$ axis. The electron is in the region between the plates for the time

$$\frac{l}{v_{0y}} = 2 \times 10^{-8} \text{ sec}$$

Then, from Eqs. (2-47) and (2-54), it is found that

$$\omega = \frac{eB}{m} = 1.76 \times 10^9 \text{ rad/sec}$$

$$u = \frac{\mathcal{E}}{B} = 10^6 \text{ m/sec}$$

$$Q = \frac{u}{\omega} = 5.68 \times 10^{-4} \text{ m}$$

$$= 0.0568 \text{ cm}$$

$$\theta = \omega t = (1.76 \times 10^9)(2 \times 10^{-8}) = 35.2 \text{ rad}$$

Since there are $2\pi$ rad/revolution, the electron goes through five complete cycles and enters upon the sixth before it emerges from the plates. Thus

$$35.2 \text{ rad} = 10\pi + 3.8 \text{ rad}$$

Since 3.8 rad equals 218 deg, then Eqs. (2-55) yield

$$x = Q(1 - \cos\theta) = 0.0568(1 - \cos 218°) = 0.103 \text{ cm}$$

$$z = Q(\theta - \sin\theta) = 0.0568(35.2 - \sin 218°) = 2.03 \text{ cm}$$

so that the distance from the tube axis is

$$r = \sqrt{x^2 + z^2} = 2.03 \text{ cm}$$

**2-15. Perpendicular Electric and Magnetic Fields (Arbitrary Initial Velocities).** It was shown in the preceding section that the motion along the direction of the magnetic field (the $Y$ axis) was one of constant speed as given by Eq. (2-45). It will now be assumed that $v_{0y} = 0$, and the path in the $XZ$ plane for arbitrary initial velocity components $v_{0x}$ and $v_{0z}$ will be investigated.

In the foregoing section it was demonstrated that $u$ reacting with $B$ gives a force just equal in magnitude and opposite in direction to $\mathcal{E}$. Hence, if a new system of coordinates $(X', Z')$ is introduced which moves with a speed $u$ along the $+Z$ axis, then this primed system may be considered as having only a magnetic field but no electric field. A further justification

of this procedure is found in Eq. (2-48). Thus, if

$$z = z' + ut \qquad \text{and} \qquad x = x' \qquad (2\text{-}58)$$

so that

$$v_z = v_z' + u \qquad \text{and} \qquad v_x = v_x' \qquad (2\text{-}59)$$

then Eqs. (2-58) become

$$\frac{dv_x'}{dt} = -\omega v_z' \qquad \text{and} \qquad \frac{dv_z'}{dt} = \omega v_x' \qquad (2\text{-}60)$$

which are identical with Eqs. (2-48) with $u = 0$. However, $u$ vanishes only if $\mathcal{E} = 0$, and in this case Eqs. (2-60) represent the equation of mo-



FIG. 2-16. The path of an electron in perpendicular electric and magnetic fields is a circle in a coordinate system which moves with a speed $u = \mathcal{E}/B$ along the axis normal to both $\mathcal{E}$ and **B**.

tion of an electron in only a uniform magnetic field. Hence, the path in the primed system is a circle with constant angular velocity $\omega$, constant peripheral speed $v_0'$, and a radius $Q' = v_0'/\omega$. These relationships are shown in Fig. 2-16. The center of the circle $C'$ is along the normal to the initial velocity vector $\mathbf{v}_0'$, which is obtained by subtracting **u** from $\mathbf{v}_0$. In other words, $\mathbf{v}_0'$ is a vector whose $X$ component is $v_{0x}$ and whose $Z$ component is $v_{0z} - u$. The point $P'(x', z')$ is the position of the electron in the primed system at a time $t = \theta/\omega$. The position of the point $P(x, z)$ in space is obtained from Eqs.

(2-58), which state that every point in the primed plane must be translated along the $Z$ axis an amount $ut = Q\omega t = Q\theta$. Note that, in general, $Q'$ is not equal to $Q$.

*Example.* Consider perpendicular fields oriented as in Fig. 2-14, with $\mathcal{E} = 10^4$ volts/m and $B = 1$ milliweber/m². Assume that the initial velocity of the electron has components $v_{0x} = 1.55 \times 10^7$ m/sec, $v_{0y} = 0$, and $v_{0z} = 1.78 \times 10^7$ m/sec. Plot the path.

*Solution.* The important constants of the motion are calculated as follows:

$$\omega = \frac{eB}{m} = (1.76 \times 10^{11})(10^{-3}) = 1.76 \times 10^8 \text{ rad/sec}$$

$$u = \frac{\mathcal{E}}{B} = \frac{10^4}{10^{-3}} = 10^7 \text{ m/sec}$$

$$Q = \frac{u}{\omega} = \frac{10^7}{1.76 \times 10^8} \text{ m} = 5.68 \text{ cm}$$

$$v_{0x}' = v_{0x} = 1.55 \times 10^7 \text{ m/sec}$$
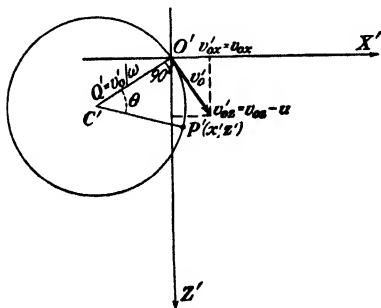
$$v_{0z}' = v_{0z} - u = 1.78 \times 10^7 - 10^7 = 0.78 \times 10^7 \text{ m/sec}$$

$$v_0' = [(v_{0x}')^2 + (v_{0z}')^2]^{\frac{1}{2}} = [(1.55)^2 + (0.78)^2]^{\frac{1}{2}} \times 10^7$$

$$= 1.73 \times 10^7 \text{ m/sec}$$

$$Q' = \frac{v_0'}{\omega} = \frac{1.73 \times 10^7}{1.76 \times 10^8} \text{ m} = 9.83 \text{ cm}$$

The circular path in the primed system is shown in Fig. 2-17a.  The circle is divided into eight 45-deg segments labeled 0 through 7, respectively.  Each point in the primed
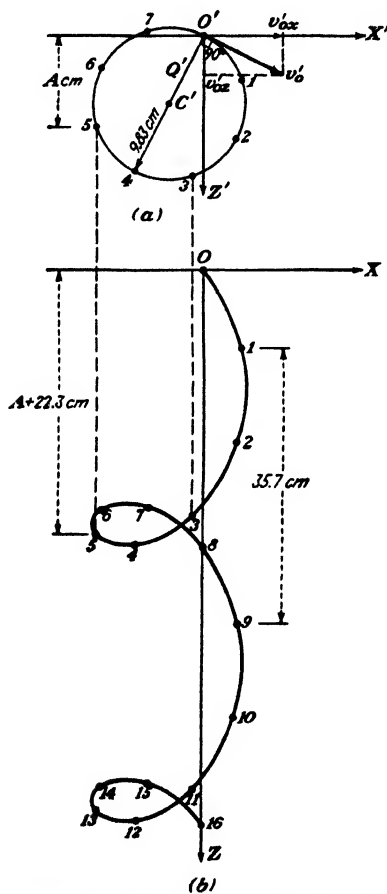


FIG. 2-17.  A problem illustrating the prolate cycloidal path of an electron in perpendicular electric and magnetic fields.

system must be displaced an amount $Q\theta = 5.68\theta$ along the $Z$ axis to find its corresponding $Z$ coordinate. These are calculated in the accompanying table.

| Point | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $\theta$ rad | 0 | 0.784 | 1.57 | 2.35 | 3.14 | 3.92 | 4.71 | 5.48 |
| $Q\theta$ cm | 0 | 4.46 | 8.92 | 13.4 | 17.8 | 22.3 | 26.8 | 31.2 |

The path is indicated in Fig. 2-17b, which is obtained from Fig. 2-17a by using a pair of dividers to displace corresponding points by the amount indicated in the table. For example, point 5, which is at some distance, say $A$, from the $X'$ axis, is put at a distance $A + 22.3$ from the $X$ axis.

Two complete cycles are indicated. Points in the second cycle are obtained from corresponding points in the first cycle by the displacement $2\pi Q = 35.7$ cm. Thus, point 9 is 35.7 cm below point 1, etc.

The above theory illustrates that the primed system moves with a speed $u$ in the $+Z$ direction relative to the unprimed system. One method of obtaining this translation is illustrated in the graphical construction of Fig. 2-17. Another method is to make $C'$ the center of a circle (of radius $Q$) which rolls along a line parallel to the $Z$ axis. Then the linear velocity of $C'$ will be $Q\omega = u$. This is illustrated in Fig. 2-18. The path, called a *trochoid*,[4] is the locus of a point on a "spoke" of a wheel rolling on a straight line. If the length $Q'$ of the spoke is greater than
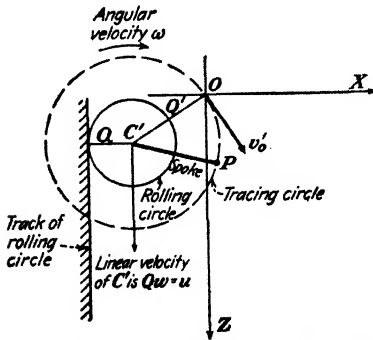


Fig. 2-18. The locus of the point $P$ at the end of a "spoke" of a wheel rolling on a straight line is a trochoid.
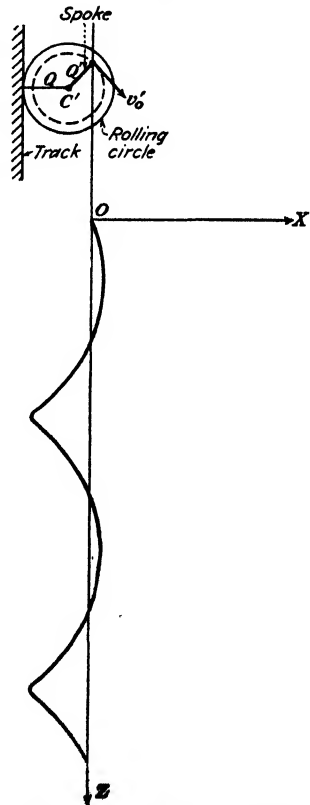


Fig. 2-19. A curtate cycloid.

the radius $Q$ of the rolling circle, the trochoid is called a *prolate cycloid* [4] (illustrated in Fig. 2-17). If $Q' = Q$, the path is called a *common cycloid* (this was the case considered in Sec. 2-14). If $Q'$ is less than $Q$, the path is called a *curtate cycloid*.[4] This is illustrated in Fig. 2-19, and it should be noted that the "loops" of Fig. 2-17 are missing.

In addition to the geometrical methods just described, the path can also be obtained analytically by integrating Eq. (2-48) directly for the case where the initial velocities differ from zero. The resulting equations are given in Prob. 2-37. The trochoidal nature of the path is, however, completely hidden in these equations.

**2-16. Constant Electric and Magnetic Fields Making an Arbitrary Angle with Each Other.** Analyses of the motion of an electron in a simple electric field that is constant both in magnitude and in direction, in a constant magnetic field, in combined parallel constant electric and magnetic fields, and in combined perpendicular constant electric and magnetic fields have been given. The only remaining possible combination of constant electric and magnetic fields is that illustrated in Fig. 2-20. The fields make an angle $\varphi$ with each other. For purposes of analysis, the plane determined by **B** and $\mathcal{E}$ is chosen as the $XY$ plane,



Fig. 2-20. Constant electric and magnetic fields at an angle $\varphi$ with each other.

and **B** is taken along the $-Y$ direction. This orientation of axes permits the application of the results of Sec. 2-14.

The electric field may be expressed with respect to the chosen frame of reference by two components

$$\mathcal{E}_x = + \mathcal{E} \sin \varphi \qquad \mathcal{E}_y = - \mathcal{E} \cos \varphi \qquad (2\text{-}61)$$

Careful thought will show that the equations of motion are those given by Eqs. (2-46) in which $\mathcal{E}$ is replaced by $- \mathcal{E} \sin \varphi$ and in which the force in the $Y$ direction is no longer zero but is $e(\mathcal{E} \cos \varphi)$. The solution of the motion is then given by Eqs. (2-55) with $\mathcal{E}$ replaced by $- \mathcal{E} \sin \varphi$ and with the expression for $y$ being.

Here, the quantity

$$\left. \begin{array}{l} y = v_{0y}t + \tfrac{1}{2}a_y t^2 \\[2mm] a_y = \dfrac{e(\mathcal{E} \cos \varphi)}{m} \end{array} \right\} \qquad (2\text{-}62)$$

is the constant acceleration along the $Y$ axis. That is, the motion is identical with that for the perpendicularly directed fields except for the added acceleration component along the $Y$ direction. For example, if the electron

starts at rest, the projection of the path in the $XZ$ plane will still be a cycloid.

*Example.*   The fields in Fig. 2-20 have the following values:

$$\mathcal{E} = 5 \text{ kv/m}, \qquad B = 1 \text{ milliweber/m}^2, \qquad \text{and angle } \varphi = 20°$$

If an electron is released with zero velocity at the origin, where will it expose a photographic plate which is perpendicular to the $Z$ axis at a distance of 8.00 cm from the origin?

*Solution.*   The projection of the path in the $XZ$ plane is a common cycloid, as discussed above.   The orientation of this cycloid is found by considering the initial effects
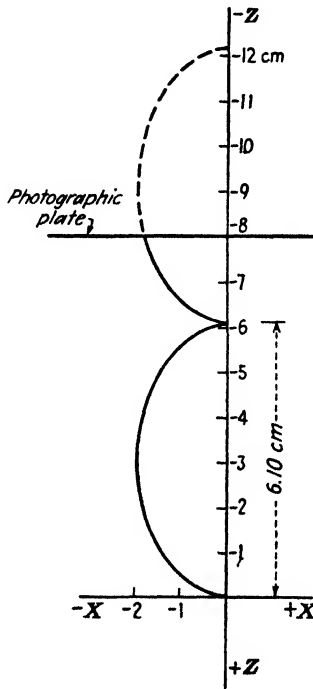


Fig. 2-21.   A problem illustrating cycloidal motion.

of the field on the electron.   Thus, first the electric field accelerates it in the *minus* $X$ direction.   Applying the direction rule of Sec. 2-9 for an *electronic* velocity along $-X$ and a magnetic field along $-Y$ gives a magnetic force along $-Z$.   Hence, the projection of the path is as shown in Fig. 2-21.

It is now seen that the photographic plate will not be exposed unless it is located along the $-Z$ axis as shown.   The equation of the cycloid is

$$-x = Q(1 - \cos\theta) \qquad\qquad -z = Q(\theta - \sin\theta)$$

The radius of the rolling circle is calculated as follows:

$$u = \frac{\mathcal{E}\sin\varphi}{B} = \frac{5 \times 10^3 \sin 20°}{10^{-3}} = 1.71 \times 10^6 \text{ m/sec}$$

$$\omega = \frac{eB}{m} = (1.76 \times 10^{11})(10^{-3}) = 1.76 \times 10^8 \text{ rad/sec}$$

$$Q = \frac{u}{\omega} = \frac{1.71 \times 10^6}{1.76 \times 10^8} = 0.971 \times 10^{-2} \text{ m} = 0.971 \text{ cm}$$

The distance traveled along the $-Z$ direction between cusps is $2\pi Q = 6.10$ cm. Thus, the plate is located beyond the first cycle, as indicated. The angle $\theta'$ *in the second cycle* is given by the equation

$$8.00 - 6.10 = (0.971)(\theta' - \sin\theta') \quad \text{cm} \quad \text{or} \quad \theta' - 1.95 = \sin\theta'$$
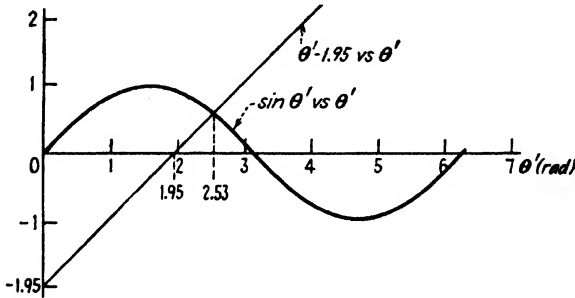


FIG. 2-22. Solution of $\theta' - 1.95 = \sin\theta'$.

This transcendental equation is solved graphically by plotting $\theta' - 1.95$ versus $\theta'$ and $\sin\theta'$ versus $\theta'$ and finding the point of intersection. The result, as given in Fig. 2-22, is $\theta' = 2.53$ rad, or 145 deg. Hence

$$\theta = \theta' + 2\pi = 2.53 + 6.28 = 8.81 \text{ rad}$$

As a check on the graphical solution, note that

$$z = -(0.971)(8.81 - \sin 145°) = -8.00 \text{ cm}$$

as it should. The $X$ coordinate of the exposed point is

$$x = -(0.971)(1 - \cos 145°) = -1.76 \text{ cm}$$

The time $t_1$ it takes the electron to reach the plate is

$$t_1 = \frac{\theta}{\omega} = \frac{8.81}{1.76 \times 10^8} = 5.01 \times 10^{-8} \text{ sec}$$

The $Y$ component of acceleration is

$$a_y = \frac{e\mathcal{E}_y}{m} = (1.76 \times 10^{11})(5 \times 10^3)(\cos 20°)$$

$$= 8.28 \times 10^{14} \text{ m/sec}^2$$

The plate is exposed at

$$y = \tfrac{1}{2}a_y t_1^2 = \tfrac{1}{2}(8.28 \times 10^{14})(5.01 \times 10^{-8})^2$$

$$= 1.04 \text{ m}$$

**2-17. Nonuniform Fields.** If the fields are not constant, it is still possible to write down the differential equations of motion but it is seldom possible to obtain an exact solution. The particular case in which the tube possesses cylindrical symmetry is amenable to analysis. However, in most practical tubes the fields themselves are seldom known exactly, but only the voltages applied to a certain electrode configuration are given. Because of the importance of this type of problem approximate graphical, numerical, and experimental methods of solution have been devised.[5]

One direct experimental method for finding the path of an electron in a given two-dimensional electrode configuration (without first finding the field) is called the "rubber-model" method. In this method, a large-scale model is made of the electrode assembly, the height of the various electrodes being adjusted to be in proper proportion to the negative potential of the electrodes. A rubber membrane is pressed down over this configuration in such a way that it makes contact with the top of every electrode. The surface of the diaphragm is then no longer flat but has an elevation that is proportional to the potential at each point in space. A small ball is projected onto this model with an initial speed and direction appropriate to the original electronic problem, and the subsequent path of the rolling ball is photographed with a motion-picture camera. It can be shown that the motion of the ball rolling on the stretched diaphragm under the action of the earth's gravitational field (subject to certain conditions usually approximated in practice) is identical with that of the electron in the original electrode configuration.[6]

**2-18. Electron Optics.** Because of the close analogy that exists between the paths of charged particles in electric and magnetic fields and the path of light rays in passing through lenses or through media of varying index of refraction, the foregoing analyses may be called "electron optics" or, more accurately, "geometrical electron optics." To note this close analogy,[7] consider the regions on both sides of an equipotential surface $S$, as shown in Fig. 2-23. Suppose that the electric potential to the left of the surface is $V^-$ and that to the right of $S$ is $V^+$. Suppose that an electron is moving in the direction $PQ$ with a velocity $v_1$. At the surface $S$ a force exists in the direction normal to the equipotential. Because of this force, the velocity of the electron increases to $v_2$ after it has passed $S$. Only the normal component of velocity $v_n$ changes, since no work is done by moving the particle along an equipotential. That is, the tangential component of the velocity $v_t$ on both sides of the equipotential remains unchanged. It

follows from Fig. 2-23 that

$$v_t = v_1 \sin i = v_2 \sin r$$

where $i$ and $r$ may be considered as the angles of incidence and refraction of the electron ray, respectively.   Then

$$\frac{\sin i}{\sin r} = \frac{v_2}{v_1} \tag{2-63}$$

In geometrical light optics the ratio of the velocities of propagation in two media is called the *index of refraction* $\mu$.   This resultant equation is then recognized as Snell's law of refraction, $\sin i/\sin r = \mu$.   It therefore
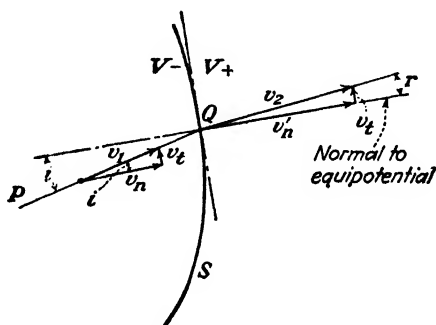


FIG. 2-23.   To demonstrate the similarity between "geometrical electron optics" and "geometrical light optics."

follows that the refraction of an electronic beam at an equipotential surface obeys the same law as the bending of a light beam at a refracting surface.   Because of this, the electron-lens system and an optical-lens system may be considered to be roughly analogous.   It must be kept in mind, however, that electron lenses cannot be sharply defined, the region actually being one of continuously varying index of refraction.   Furthermore, the index of refraction of the electron lens can readily be varied by changing the potentials applied to the electrodes that constitute the lens.   This is the electron-lens arrangement for focusing the electron beam in a cathode-ray tube and will be further discussed in Sec. 3-6.

## PROBLEMS

**2-1.** *a.* An electron is emitted from a thermionic cathode with a negligible initial velocity and is accelerated by a potential of 1,000 volts.   Calculate the final velocity of the particle.

*b.* Repeat the foregoing for the case of a deuterium ion (heavy hydrogen ion   atomic weight 2.01) that has been introduced into the electric field with an initial velocity of $10^6$ m/sec.

**2-2.** An electron starts at rest at the negative plate of a plane-parallel capacitor across which is 2,000 volts. The distance between the plates is 3 cm.

  *a.* How long has the electron been traveling when it acquires a speed of $10^7$ m/sec?

  *b.* How far has the electron traveled before it acquires this speed?

  *c.* What potential has the electron fallen through when it acquires this speed?

**2-3.** *a.* The distance between the plates of a plane-parallel capacitor is 1 cm. An electron starts at rest at the negative plate. If a direct voltage of 1,000 volts is applied, how long will it take the electron to reach the positive plate?

  *b.* If a 60-cycle sinusoidal voltage of peak value 1,000 volts is applied, how long will the time of transit be? Assume that the electron is released with zero velocity at the instant of time when the applied voltage is passing through zero. (HINT: Expand the sine function into a power series. Thus, $\sin \theta = \theta - \dfrac{\theta^3}{3!} + \dfrac{\theta^5}{5!} - \cdots$.)

**2-4.** An electron starts at rest at the bottom of a plane-parallel capacitor whose plates are 5 cm apart. The applied voltage is zero at the instant the electron is released, and it increases linearly from zero to 10 volts in $10^{-7}$ sec.

  *a.* If the upper plate is positive, what speed will the electron attain in $0.5 \times 10^{-7}$ sec?

  *b.* Where will it be at the end of this time?

  *c.* With what speed will the electron strike the positive plate?

**2-5.** An electron having an initial kinetic energy of $10^{-16}$ joule at the surface of one of two parallel-plane electrodes and moving normal to the surface is slowed down by the retarding field caused by a 400-volt potential applied between the electrodes.

  *a.* Will the electron reach the second electrode?

  *b.* What retarding potential would be required for the electron to reach the second electrode with zero velocity?

**2-6.** An electron is released with zero initial velocity from the lower of a pair of horizontal plates which are 3 cm apart. The accelerating potential between these plates increases from zero linearly with time at the rate of 10 volts/$\mu$sec. When the electron is 2.8 cm from the top plate, a reverse voltage of 50 volts is applied.

  *a.* What is the instantaneous potential between the plates at the time of the potential reversal?

  *b.* With which electrode does the electron collide?

  *c.* What is the time of flight?

  *d.* What is the impact velocity of the electron?
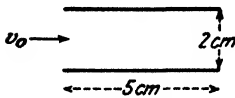
**2-7.** A 100-ev hydrogen ion is released in the center of the plates, as shown in the figure. The voltage between the plates varies linearly from 0 to 50 volts in $10^{-7}$ sec and then drops immediately to zero and remains at zero. The separation between the plates is 2 cm. If the ion enters the region between the plates at time $t = 0$, how far will it be displaced from the $X$ axis upon emergence from between the plates?

$v_0 \longrightarrow$     $2 cm$

$\longleftarrow$-----$5 cm$-----$\longrightarrow$

PROB. 2-7.

**2-8.** Consider a plane-parallel diode across which there is an impressed voltage $E_0 + E_1 \sin \omega t$. If an electron leaves the cathode with a perpendicular velocity $v_0$ at time $t_0$, derive the expression for the position of the electron at any subsequent time $t$.

**2-9.** A plane triode consists of a cathode and an anode 4 cm apart, and an open grid midway between them. The grid is maintained at a potential of 5 volts above the cathode, and the plate is at 20 volts above the cathode. How long will it take an electron starting at rest at the cathode to reach the anode? Assume that the potential varies linearly from the cathode to the grid and also linearly from the grid to the anode.

**2-10.** 100-volt electrons are introduced at $A$ into a uniform electric field of $10^4$ volts/m. The electrons are to emerge at the point $B$ in time $4.77 \times 10^{-9}$ sec.

  a. What is the distance $AB$?

  b. What angle does the electron gun make with the horizontal?
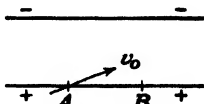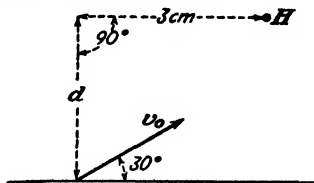


PROB. 2-10.                PROB. 2-11.

**2-11.** Electrons are projected into the region of constant electric-field intensity of magnitude $5 \times 10^3$ volts/m that exists vertically. The electron gun makes an angle of 30 deg with the horizontal. It ejects the electrons with an energy of 100 ev.

  a. How long does it take an electron leaving the gun to pass through a hole $H$ at a horizontal distance of 3 cm from the position of the gun? Refer to the figure. Assume that the field is downward.

  b. What must be the distance $d$ in order that the particles emerge through the hole?

  c. Repeat the foregoing for the case where the field is upward.

**2-12.** In a certain plane-parallel diode the potential $V$ is given as a function of the distance $x$ between electrodes by the equation

$$V = kx^{\frac{4}{3}}$$

where $k$ is a constant. Find an expression for the time that it will take an electron that leaves the cathode with zero initial velocity to reach the anode, a distance $d$ away.

**2-13.** a. Through what potential must an electron fall, if relativistic corrections are not made, in order that it acquire a speed equal to that of light?

  b. What speed does the electron actually acquire in falling through this potential?

**2-14.** Calculate the ratio $m/m_0$ for 2-Mev electrons and also for 2-Mev deuterons (atomic weight 2.01).

**2-15.** An electron starts at rest in a constant electric field. Using the relativistic expression for the mass, find the velocity and the displacement of the particle at any time $t$.

**2-16.** A current of 30 amp is passing through a No. 8 (AWG) copper wire (diameter 128.5 mils). If there are $5 \times 10^{28}$ free electrons per cubic meter in copper, with what average drift speed does an electron move in the wire?

**2-17.** What transverse magnetic field acting over the entire length of a cathode-ray tube must be applied to cause a deflection of 3 cm on a screen that is 15 cm away from the anode, if the accelerating voltage is 2,000 volts?
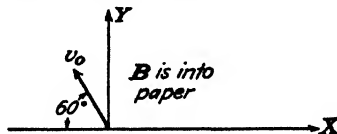


PROB. 2-18.

**2-18.** A 100-volt electron is introduced into the region of uniform magnetic-field intensity of 5 milliwebers/m², as shown.

  a. At what point does the electron strike the $XZ$ plane?

  b. What are the velocity components with which the electron strikes the $XZ$ plane?

**2-19.** Two 50-ev electrons enter a magnetic field of 2.0 milliwebers/m² as shown, one at 10 deg, the other at 20 deg.   How far apart are these electrons when they have traversed

a.   One revolution of their helical paths?

b.   Two revolutions of their helical paths?
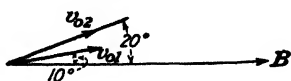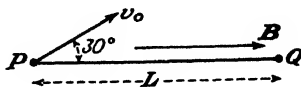


PROB. 2-19.                    PROB. 2-20.

**2-20.** An electron is injected into a magnetic field with a velocity of $10^7$ m/sec in a direction lying in the plane of the paper and making an angle of 30 deg with $B$, as shown in the figure.   If the length $L$ is 0.1 m, what must be the value of $B$ in order that the electron pass through the point $Q$?

**2-21.** Deuterons (ionized heavy hydrogen atoms—atomic weight 2.01) that are produced in an arc chamber are accelerated by falling through a potential of 100 kv.

a.   Through what angle is the direction of the beam deflected if the ions pass through a magnetic field of 800 gauss that is confined to a region 7 cm long?

b.   If the particles pass between a pair of plates 7 cm long and 2 cm apart between which a potential of 800 volts is maintained, what is the angle of deflection of the beam?

**2-22.** Electrons emerge from a hole in an anode of a cathode-ray tube in a diverging cone of small angle.   With 900 volts between the cathode and the anode, the minimum longitudinal magnetic field that is required to cause the electron beam to come to a focus on the screen is 2.5 milliwebers/m².   If the anode voltage is decreased to 400 volts, what minimum magnetic field will now be necessary to focus the beam?   What is the next higher value of magnetic field at which a focus will be obtained?

**2-23.** Electrons emerge from the hole in the anode of a cathode-ray tube in all directions within a cone of small angle.   The accelerating voltage is 300 volts.   The distance from the anode to the screen is 22.5 cm.

The tube is placed in a 40-cm-long solenoid having a diameter of 12 cm and wound with 24 turns of wire per inch.   The tube and solenoid axes coincide.   The maximum current rating of the solenoid is 5 amp.   For what values of current in the solenoid will the beam of electrons come to a focus as a spot on the screen?

**2-24.** Refer to Sec. 2-12 on Magnetic Focusing.   Show that the coordinates of the electron *on the screen* are

$$x = \left(\frac{v_{0x}L}{v_{0y}\alpha}\right)\sin\alpha \quad \text{and} \quad z = \left(\frac{v_{0x}L}{v_{0y}\alpha}\right)(1 - \cos\alpha)$$

where $\alpha = eBL/mv_{0y}$ and the other symbols have the meanings given in the text.

Let $x' = (\pi v_{0y}/Lv_{0x})x$ and $z' = (\pi v_{0y}/Lv_{0x})z$, and plot $z'$ versus $x'$ for intervals of $\alpha$ equal to $\pi/4$.   This will give the path that the electrons will trace out on the screen as the magnetic-field intensity is increased from zero.   Plot enough points so that the path corresponding to two complete spirals will be obtained.
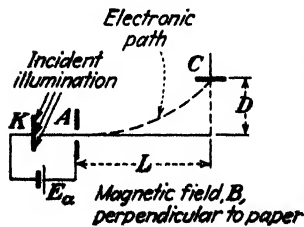


PROB. 2-25.

**2-25.** Lenard's apparatus for measuring $e/m$ for photoelectrons is shown in the sketch.   Electrons are released from the cathode $K$ under the influence of the incident illumination.   The electrons are accelerated by the potential $E_a$ volts to the anode $A$.   They pass through the hole in the anode and are deflected by a

transverse magnetic field so that they are collected at $C$.   Show that

$$\frac{e}{m} = \frac{2E_a}{R^2B^2} \quad \text{coulombs/kg}$$

where $R = (D^2 + L^2)/2D$ meters is the radius of the path.

(See G. P. Harnwell and J. J. Livingood, "Experimental Atomic Physics," p. 117, McGraw-Hill Book Company, Inc., New York, 1923.)

**2-26.** Show that it is justifiable to neglect the earth's gravitational field in considering the motion of an ion in electric- and magnetic-field configurations.  To do this, calculate the electric-field intensity that will exert a force on a singly ionized molecule (say, Hg) equal to the force of gravitation on this ion.  Also, calculate the maximum speed with which this ion can travel perpendicular to the earth's magnetic field (assume a field intensity of 0.6 gauss) if the gravitational force is to be at least 1 per cent of the magnetic force.

**2-27.** Given a uniform electric field of $5 \times 10^3$ volts/m parallel to and in the same direction as a uniform magnetic field of 1.2 milliwebers/m$^2$.  300-ev electrons enter the region where these fields exist, at an angle of 30 deg with the direction of the fields.  A photographic plate is placed normal to the direction of the fields at a distance of 1.6 cm from the electron gun as shown in the figure.

*a.* At what point do the electrons strike the plate?

*b.* With what velocity components do they strike the plate?

*c.* Repeat parts *a* and *b* for the case where the direction of the electric field is reversed.
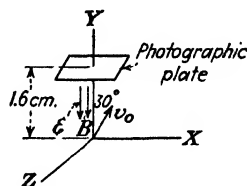
PROB. 2-27.

**2-28.** A positive hydrogen ion enters a region containing parallel electric and magnetic fields in a direction perpendicular to the lines of force.  The electric-field strength is $10^4$ volts/m, and the magnetic-field strength is 0.1 weber/m$^2$.  How far along the direction of the fields will the ion travel during the second revolution of its helical path?

**2-29.** Given a uniform electric field of $10^4$ volts/m parallel to and in the same direction as a uniform magnetic field of $B$ webers per square meter.  300-volt electrons enter the region where these fields exist at an angle of 60 deg with the direction of the fields.  If the electron reverses its direction of travel along the lines of force at the end of the first revolution of its helical path, what must be the strength of the magnetic field?

**2-30.** In Fig. 2-12 what must be the relationship between $\mathcal{E}$, $B$, $\theta$, $v_0$ if the electron is to return to the origin?  The 30-deg angle in the figure is $\theta$.

PROB. 2-31.

**2-31.** Given a uniform electric field of $2 \times 10^4$ volts/m and a uniform magnetic field of 0.03 weber/m$^2$ parallel to each other and in the same direction.  Into this region are released 150-ev hydrogen ions in a direction normal to the fields.  A photographic plate is placed normal to the initial direction of the ions at a distance of 5.0 cm from the gun, as shown in the figure.

*a.* How long after leaving the gun will the ions hit the plate?

*b.* What are the coordinates of the point at which the photographic plate is exposed?

*c.* Repeat for the case where the photographic plate is perpendicular to the $Y$ axis and 5.0 cm from the origin (instead of perpendicular to the $X$ axis).

*d.* Repeat for the case where the photographic plate is perpendicular to the $Z$ axis and 5.0 cm from the origin.
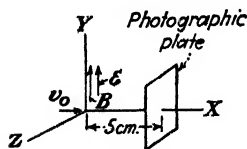
**2-32.** An electron starts from rest at the center of the negative plate of a parallel-plate capacitor across which is a voltage of 100 volts. Parallel to the plates is a constant magnetic field of 1.68 milliwebers/m².

a. If the distance between the plates is 1 cm, how far from the center does the electron strike the positive plate?

b. How long will it take the electron to reach the positive plate?

**2-33.** An electron is released at the point $O$ with a velocity $v_0$ parallel to the plates of a parallel-plate capacitor. The distance between the plates is 1 cm, and the applied potential is 100 volts.

a. What magnitude and direction of magnetic field will cause the electron to move in the cycloidal path indicated? Note that $O$ is midway between the plates and that the cusps are on the negative plate.

b. What must be the value of $v_0$ in order that this path be followed?
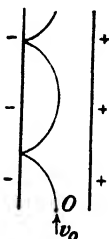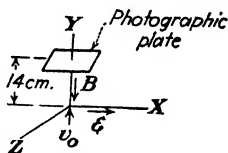


PROB. 2-33.        PROB. 2-34.

**2-34.** Consider the configuration of perpendicular electric and magnetic fields shown in the figure. An ion gun fires 100-ev hydrogen ions along the $Y$ axis as shown. $B = 0.05$ weber/m², and $\varepsilon = 5 \times 10^3$ volts/m.

a. What are the coordinates of the point at which the photographic plate is exposed?

b. Repeat for the case where the photographic plate is perpendicular to the $X$ axis (and at a distance of 14 cm from the origin) instead of perpendicular to the $Y$ axis.

c. Repeat for the case where the photographic plate is perpendicular to the negative $Z$ axis and at a distance of 14 cm from the origin.



PROB. 2-35.

**2-35.** An apparatus for verifying the relativistic variation of mass with velocity [Eq. (2-25)] is shown in the sketch. The electronic source $S$ of high-velocity electrons is situated between the two very closely spaced capacitor plates $CC'$. The entire apparatus (the source, the capacitor plates, and the photographic plate $PP'$) is subjected to a transverse magnetic field of intensity $B$ webers per square meter. Show that if the electric-field intensity between the plates is $\varepsilon$ volts/m only those electrons having a speed $v = \varepsilon/B$ meters per second will leave the region between the plates. Show that for the electrons with this particular speed the ratio of charge to mass is

$$\frac{e}{m} = \frac{\varepsilon}{B^2 R} \qquad \text{coulombs/kg}$$

where the radius of the circular path $R$ is given by $R = (L^2 + D^2)/2D$ meters.

By changing either $B$ or $\varepsilon$ a new value of $v$ and the corresponding value of $e/m$ is obtained, etc. (See Harnwell and Livingood, "Experimental Atomic Physics," p. 120.)
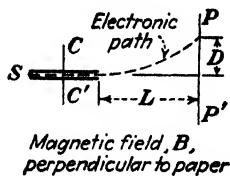
**2-36.** An electron starts at rest in perpendicular electric and magnetic fields. Show that the speed at any instant is given by

$$v = 2u \sin \frac{\theta}{2}$$

and that the distance $d$ traveled *along the cycloidal path* is

$$d = 4Q \left( 1 - \cos \frac{\theta}{2} \right)$$

The symbols have the meaning given in Sec. 2-14.

**2-37.** In Sec. 2-14 the equations of motion in perpendicular electric and magnetic fields are considered, the initial velocities $v_{0x}$ and $v_{0z}$ being taken as zero. Show, by direct integration of Eq. (2-48), that if arbitrary initial velocities are assumed the position of the electron at any time $t$ is given by the equations

$$x = \frac{v_{0x}}{\omega} \sin \omega t + \left( \frac{u}{\omega} - \frac{v_{0z}}{\omega} \right) (1 - \cos \omega t)$$

$$y = v_{0y}t$$

$$z = \frac{v_{0x}}{\omega} (1 - \cos \omega t) - \left( \frac{u}{\omega} - \frac{v_{0z}}{\omega} \right) \sin \omega t + ut$$

Show that these equations represent a trochoidal path.

**2-38.** Using the method of construction explained in Sec. 2-15, obtain the path of an electron *starting at rest* in perpendicular electric and magnetic fields. Show that this is the common cycloid.

**2-39.** From the geometry of Fig. 2-17 show that the coordinates $x'$ and $z'$ are given by

$$x' = Q' \cos (\theta - \theta_0) - Q' \cos \theta_0$$

and

$$z' = Q' \sin (\theta - \theta_0) + Q' \sin \theta_0$$

where $\theta_0$ is the angle that line $C'O'$ makes with the $X'$ axis.

Show that these equations are equivalent to those found by direct integration in Prob. 2-37.

**2-40.** Consider perpendicular fields oriented as in Fig. 2-14, with $\varepsilon = 10^5$ volts/m and $B = 2$ milliwebers/m$^2$. Assume that the initial velocity components are $v_{0x} = 10^7$ m/sec, $v_{0y} = 0$, and $v_{0z} = 2 \times 10^7$ m/sec. Plot the path.

**2-41.** The fields are as in Prob. 2-40. Plot the path if

*a.* $v_{0x} = 10^7$ m/sec and $v_{0y} = v_{0z} = 0$.

*b.* $v_{0x} = v_{0y} = 0$ and $v_{0z} = 10^7$ m/sec.

*c.* $v_{0x} = v_{0y} = 0$ and $v_{0z} = 5 \times 10^7$ m/sec.

**2-42.** A uniform magnetic field of $B$ webers per square meter exists in the $Y$ direction, and a uniform electric field of $10^4$ volts/m makes an angle of 60 deg with $B$ and lies in the $XY$ plane as indicated. A 400-ev electron starts at the origin, moving up to the $Y$ axis.
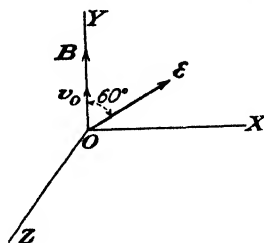
*a.* Describe clearly the exact motion of the electron, including a sketch of the path.

PROB. 2-42.

*b.* Calculate the value (or values) of $B$ which will cause the electron to return to the $XZ$ plane at some point along the $Z$ axis.

**2-43.** An electron which was at rest at the origin at time $t = 0$ strikes the photographic plate at time $t = 0.5 \times 10^{-8}$ sec. Find the $x$, $y$, and $z$ coordinates of the point where it hits the plate. (In the figure, $\mathcal{E}$ is parallel to the $YZ$ plane, $\mathbf{B}$ is parallel to the negative $Y$ axis, and the plate is perpendicular to the $X$ axis.)
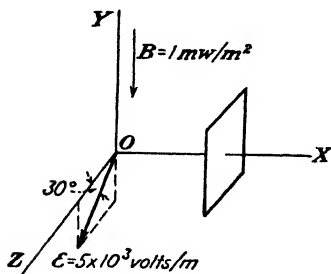


PROB. 2-43.

**2-44.** Uniform electric and magnetic fields of $10^5$ volts/m and 10 milliwebers/m², respectively, are inclined at an angle of 30 deg with respect to each other. If an electron is released with zero initial velocity, how far from its initial position will it be at the end of $10^{-9}$ sec?

**2-45.** Show by integrating Poisson's equation in cylindrical coordinates (see Appendix VI) that the potential at any point $r$ from the axis of a space-charge-free cylindrical capacitor is

$$V = \frac{E_b}{\ln (r_a/r_k)} \ln \frac{r}{r_k}$$

where $E_b$ is the potential difference between the cathode and the anode, $r_a$ is the anode radius, and $r_k$ is the cathode radius.

**2-46.** *a.* If the potential at any point in space is $V(x, y, z)$ write down the differential equations of motion of an electron in this field.

*b.* If the magnetic-field components $B_x(x, y, z)$, $B_y(x, y, z)$, and $B_z(x, y, z)$ are added to the electric field in part *a*, write down the modified equations.

### REFERENCES

1. KAUFMANN, W., *Ann. Physik*, **19**, 487, 1906.
   BUCHERER, A. H., *ibid.*, **28**, 513, 1909.
   PERRY, C. T., and E. L. CHAFFEE, *Phys. Rev.*, **36**, 904, 1930.
   ROGERS, M. M., A. W. McREYNOLDS, and F. T. ROGERS, JR., *ibid.*, **57**, 379, 1940.
2. BUSCH, H., *Physik. Z.*, **23**, 438, 1922.
3. WOLF, F., *Ann. Physik*, **83**, 849, 1927.
   GOEDICKE, E., *ibid.*, **36**, 47, 1939.
4. JAMES, G., and R. C. JAMES, "Mathematics Dictionary," D. Van Nostrand Company, Inc., New York, 1949.
5. ZWORYKIN, V. K., and G. A. MORTON, "Television," John Wiley & Sons, Inc., New York, 1940.
   SPANGENBERG, K. R., "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.
6. ZWORYKIN and MORTON, *op. cit.*, p. 83.
   BRÜCHE, E., and A. RECKNAGEL, *Z. tech. Physik*, **17**, 126, 1936.
7. ZWORYKIN, V. K., G. A. MORTON, E. G. RAMBERG, J. HILLIER, and A. W. VANCE, "Electron Optics and the Electron Microscope," John Wiley & Sons, Inc., New York, 1945.

#### General References

MIT Staff: "Applied Electronics," John Wiley & Sons, Inc., New York, 1943.
RYDER, J. D.: "Electronic Engineering Principles," Prentice-Hall, Inc., New York, 1947.

# CHAPTER 3

# APPLICATIONS OF THE MOTION OF PARTICLES IN APPLIED FIELDS

This chapter will present a discussion of a number of the more important electronic devices that depend for their operation on the theory developed in the preceding chapter.

**3-1. Electrostatic Deflection in Cathode-ray Tubes.** The essentials of a cathode-ray tube for electrostatic deflection are illustrated in Fig. 3-1.
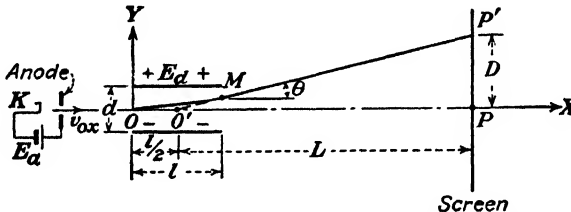


Fig. 3-1. Electrostatic deflection in a cathode-ray tube.

A more detailed discussion of the elements of the tube will follow later. It will be assumed that a constant voltage $E_d$ is applied between the deflecting plates as shown. The initial velocity of the electrons $v_{0x}$ results from the application of an accelerating potential $E_a$. The magnitude of this velocity is given by Eq. (2-12), *viz.*,

$$v_{0x} = \sqrt{\frac{2eE_a}{m}} \qquad \text{m/sec} \qquad (3-1)$$

on the assumption that the initial velocities of emission of the electrons from the cathode are negligible.

Since no field is supposed to exist in the region from the anode to the point $O$, the electrons will move with a constant velocity $v_{0x}$ in a straight-line path. In the region between the plates the electrons will move in the parabolic path given by $y = \frac{1}{2}(a_y/v_{0x}{}^2)x^2$ according to Eq. (2-20). The path is a straight line from the point of emergence $M$ at the edge of the plates to the point $P'$ on the screen, since this region is field-free.

The straight-line path in the region from the deflecting plates to the screen is, of course, tangent to the parabola at the point $M$. The slope of

the line at this point, and so at every point between $M$ and $P'$, is [from Eq. (2-20)]

$$\tan \theta = \left.\frac{dy}{dx}\right]_{x=l} = \frac{a_y l}{v_{0x}^2}$$

From the geometry of the figure, the equation of the straight line $MP'$ is found to be

$$y = \frac{a_y l}{v_{0x}^2}\left(x - \frac{l}{2}\right) \tag{3-2}$$

since $x = l$ and $y = \frac{1}{2}a_y l^2/v_{0x}^2$ at the point $M$.

When $y = 0$, $x = l/2$, which indicates that when the straight line $MP'$ is extended backward it will intersect the tube axis at the point $O'$, the center point of the plates. This is a remarkable result since it means that $O'$ is, in effect, a virtual cathode; and, regardless of the applied potentials $E_a$ and $E_d$, the electrons appear to emerge from this "cathode" and move in a straight line to the point $P'$.

At the point $P'$, $y = D$, and $x = L + \frac{1}{2}l$. Equation (3-2) reduces to

$$D = \frac{a_y l L}{v_{0x}^2}$$

By inserting the known values of $a_y(= eE_d/dm)$ and $v_{0x}$, this becomes

$$D \doteq \frac{lLE_d}{2dE_a} \tag{3-3}$$

This result shows that the deflection on the screen of a cathode-ray tube is directly proportional to the deflecting voltage $E_d$ applied between the plates. Consequently, a cathode-ray tube may be used as a linear-voltage indicating device.

The *electrostatic-deflection sensitivity* of a cathode-ray tube is defined as the deflection (in meters) on the screen per volt of deflecting voltage. Thus

$$S \equiv \frac{D}{E_d} = \frac{lL}{2dE_a} \qquad \text{m/volt} \tag{3-4}$$

This expression shows that, the higher the sensitivity, the greater will be the deflection of the cathode-ray beam per unit applied potential. The *deflection factor* of the tube is, by definition, the reciprocal of the sensitivity and is, therefore,

$$G \equiv \frac{1}{S} = \frac{2dE_a}{lL} \qquad \text{volts/m} \tag{3-5}$$

This expression gives a measure of the potential that must be applied to the deflecting plates in order to give unit deflection on the screen.

An inspection of Eq. (3-4) shows that the sensitivity is independent of both the deflecting voltage $E_d$ and the ratio $e/m$. Furthermore, the sensitivity varies inversely with the accelerating potential $E_a$.

The idealization made in connection with the foregoing development, *viz.*, that the electric field between the deflecting plates is uniform and does not extend beyond the edges of the plates, is never met in practice. Consequently, the effect of fringing of the electric field may be enough to necessitate corrections amounting to as much as 40 per cent[1] in the results obtained from an application of Eqs. (3-4) and (3-5).

Typical values of deflection factors range from approximately 50 to 250 volts/in. corresponding to sensitivities of 0.5 to 0.1 mm/volt.

**3-2. Experimental Determination of Sensitivity.**  The simplest procedure for determining the sensitivity of a cathode-ray tube experimentally is to apply an a-c (60-cycle) voltage to the deflecting plates.  As far as an individual electron is concerned, this is equivalent to a constant potential since, during the time interval during which it is in the region between the plates, the potential remains substantially constant.  To verify this, one need only recall that, if the electron were accelerated even by an anode potential of only 100 volts, its speed would be $5.93 \times 10^6$ m/sec.  The length of time it would be in the region between plates 1 cm long would be $l/v = 0.01/(5.93 \times 10^6)$, or less than $10^{-8}$ sec.  Certainly the 60-cycle voltage may be considered to remain unchanged for such a short time interval.

This analysis furnishes the justification for determining the electrostatic sensitivity by means of an alternating instead of a static field between the plates.  It must be noted that the expression for the deflection sensitivity must be modified when the time of passage of an electron between the plates is comparable with the period of the a-c deflecting voltage.  However, since the time of passage of the electron between the plates is of the order of $10^{-8}$ sec, the deflection sensitivity is independent of frequency for frequencies less than about 10 megacycles/sec.

However, what will be seen on the screen when the deflecting voltage is sinusoidal?  Since each succeeding electron in the cathode-ray stream arrives between the plates at a different phase of the applied voltage, then the deflecting voltage will be different in magnitude for the different electrons.  This circumstance results in different deflections so that the electrons, instead of striking one spot on the screen, will be spread out, the image on the screen being a line.  The eye cannot, of course, detect the individual rain of electrons on the screen but owing to the persistence of vision sees only the integrated result, a line.  Since the applied a-c poten-

tial reverses twice each cycle, the electron beam will be deflected upward during one half period and downward during the other half period. The total length of the line will correspond to that resulting from the application of a d-c voltage equal to $2E_m$, where $E_m$ is the maximum value of the applied a-c potential.

It is a very simple matter to measure the length of the resulting line with a transparent scale, particularly since the end points appear brighter than the rest of the line. The bright end points result from the fact that more electrons are deflected toward these points than to any other point in the line. This is so because, within one cycle of an a-c wave, the voltage is in the neighborhood of its maximum or its minimum value for a large portion of its period.

If the applied deflecting potential is not a pure sinusoidal wave but contains harmonics, the resulting line on the screen will possess, in accordance with the foregoing discussion, gradations in intensity. The brightest portions of the line will correspond to regions where the magnitude of the applied potential wave is changing least with respect to the time, and more electrons are acted upon by this field strength than by any other. This condition is illustrated in Fig. 3-2, which shows an oscillogram of a nonsinusoidal wave, both as a function of time and with the time axis compressed.



Fig. 3-2. The cause of variations in intensity of oscillograms.

**3-3. The Use of Inclined Deflecting Plates.**[2] If one gradually increases the alternating potential applied to the deflecting plates of a cathode-ray tube, the length of the line on the screen will continue to increase within certain limits and then remain constant, although the bright end points, of which mention was previously made, will disappear. This is on the assumption that the screen is sufficiently large so that the line has not already spread out over the entire screen before this effect is noted. The explanation of this effect is evident. The condition arises when the deflecting field is so large that the electrons strike the deflecting plates and so never reach the screen.

To remedy this and so make larger deflections possible, the plates of the deflecting system are inclined with respect to each other, as illustrated in Fig. 3-3. The photograph of the modern cathode-ray tube, shown in Fig. 3-8, clearly shows the inclined deflecting plates. Evidently, the increased distance between the inclined portion of the plates will cause a weakened field for a given deflecting potential, thereby resulting in a smaller deflection and so in a decreased sensitivity. This weakened field is
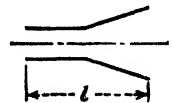


Fig. 3-3. Inclined plates in a cathode-ray tube.

counteracted either by decreasing the separation of the plates (thereby increasing the field) or by increasing $l$ (thereby subjecting the electron to a deflecting force for a longer interval of time). In this way the sensitivity of the tube may actually be increased by 20 to 30 per cent, and at the same time deflections over the full screen are possible.

**3-4. The Intensifier Tube.** It is shown in Sec. 3-1 that the sensitivity varies inversely with the accelerating potential $E_a$. Consequently, greater sensitivity prevails for smaller values of $E_a$. However, smaller values of $E_a$ cause a decrease of the luminosity of the spot on the screen, since the luminosity depends upon the energy carried by the beam of electrons (see Sec. 3-10). It is therefore necessary to reach a compromise between luminosity and sensitivity in the design of a cathode-ray tube.
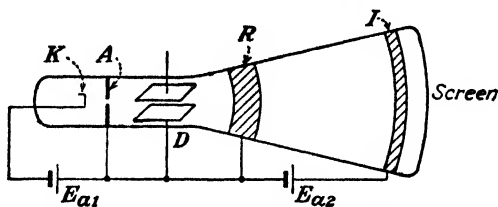


FIG. 3-4. A cathode-ray tube provided with a postaccelerating electrode.

It is however possible to obtain high sensitivity and at the same time high luminosity by the use of an additional electrode called the *intensifier*,[3] *postaccelerating electrode*, or *third anode* which is placed close to the screen. The general appearance of such a tube is shown in Fig. 3-4. This extra electrode $I$ is usually in the form of a conducting ring on the inner surface of the glass near the screen. It serves to give the electrons an added acceleration toward the screen subsequent to their deflection by the deflecting plates. That is, the sensitivity is increased through the use of a lower accelerating potential $E_{a1}$ prior to the deflection of the beam, and the intensity of the spot on the screen is increased through the use of the intensifier voltage $E_{a2}$, which speeds them up subsequent to the deflection. As usually operated, the potential between the intensifier electrode and the accelerating anode is approximately equal to that between the anode and the cathode. This provides full brilliance with a sensitivity approximately twice that of a tube not provided with the intensifier electrode. In order to avoid a loss in sensitivity resulting from an axial acceleration in the region of the deflecting plates caused by the field of the intensifier electrode, the anode is connected to a conducting ring $R$ painted on the inner surface of the glass beyond the deflecting plates (see Fig. 3-4). This ring may take the form of an aquadag coating extending over quite a region on both sides of the deflecting plates. This renders the deflecting plate space free from an axial component of force. Such an axial force

would, of course, shorten the time the electron would be under the influence of the deflecting voltage, which, in turn, would give a decreased deflection.

If the ratio of intensifier to anode potential is much greater than 2:1, then it is found that distortions are introduced.[4] These are due in a large measure to the nonaxial components of electric field caused by the intensifier. The distortions are minimized by changing the shape of the envelope from a tapered to a cylindrical form and by applying the intensi-
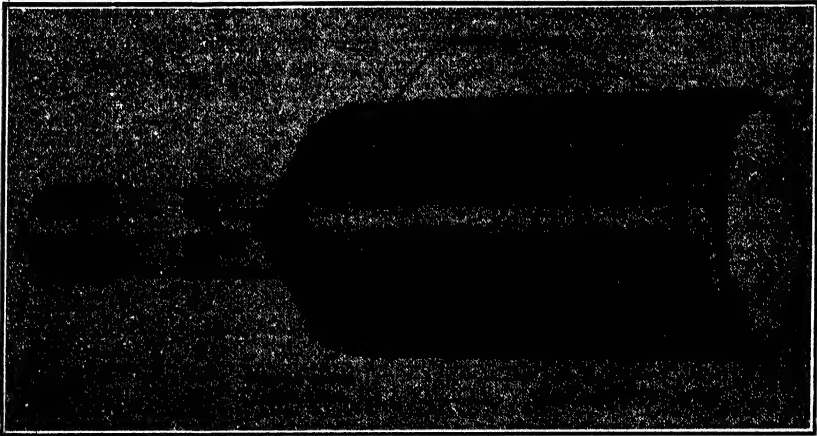


FIG. 3-5.   Type 5RP multiband intensifier tube.   (*Courtesy of A. B. Du Mont Laboratories, Inc.*)

fier voltage in several equal steps.   Figure 3-5 shows such a tube operating satisfactorily at an intensifier-to-anode voltage of 10:1.   The electrons are accelerated through a potential of 20 kv before striking the screen and yet the sensitivity is as high as 0.15 mm/volt.   This tube has sufficient intensity to photograph very rapid transients, *writing speeds* (see Sec. 3-10) as high as 400 in./μsec having been recorded.   Before the advent of the intensifier-type tube this could be done only with a demountable-type tube in which the photographic plate was placed inside of the tube envelope so that the electron beam exposed it directly.[5]   This required that the system be connected continually to vacuum pumps, a distinct disadvantage over the sealed-off intensifier-type tube.

A second application of the intensifier tube is as a projection oscillograph.

**3-5. Magnetic Deflection in Cathode-ray Tubes.**   The illustrative example in Sec. 2-11 immediately suggests that a cathode-ray tube may employ magnetic as well as electric fields in order to accomplish the de-

flection of the electron beam.   However, as it is not feasible to use a field extending over the entire length of the tube, a short coil furnishing a transverse field in a limited region is employed, as shown in Fig. 3-6.   The magnetic field is taken as pointing out of the paper, and the beam is deflected upward.  It is assumed that the magnetic-field intensity $B$ is uniform in the restricted region shown and is zero outside of this area.   Hence, the electron moves in a straight line from the cathode to the boundary $O$ of the magnetic field.   In the region of the uniform magnetic field the electron experiences a force of magnitude $evB$, where $v$ is the speed.
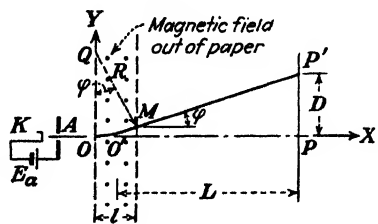
The path $OM$ will be the arc of a circle whose center is at $Q$.  The speed of the particles will remain constant and equal to

$$v = v_{0x} = \sqrt{\frac{2eE_a}{m}} \qquad \text{m/sec} \tag{3-6}$$

The angle $\varphi$ is, by definition of radian measure, equal to the length of the arc $OM$ divided by $R$, the radius of the circle, or approximately *

$$\varphi \approx \frac{l}{R} \tag{3-7}$$

where, by Eq. (2-38),

$$R = \frac{mv}{eB} \qquad \text{m} \tag{3-8}$$

In most practical cases, $L$ is very much larger than $l$, so that little error will be made in assuming that the straight line $MP'$, if projected backward, will pass through the center $O'$ of the region of the magnetic field. Then

$$D \approx L \tan \varphi \tag{3-9}$$

These assumed conditions presuppose that the angle $\varphi$ is a small quantity, thereby permitting the approximation $\tan \varphi \approx \varphi$.  By Eqs. (3-6), (3-7), and (3-8), Eq. (3-9) now becomes

$$D \approx L\varphi = \frac{lL}{R} = \frac{lLeB}{mv} = \frac{lLB}{\sqrt{E_a}} \sqrt{\frac{e}{2m}}$$



FIG. 3-6.   Magnetic deflection in a cathode-ray tube.

* Either the symbol $\approx$ or $\doteq$ is used to designate "approximately equal to."

The deflection per unit magnetic-field intensity, $D/B$, given by

$$\frac{D}{B} = \frac{lL}{\sqrt{E_a}} \sqrt{\frac{e}{2m}} \qquad m/(weber/m^2) \qquad (3\text{-}10)$$

is called the *magnetic-deflection sensitivity* of the tube. It is observed that this quantity is independent of $B$. This condition is analogous to the electric case for which the electrostatic sensitivity is independent of the deflecting potential. However, in the electric case the sensitivity varies inversely with the anode voltage, whereas it here varies inversely with the square root of the anode voltage. Another important difference is in the appearance of $e/m$ in the expression for the magnetic sensitivity, whereas this ratio did not enter into the final expression for the electric case. Because the sensitivity increases with $L$, the deflecting coils are placed as far down the neck of the tube as possible, usually directly after the accelerating anode.

**3-6. Focusing the Beam.** The essentials of a high-vacuum tube are illustrated in Fig. 3-7. (The pressure is about $10^{-6}$ mm Hg.) The elec-
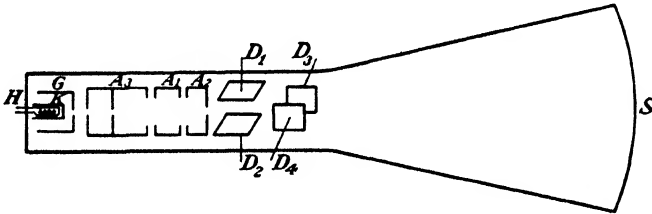


Fig. 3-7. The elements of a high-vacuum cathode-ray tube. (*Courtesy of A. B. Du Mont Laboratories, Inc.*)

trons are emitted from the indirectly heated oxide-coated cathode $K$, the filament power being supplied to the heater $H$. The heater is generally noninductively wound. The intensity of the beam is controlled by varying the negative bias on the grid $G$, which is usually a nickel cylinder provided with a small, centrally located hole coaxial with the tube axis.

Because of the mutual repulsion of the electrons a method for focusing the beam into a small spot must be provided. The electrostatic method of doing this is indicated in Fig. 3-7. The electrode $A_3$ is a metal cylinder that contains several baffles in order to collimate the beam and accelerate it to its final velocity. The electrodes $A_1$ and $A_2$ are cylinders coaxial with $A_3$. The cylinder $A_2$ is electrically connected to $A_3$, but $A_1$ is at a lower voltage. Focusing of the electron beam is accomplished by varying the potential between $A_1$ and $A_2$, thereby changing the index of refraction of the electron rays (see Sec. 2-18). Because of this, $A_1$ is referred to as the "focusing electrode" or the "first anode," $A_2$ is called the "second anode," and $A_3$ is known as the "preaccelerating electrode." The com-

1. Key
2. Pins
3. Base
4. Heater element
5. Cathode
6. Control grid

7. Preaccelerating
   electrode
8. Electron beam
9. Focusing electrode
10. Ceramic support
11. Spider support
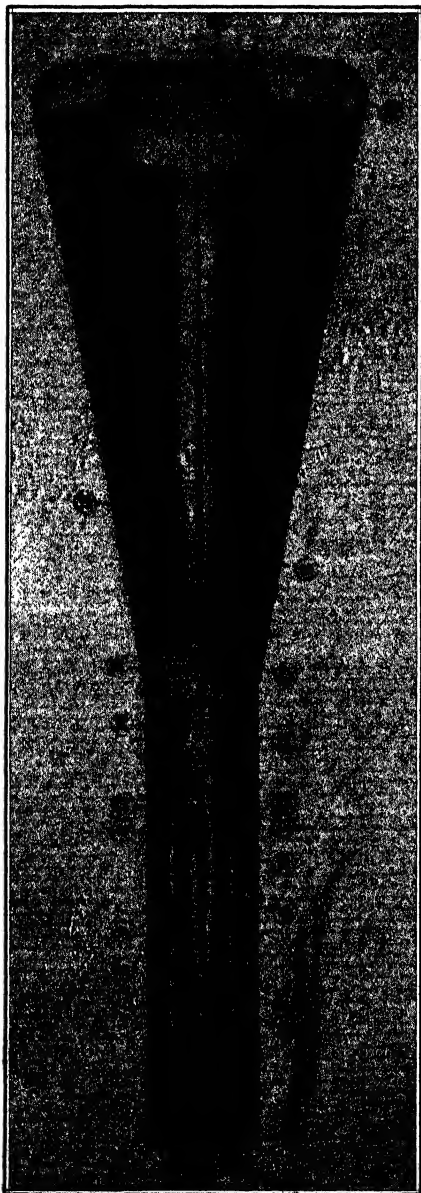
12. Mica support
13. Accelerating
    electrode
14. Deflection plate D3
15. Deflection plate D4
16. Deflection plate D1

17. Deflection plate D2
18. Internal conductive
    coating
19. Flourescent screen
    material
20. Pattern

FIG. 3-8.  Complete modern electrostatic cathode-ray tube.  (*Courtesy of A. B. Du Mont Laboratories, Inc.*)

bination of $H$, $K$, $G$, $A_1$, $A_2$, and $A_3$ is called an "electron gun," [6] an appropriate name because of the analogy with ordinary ballistics. A complete picture of an electrostatic cathode-ray tube is given in Fig. 3-8. The lens action of the electron gun on the electron beam is clearly indicated in the diagram.

In addition to the electrostatic method of focusing, just described, it is possible to focus the beam magnetically. The method of Sec. 2-12 employing a longitudinal magnetic field over the entire length of the tube is not too practical. Hence, a short coil is placed over the neck of the tube. One type of coil and the corresponding magnetic lines are shown in Fig. 3-9.
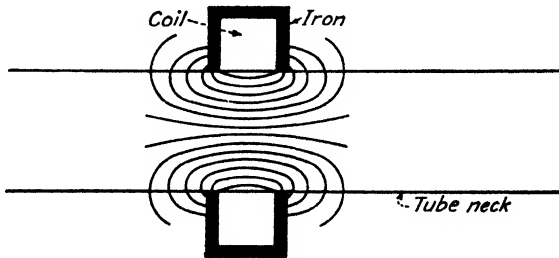


FIG. 3-9.   Magnetic focusing coil.

There are two components of force on the electron, one due to the axial component of velocity and the radial component of field and the second due to the radial component of the velocity and the axial component of field. Because the axial component of velocity is so much greater than the transverse component, then, to a good approximation, only the former force need be taken into account. The analysis is complicated,[7] but it can be seen qualitatively that the motion will be a rotation about the axis of the tube and, if conditions are correct, the electron on leaving the region of the coil may be turned sufficiently so as to move in a line toward the center of the screen. A rough adjustment of the focus is obtained by positioning the coil properly along the neck of the tube. The fine adjustment of focus is made by controlling the coil current.

A picture of a tube with magnetic deflection and focus is shown in Fig. 3-10. Note that the preaccelerating electrode and the focusing electrode of the electrostatic tube are missing and instead an electrode called the "screen grid" is used. This is held at a fixed positive voltage, usually 250 volts, with respect to the cathode and allows the tube to be operated over a wide range of accelerating voltages without appreciably changing the control-grid characteristics of the tube.

**3-7. Charges in a Cathode-ray Tube.** There is the likelihood of the accumulation of electrons on the surface of the glass in a vacuum tube. If these charges are not symmetrically distributed, they may cause spurious
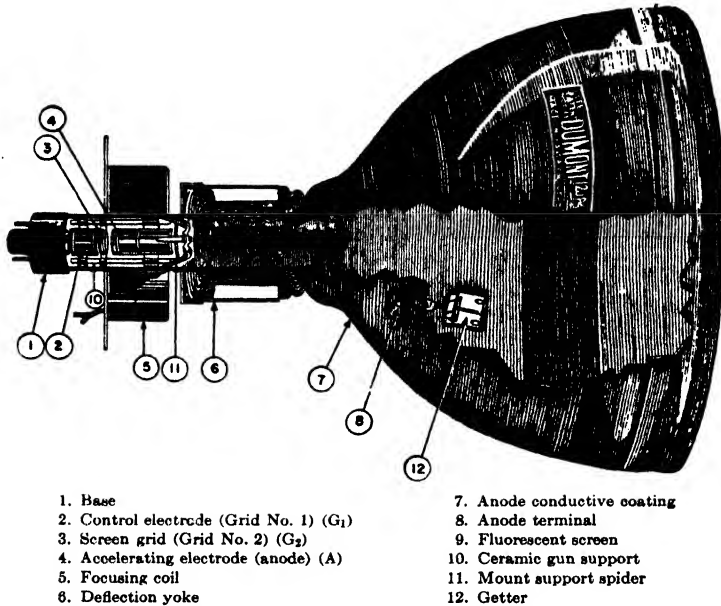
1. Base
2. Control electrode (Grid No. 1) ($G_1$)
3. Screen grid (Grid No. 2) ($G_2$)
4. Accelerating electrode (anode) (A)
5. Focusing coil
6. Deflection yoke
7. Anode conductive coating
8. Anode terminal
9. Fluorescent screen
10. Ceramic gun support
11. Mount support spider
12. Getter

FIG. 3-10. Typical cathode-ray tube with magnetic focusing and deflection. (*Courtesy of A. B. Du Mont Laboratories, Inc.*)

deflections of the electron beam.   This effect can sometimes be observed by touching the glass envelope of the tube of an oscilloscope with one's fingers.   This disturbs the charge distribution on the glass, resulting in a deflection of the spot.

In order to assist in the removal of undesirable stray charges, the inside surface of the glass envelope of most tubes is given a coating of a conducting material, which is then electrically connected to the second anode (see Figs. 3-8 and 3-10).   Usually an aqueous solution of graphite, known as *aquadag* or *dixonac*, is used.   In some magnetic tubes the anode is dispensed with entirely, the conducting coating on the glass walls being used as the final accelerating electrode.

Since both the screen material and the uncoated portion of the glass envelope are good electrical insulators, what happens when the electron beam strikes the screen?   The answer is found in the secondary-emission properties of the screen material and the glass on which it is deposited. For the materials usually employed, between one and two secondary electrons are emitted for each incident electron.   However, in the equilibrium state as many electrons must leave the screen as reach it.   Experimentally, it is found that the screen acquires a potential of 1 or 2 volts positive with respect to the second anode.   Thus, although more secondary electrons

are emitted than the number contained in the primary beam, not all of these can surmount the retarding field between the screen and the second anode and some will return to the screen. Thus, the equilibrium condition of zero net current to the screen is attained.

In addition to electrons, many types of *negative* ions (oxygen, carbon, chlorine, etc.) have been discovered [8] in the cathode-ray tube. The chief source of these ions is the oxide-coated cathode. In Sec. 3-5 it is shown that the deflection due to a magnetic field varies as $(e/m)^{\frac{1}{2}}$, where $m$ is the mass of the particle. Since these ions are thousands of times heavier than the electron, they are deflected very little and hence they continually strike a small area at the center of the screen. This soon results in the appearance of a dark stain or blemish. This "ion burn" is found principally in tubes with electrostatic focusing and magnetic deflection. If magnetic focusing is used, then the coil current for proper focusing depends upon the mass of the particle. If the electron beam is focused, then the ions are not in focus and they spread out over a considerable portion of the screen. Hence, the concentration is usually not great enough to cause a severe ion burn.

It is shown in Sec. 3-1 that the electrostatic deflection of a particle is independent of $e/m$. Hence, in a tube with electrostatic focusing and electrostatic deflection the ions travel with the electrons. They are not concentrated and do not produce an ion burn.

In order to avoid this ion difficulty, several forms of "ion trap" have been invented.[9] The simplest of these uses an electron gun inclined at an angle of about 30 deg with respect to the tube axis. A magnetic field transverse to the gun bends the electrons back to the axis but (because of their much greater mass) has very little effect on the ions so that they strike the ion trap and do not get to the screen.

**3-8. Comparison of Electric and Magnetic Tubes.** Electrostatic tubes are used in all oscilloscope applications where operation over a wide range of frequency is desired. The upper frequency limit (about 10 megacycles) is determined by the capacitance between the deflecting plates (a few micromicrofarads) and the capacitance of the associated amplifier and leads. Higher frequencies can be observed by using tubes specially designed for the purpose and having very small plates brought out directly through the glass neck to coaxial connectors.

Magnetic deflection has a much smaller band pass (about 10 kc) because of the inductance of the coils. Thus, the same input voltage at different frequencies will result in different coil currents and hence different deflections. For this reason magnetic deflection is limited essentially to tubes involving a constant sweep frequency, such as television or radar. (In these applications the high-frequency signals are applied to the grid or to the cathode to give intensity variations of the picture.)

Magnetic focusing is almost always used with magnetic deflection. This combination has the following advantages: A high-intensity beam can be produced with a minimum tube length and a minimum distortion of the spot as a result of deflection.  As explained in Sec. 3-7, ion burn is not severe although sometimes an ion trap is used.  The magnetic system is structurally simpler than the electrostatic, and hence production costs are lower.  The magnetic tube is also more rugged, which is an advantage in mobile equipment.  It is seen from Eq. (3-10) that the sensitivity varies inversely as the square root of the accelerating voltage, and hence for the high anode voltages needed to give a bright picture the deflecting field need not be increased proportionately.  Thus, to obtain the same deflection at 16 kv as at 1 kv, the deflecting field need be increased only four times.

Electrostatic tubes require very little power, and the auxiliary circuits are somewhat simpler.  Another advantage is that the focus depends much less upon power-supply variations than in a magnetic tube.  This is due to the fact that the focusing-electrode voltage is obtained from a bleeder across the accelerating-voltage power supply (see Fig. 3-12), and hence these two voltages will vary in proportion and the spot will stay focused for small variations in d-c supply.

For some applications (for example, one type of radar indicator) it is necessary to have a rotating radial deflection.  This is easily obtained with magnetic deflection since the sweep coil can be rotated mechanically around the neck of the tube.

**3-9. Sweep Action in a Cathode-ray Tube.**  An electrostatic tube has two sets of deflecting plates which are at right angles to each other in space (as indicated in Figs. 3-7 and 3-8).  These plates are referred to as the *vertical* and *horizontal* plates because the tube is oriented in space so that the potentials applied to these plates result in vertical and horizontal deflections, respectively.  The reason for having two sets of plates will now be discussed.

Suppose that a sinusoidal voltage is impressed across the vertical plates. This will result in a straight vertical line being traced on the fluorescent



Fig. 3-11.  Sweep voltage for a cathode-ray tube.

screen.  Suppose now that the sinusoidal wave is removed and that a "saw-tooth" wave is impressed across the horizontal plates.  A saw-tooth wave of the type here considered is shown in Fig. 3-11.  This voltage increases linearly for a time $T$ and then suddenly drops to zero.  It then repeats its linear rise for the next cycle and again falls to zero, etc.  This is called a *sweep voltage*, and the horizontal plates to which this potential is

applied are often referred to as the *sweep-circuit* plates. The application of such a potential to the horizontal plates will cause a sidewise force to act on the cathode-ray beam. As a result of this force, the beam will be deflected across the screen in the time $T$ as the voltage increases linearly but will return to its starting point very quickly as the saw-tooth wave rapidly falls to zero at the end of each period.

If the sinusoidal voltage is impressed across the vertical plates when, simultaneously, the sweep voltage is impressed across the horizontal plates, the sinusoidal voltage, which of itself gives rise to a vertical line, will now be spread out and will appear as a sinusoidal trace on the screen. The pattern will appear stationary only if the time $T$ is equal to or is some multiple of the time for one cycle of the wave on the vertical plates. It is then necessary that the frequency of the sweep circuit be adjusted to synchronize with the frequency of the applied wave.

This discussion is readily understood in the light of the theory of motion of an electron in a uniform electric field. When passing between the vertical plates, the electrons receive a deflection that is proportional to the voltage on this set of plates at this instant. The electron now reaches the region of the horizontal plates where it receives an additional sidewise deflection that is proportional to the potential on these plates. Since, for a linear sweep, this sweep voltage is proportional to the time, the horizontal deflection is proportional to the time. Hence, a graph of voltage vs. time appears on the screen.

Actually, of course, the voltage impressed on the vertical plates may have any wave form. Consequently, a system of this type provides an almost inertialess oscilloscope for viewing arbitrary wave shapes. This is one of the most common uses for cathode-ray tubes. If a nonrepeating sweep voltage is applied to the horizontal plates, it is possible to study transients on the screen. This requires a system for synchronizing the sweep with the start of the transient.

A magnetic tube has two coils oriented at right angles to perform the same functions as the two sets of plates in an electrostatic tube.

**3-10. Screens for Cathode-ray Tubes.**[10] *Luminescence* is the production of light by means other than heating (for example, by the bombardment of a screen by electrons). *Fluorescence* is luminescence during excitation, and *phosphorescence* is luminescence after the excitation has stopped (when the electron beam has been turned off). The phosphorescence, or persistence, of a screen material may be classified generally as short (microseconds), medium (milliseconds), or long (seconds)

Standard cathode-ray tubes are classified with a phosphor code designation such as P1, P2, etc. The type P1 produces a green trace of medium persistence and is well suited for general-purpose oscilloscope work. The type P2 produces a bluish-green trace with a long-persistent yellow phos-

phorescence and is useful for visual observations of transient signals, low-frequency recurrent signals, and radar pulses.    The type P4 produces a medium persistence white phosphorescence.    This is the screen used in television tubes.    The type P7 produces a blue fluorescence and a long-persistent yellow phosphorescence.    It is excellent for radar work and for transient oscillography.    The type P11 produces a very short blue persistence and is used for photographic work, particularly of high-speed phenomena.    Complete characteristics of these and other types of screens are available from the manufacturers.

The luminous-output characteristics of the screen will depend upon a number of factors, such as the beam current, the accelerating voltage (the product of these two gives the beam power), the size of the spot, the chemical and physical make-up of the screen, the side from which the screen is viewed, and the length of time a given area on the screen is bombarded by the electrons.    In a given cathode-ray tube the luminescence can be altered by varying the energy with which the impinging electrons bombard a given spot on the screen or by controlling the number of electrons that get to the given area in a given time, *i.e.*, the electron current.    To increase the energy of the impinging electrons, one must increase the accelerating potential.    The beam current is most easily varied by means of variations of the grid-to-cathode potential.

Strictly speaking, this discussion applies to a stationary beam.    If a pattern exists on the screen, local variations in intensity may result from variations in the rate of arrival of the electrons at various points on the screen (see Fig. 3-2).    This rate is controlled by the wave shape of the potential applied to the deflecting plates.    It is because of this that one talks about a "writing" or "tracing" speed.    That is, if the writing rate is too great over a portion of a curve, this part may not be at all visible on the screen.    For example, with a well-designed sweep circuit, the sharp vertical trace at the end of each period should not be visible (see Fig. 3-11).

**3-11. Cathode-ray-tube Connections.**    The connections for a typical electrostatic cathode-ray tube are shown in Fig. 3-12.    The d-c supply for the accelerating and focusing system is obtained from a thermionic-tube rectifier.    The resistors $R_1$, $R_2$, and $R_3$ represent a voltage divider providing the correct voltages for operation.    Focusing of the beam is accomplished by adjusting the ratio between the voltages on anodes 2 and 1. The magnitude of the voltage on the first focusing electrode is usually about one-fourth to one-fifth that of the accelerating anode $A_2$.    By varying the voltage on anode 2, regulation of the spot size and intensity may be accomplished.    If the voltage of anode 2 is increased, the electron speed is increased, thereby decreasing the spot size and increasing the intensity. The current, or number of electrons per second, to anode 2 may be increased by decreasing the bias voltage applied to the control grid.    Thus,

in an oscilloscope operating at a fixed accelerating voltage the intensity of the beam is controlled by adjusting the grid voltage.

It is noted that in Fig. 3-12 the final anode and one plate of each set are grounded. This means that the cathode is below ground by an amount equal to the accelerating voltage. The chief reason for this arrangement is that the wave form to be observed is applied to the plates. Hence, to protect the operator from high-voltage shock when making this connection, the plates should be near ground potential. Also, the deflecting voltages are measured with respect to ground and hence high-voltage
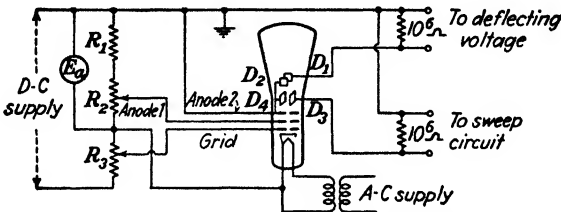


Fig. 3-12.   Connections for a typical electrostatic high-vacuum cathode-ray tube.

blocking capacitors and additional shunt capacitances with respect to ground are avoided by this connection.

In applications, such as television or radar, where the signal is used to control the intensity of the picture the operator must have access to the cathode-grid region of the tube. For reasons of safety in these cases, it is customary to ground this portion of the circuit. The anodes are then at high voltages (suitably insulated) with respect to ground.

In tubes using the intensifier principle a separate high-voltage supply is ordinarily used for the intensifier electrodes.

In Fig. 3-12 one plate of each set is shown internally connected to the second anode. This reduces by two the relatively large number of external connections that must be made to the tube. However, this simplified design results in the two sets of plates not being independent of one another. Consequently defocusing of the spot and distortion of the sweep may result.[11] These difficulties are avoided by bringing out each plate lead separately and applying the deflecting voltages in a balanced manner, i.e., one plate is driven positively while the other is driven negatively the same amount with respect to ground.

**3-12. The Magnetic and Mass Spectrographs.** The charge, mass, and velocity of electrons and ions emitted from various sources under different types of excitation have been determined experimentally by applying known electric and magnetic fields to regions through which these particles are directed and studying the subsequent motions. Several applications of the theory of Sec. 2-11 to these problems will now be given.

Suppose that $PP'$ (Fig. 2-7), which is reproduced in Fig. 3-13, is a photographic plate with a small opening at $O$ through which charged particles may be introduced into the magnetic field. If the particle is an electron, it will be deflected to the right as shown (the reader should verify this by means of the direction rule) and will expose the plate at $O'$. The distance $OO'$ equals the diameter of the circular path and if $e/m$ is known, $v$ can be calculated with the aid of Eq. (2-38). It is in this way that the velocity of the electrons emitted from radioactive matter (the $\beta$-rays) is measured.

If, instead of the radioactive substance at $O$, there is a metal surface being bombarded by electrons, the velocity of the secondary electrons emitted can be measured in the manner indicated above. Similarly, if the surface at $O$ has radiation, either visible or invisible, falling on it, then the velocity of the emitted photoelectrons can be studied. An apparatus of the type here considered is known as a *magnetic spectrograph*.

If the particles entering the field at $O$ have a known energy, for example, if the potential through which they have fallen is known, it is then possible from the setup of Fig. 3-13 to determine the mass of the particle, provided that the charge of the particle is known. If, therefore, particles possessing the same charge but differing slightly in mass are accelerated through the same potential, the photographic plate would reveal several different spots. It was shown in essentially this way that most



FIG. 3-13. The principle of the magnetic or the mass spectrograph.

chemical elements consist, not entirely of particles of one mass, but rather of a mixture of atoms of slightly different mass. Atoms of the same element that possess the same electronic charge and differ but slightly in mass are known as *isotopes*. For example, it has been found that oxygen (atomic weight 16.000) has three stable isotopes. The instruments used to measure the masses of the isotopes that are based upon these considerations, although they may differ widely in detail, are known as *mass spectrographs*.[12]

Commercial application has been made of mass spectrographs for the quantitative and qualitative measurement of the components of gaseous mixtures [13] and for the detection of leaks in large vacuum systems.[14] In the latter application the mass spectrograph is connected to the vacuum chamber, and a stream of helium is played over the surface of the system where a leak is suspected. If there actually is a leak at this point, a reading is obtained on an output meter. The mass spectrograph contains an electron gun for ionizing the helium atoms and a voltage supply for accelerating the ions before they enter the magnetic field. The sensitivity of such a system is claimed to be one part helium to a few hundred thousand parts of air.

If ions of a known mass and velocity are under consideration, then this magnetic analysis can be used to determine the charge on the ions. This has been used to study the various stages of ionization in a discharge tube.

**3-13. The Cyclotron.** The principles of Sec. 2-11 were first employed by Lawrence and Livingston [15] to develop an apparatus called a *magnetic resonator* or *cyclotron*. This device imparts very high energies (tens of millions of electron volts) to positive ions. These high-energy positive ions are then allowed to bombard some substances which then become radioactive and generally disintegrate. Because of this, the cyclotron has popularly become known as an *atom smasher*.

The basic principles upon which the cyclotron operates are best understood with the aid of Figs. 3-14 and 3-15. The essential elements are the "dees," the two halves of a shallow, hollow, metallic "pillbox" which has been split along a diameter as shown; a strong magnetic field which is normal to the plane of the paper; and a high-frequency a-c potential applied to the dees.
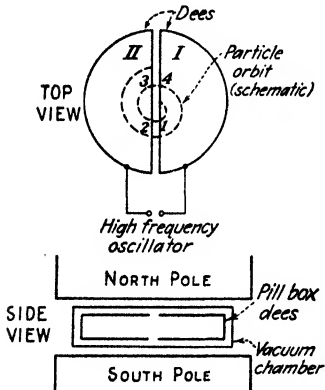


Fig. 3-14.  The cyclotron principle.

A moving positive ion released near the center of the dees will be accelerated in a semicircle by the action of the magnetic field and will reappear at point 1 at the edge of dee I. Assume that dee II is negative at this instant with respect to dee I. Then the ion will be accelerated from the point 1 to point 2 across the gap and will gain an amount of energy corresponding to the potential difference between these two points. Once the ion passes inside the metal dee II, the electric field no longer acts on the ion, while the magnetic field causes it to move · in the semicircle from point 2 to point 3. If the frequency of the applied a-c potential is such that the potential has reversed in the time necessary for the ion to go from point 2 to point 3, then dee I is now negative with respect to dee II and the ion will be accelerated across the gap from point 3 to point 4. With the frequency of the accelerating voltage properly adjusted to this "resonance" value, the ion continues to receive pulses of energy corresponding to this difference of potential again and again.

Thus, after each half revolution the ion gains energy from the electric field, resulting, of course, in an increased velocity. The radius of each semicircle is then larger than the preceding one, in accordance with Eq. (2-38), so that the path described by the whirling ion will approximate a planar spiral.

FIG. 3-15.   The component parts of a cyclotron.   (*J. J. Livingood.*)

*Example.*   Suppose that the oscillator that supplies the power to the dees of a given cyclotron imparts 50,000 volts to heavy hydrogen atoms (deuterons) each of atomic number 1 and atomic weight 2.0147, at each passage of the ions across the accelerating gap.   Calculate the magnetic-field intensity, the frequency of the oscillator, and the time that it will take for an ion introduced at the center of the chamber to emerge at the rim of the dee with an energy of 5 million electron volts (5 Mev).   Assume that the radius of the last semicircle is 15 in.

*Solution.*   The mass of the deuteron is

$$m = 2.01 \times 1.66 \times 10^{-27} = 3.34 \times 10^{-27} \text{ kg}$$

The velocity of the 5-Mev ions is given by the energy equation

$$\tfrac{1}{2}mv^2 = (5 \times 10^6)(1.60 \times 10^{-19}) = 8.00 \times 10^{-13} \text{ joule}$$

or

$$v = \left(\frac{2 \times 8.00 \times 10^{-13}}{3.34 \times 10^{-27}}\right)^{\frac{1}{2}} = 2.20 \times 10^7 \text{ m/sec}$$

The magnetic field, given by Eq. (2-38),

$$B = \frac{mv}{eR} = \frac{(3.34 \times 10^{-27})(2.20 \times 10^7)}{(1.60 \times 10^{-19})(15 \times 2.54 \times 0.01)} = 1.20 \text{ webers/m}^2$$

is needed in order to bring these ions to the edge of the dees.

The frequency of the oscillator must be equal to the reciprocal of the time of revolution of the ion. This is, from Eq. (2-40),

$$f = \frac{1}{T} = \frac{eB}{2\pi m} = \frac{1.60 \times 10^{-19} \times 1.20}{2\pi \times 3.34 \times 10^{-27}}$$

$$= 9.15 \times 10^6 \text{ cps} = 9.15 \text{ megacycles/sec}$$

Since the ions receive 5 Mev energy from the oscillator in 50-kv steps, they must pass across the accelerating gap 100 times. That is, the ion must make 50 complete revolutions in order to gain the full energy. Thus, from Eq. (2-40), the time of flight is

$$t = 50T = \frac{50 \times 1}{9.15 \times 10^6} = 5.47 \times 10^{-6} \text{ sec}$$

In order to produce a uniform magnetic field of 1.2 webers/m² over a circular area whose radius is at least 15 in., with an air gap approximately 6 in. wide, an enormous magnet is required, the weight of such a magnet being of the order of 60 tons. Also, the design of a 50-kv oscillator for these high frequencies and the method of coupling it to the dees present some difficulties, since the dees are in a vacuum-tight chamber. Further, means must be provided for introducing the ions into the region at the center of the dees and also for removing the high-energy particles from the chamber, if desired, or of directing them against a target.

The bombardment of the elements with the high-energy protons, deuterons, or helium nuclei which are normally used in the cyclotrons renders the bombarded elements radioactive. These radioactive elements are of the utmost importance to physicists, since they permit a glimpse into the constitution of nuclei. They are likewise of extreme importance in medical research, since they offer a substitute for radium. Radioactive substances can be followed through any physical or chemical changes by observing their emitted radiations. This "tracer" or "tagged-atom" technique is used in industry, medicine, physiology, and biology.

It is shown in Sec. 2-8 that if an electron falls through a potential of more than 3 kv a relativistic mass correction must be made, indicating that its mass increases with its energy. Thus, if electrons were used in a cyclotron, their angular velocity would decrease as their energy increased and they would soon fall out of step with the high-frequency field. For this reason electrons are not introduced into the cyclotron.

For positive ions whose mass is several thousand times that of the electron the relativistic correction becomes appreciable when energies of a few tens of millions of electron volts are reached. For greater energies than these the ions will start to make their trip through the dees at a slower rate and slip behind in phase with respect to the electric field. This difficulty is overcome in the *synchro-cyclotron* or *f-m cyclotron* by decreasing the frequency of the oscillator (frequency modulation) in accordance with the decrease in the angular velocity of the ion. In 1946 the 184-in. f-m cyclotron at the University of California accelerated deuterons to 200 Mev

and $\alpha$-particles to 400 Mev.[16]    The frequency is varied between 12.6 and 9.0 megacycles/sec at a modulation rate of 120 cps.    In 1951 the Columbia University f-m cyclotron accelerated protons to 385 Mev.

It is possible to give particles energies in excess of those for which the relativistic correction is important by keeping the oscillator frequency fixed provided that the magnetic field is slowly increased in step with the increase in the mass of the ions so as to maintain a constant angular velocity.    Such an instrument is called a *synchrotron*.    In 1947, electrons were accelerated to 70 Mev in a synchrotron at the research laboratory of the General Electric Co.[17]    The electrons are injected from a gun, which gives them a velocity approaching that of light.    Since the radius of the orbit is given by $R = mv/Be$ and since the ratio $m/B$ is kept constant and $v$ changes very little, there is not much of an increase in the orbit as the energy of the electron increases.    The vacuum chamber is built in the form of a "doughnut" instead of the cyclotron "pillbox."    The magnet has the form of a hollow cylinder since there is need for a magnetic field only transverse to the path.    This results in a great saving in weight and expense.    The dees of the cyclotron are replaced by a single-cavity resonator.

In 1950 two proton synchrotrons were under construction, one at the Brookhaven National Laboratories [18] and the other at the University of California.    Proton energies in the 1 to 10 Bev range are expected.

**3-14. The Betatron.**[19]    In Sec. 2-11 it is shown that no energy can be gained by a particle in a magnetic field that is constant in time.    However the situation is completely different if the field changes with time.    Consider an electron moving with a speed $v$ in a magnetic field whose instantaneous value is $B_0$.    If the rate of change of the magnetic field is small compared with the time required for the electron to traverse its path, this path will be almost circular, with a radius $r_0$, where

$$r_0 = \frac{mv}{eB_0} \tag{3-11}$$

However, the tangential velocity $v$ is not a constant, because a tangential force is acting on the electron.    This follows from the fact that an amount of flux $\Phi$ (webers) is enclosed by the electron in traversing this path, and as this flux is changing with time, an emf equal to $d\Phi/dt$ is induced, in accordance with Faraday's law of induction.    Thus work is done on the electron over the distance of the circumference of the circle of radius $r_0$.    Consequently, there is an effective tangential electric field acting on the electron given by

$$\mathcal{E} = \frac{1}{2\pi r_0} \frac{d\Phi}{dt} \tag{3-12}$$

Since the energy of the particle increases as it traverses its orbit, the

tangential velocity will change and as a result the radius will change. A particle in such a field will spiral outward as the magnetic field increases, picking up energy as it continues in its path. However, if the radial variation of the magnetic field is properly chosen, the radius $r_0$ of the electron orbit can be made to remain unchanged. This equilibrium condition will now be investigated.

Since the force in the tangential direction equals the rate of change of momentum in the tangential direction,

$$f = \frac{d}{dt}(mv) = e\mathcal{E}$$

or

$$\frac{e}{2\pi r_0}\frac{d\Phi}{dt} = \frac{d}{dt}(mv) \qquad (3\text{-}13)$$

Integrate this equation subject to the initial conditions that the electron is injected into the field when the flux is zero and with a velocity small compared with its final velocity. There results

$$\frac{e\Phi}{2\pi r_0} = mv \qquad (3\text{-}14)$$

where $\Phi$ is the final flux and $v$ is the final speed. Using Eq. (3-11), this is equivalent to

$$\Phi = 2\pi r_0{}^2 B_0 \qquad (3\text{-}15)$$

Since $\pi r_0{}^2$ is the area of the orbit, then this equation requires that the magnetic flux included within the orbit must be twice that which would be contained within the path if the field were uniformly distributed in space and equal to $B_0$. This is referred to as the 1:2 condition for orbit stability and is satisfied by having a strong central field and a tapered weaker field at the electron orbit.

The concepts discussed above are made use of in the *betatron* [19] invented by Kerst. In 1945 the betatron at the General Electric Co.[20] accelerated electrons to 100 Mev. The essential features of the betatron are shown in Fig. 3-16. The particles are injected into the doughnut-shaped vacuum chamber from an electron gun at energies of about 50 kv, near the equilibrium orbit. This gun is triggered for a few microseconds at the beginning of each cycle of the magnetic field. The electrons are accelerated during the time the magnetic field passes from zero to its peak value, or one-quarter of the cycle of the field current. If the electrons were to remain in the tube during the second quarter cycle when the field is decreasing, they would be decelerated and would give up their energy. They must be

removed before this happens. This requires destroying the 1:2 condition, in order to have the electrons leave the equilibrium orbit. It is accomplished by applying a short pulse to an auxiliary system of coils on the pole faces at the end of the first quarter cycle. This causes the electrons to spiral away from their orbit and hit a tungsten target, which then emits very penetrating X rays. The betatron is used as an X-ray tube for radiographing metal sections several feet thick.



FIG. 3-16.  The betatron principle.

It is customary now to apply a steady d-c biasing field which is equal to the peak value of the a-c field. The electrons are injected when the d-c and a-c fields cancel each other, and they remain in the accelerating chamber for one-half cycle (from peak to peak) instead of one-quarter cycle. Consequently the energy possible with the biased betatron is twice that of the unbiased type.

*Example.* The diameter of the orbit of an unbiased betatron is 66 in. The peak flux density at the orbit is 0.4 weber/m². The magnetic field varies at a 60-cycle rate.

a. What is the maximum energy acquired by the electrons?

b. What is the average energy imparted to an electron in each trip around the doughnut?

c. How many revolutions does the electron make?

d. What is the average transit time of the electron?

*Solution.* a. The energy is given by the relativistic equation (2-26),

$$eV = mc^2 - m_0c^2$$

and hence we must first find the maximum mass of the electron. From Eq. (3-11) the maximum momentum is

$$mv = eB_0r_0$$

and from Eq. (2-25)

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

so that

$$\frac{m_0v}{\sqrt{1 - v^2/c^2}} = eB_0r_0 \tag{3-16}$$

If the abbreviation $\beta = v/c$ is introduced, then

$$\frac{\beta}{\sqrt{1 - \beta^2}} = \frac{eB_0r_0}{m_0c} = \frac{(1.76 \times 10^{11})(0.4)(33 \times 2.54 \times 10^{-2})}{3 \times 10^8} = 197$$

Solving for $\beta$, a value very close to unity is obtained. Thus, the maximum electronic velocity is almost exactly equal to the velocity of light.

Putting $\beta = 1$ in the numerator of the above equation gives

$$\frac{1}{\sqrt{1-\beta^2}} = 197$$

and from Eq. (2-25) this equals $m/m_0$. Hence, the mass of the electron after acceleration is 197 times its mass at injection! Its final energy is

$$V = \frac{mc^2 - m_0c^2}{e} = \frac{m_0c^2}{e}(197 - 1)$$

$$= \frac{9 \times 10^{16} \times 196}{1.76 \times 10^{11}} = 10^8 \text{ ev, or 100 Mev}$$

*b.* The energy in electron volts given to the electron per trip around the doughnut is $d\Phi/dt$. The average value of this is the peak flux divided by the time it takes the flux to build up to this maximum value, which in this example is one-quarter of a cycle. By the 1:2 condition the peak flux is

$$2\pi r_0^2 B_0 = 2\pi(33 \times 2.54 \times 10^{-2})^2(0.4) = 1.80 \text{ webers}$$

and the time the flux takes to reach this value is $\frac{1}{4} \times \frac{1}{60} = \frac{1}{240}$ sec. Hence, the average voltage per trip is $1.80 \times 240 = 430$ volts. Since the flux varies sinusoidally with time, the voltage per trip varies from revolution to revolution. It is a maximum at injection and drops to zero at the end of the electronic journey. The average voltage per trip is the 430 volts calculated above.

*c.* The number of revolutions made by each electron during its acceleration cycle is

$$\frac{10^8 \text{ volts}}{430 \text{ volts/trip}} = 230,000 \text{ revolutions}$$

*d.* The average transit time is the

$$\left(\frac{1}{240} \text{ sec}\right)\left(\frac{1}{230,000 \text{ revolutions}}\right) = 1.79 \times 10^{-8} \text{ sec/revolution}$$

The transit time per revolution remains almost constant throughout the electron's journey. This can be seen as follows: As soon as the electron acquires an energy of 1 Mev, its velocity is 94 per cent that of light. Hence, as its energy increases from 1 to 100 Mev, its velocity changes by only 6 per cent (the increased energy manifesting itself in an increased mass rather than an increased speed). The transit time is given by $2\pi r_0/v$, and since $r_0$ and $v$ are almost constant, so is the time. For example, the transit time at the end of its journey is

$$\frac{2\pi r_0}{c} = \frac{2\pi(33 \times 2.54 \times 10^{-2})}{3 \times 10^8} = 1.75 \times 10^{-8} \text{ sec}$$

which agrees well with the average time calculated above.

The above illustrative example applies to the General Electric Co. 100-Mev betatron. The steel in the magnet of this instrument weighs 130 tons! It is of some interest to note that it is desirable to correct for the poor power factor which exists because the magnet is such a large inductive load. A tremendous bank of capacitors is used for this purpose. These constitute a large fraction of the total cost and space.

Since the electrons make 230,000 revolutions, then the effects of the mutual repulsions of the electrons and the scattering of the electrons by residual gases must be counterbalanced. It has been shown [21] that proper focusing can be obtained if the magnetic-field intensity in the region of the orbit varies as $1/r^n$, where $0 < n < 1$ provided that the field at the center of the magnet is so chosen as to satisfy the 1:2 condition.

Since the phase of the electron in its orbit is immaterial to the operation of the betatron, the relativistic change of mass with velocity introduces no complications.

In the synchrotron discussed in Sec. 3-13 the magnetic field varies with time, and hence some betatron action takes place in this instrument, also.

The question of the ultimate limit of the energies that might be attained from different types of particle accelerators is considered by Schiff.[22] He concludes that the practical limit to the betatron is approximately 500 Mev.

**3-15. Secondary-emission Multiplier.** An illustration of the practical use to which the theory of the cycloidal paths described in Sec. 2-14 has been applied is to be found in the secondary-emission multiplier devised by Zworykin, Morton, and Malter.[23]

It has already been mentioned in Sec. 1-3 that secondary electrons are emitted when a beam of high-velocity electrons is directed against a metal.
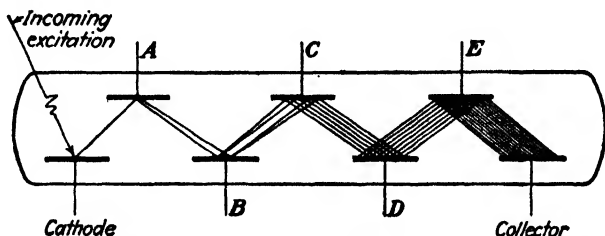


FIG. 3-17. Principle of operation of the secondary-emission multiplier.

More details of the mechanism of secondary emission are found in Sec. 4-13. For the present application, it is sufficient to know that the secondary-emission ratio, *i.e.*, the number of secondary electrons per incident primary electron, from a specially prepared Cs-CsO-Ag surface will be as high as 3 or 4 to 1, although the ratio depends upon the velocity of the impinging electron.

The principles of operation of the multiplier may best be understood by reference to Fig. 3-17. Light impinges upon the cathode and emits electrons photoelectrically. The electrons so emitted are accelerated toward plate $A$ by an applied electric field between the cathode and $A$. Upon colliding with $A$, secondary electrons are liberated. By making the potential of plate $B$ more positive than $A$, the secondary electrons from $A$ travel to plate $B$. In the same way, the secondary electrons released from plate $B$

travel to plate $C$, and so on, until the electrons are finally collected by the collector. If the ratio of secondary to primary electrons is $\gamma$ and if there are $n$ secondary-emission plates, the current at the collector is

$$i = i_0\gamma^n$$

where $i_0$ is the initial current at the cathode. The over-all gain is then $\gamma^n$ and may be very high (of the order of $10^6$ or more).

Although the multiplier shown schematically above is suitable to explain
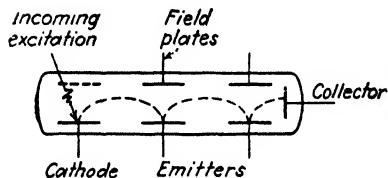


Fig. 3-18. The approximate cycloidal path in a magnetic secondary-emission electron multiplier. The magnetic field is perpendicular to the plane of the paper.

the theory of operation of the device, it would not operate successfully. The electrons from the cathode, instead of progressing successively from one emitter plate to the next, would simply be accelerated to the collector with no resultant gain. In order both to focus and to direct the beam from emitter plate to emitter plate, crossed electric and magnetic fields are employed.

The magnetic multiplier illustrated in Fig. 3-18 employs both electric and magnetic fields. With the magnetic field perpendicular to the plane of the paper, the electrons move from emitter to emitter in practically cycloidal paths, as shown. However, if an electron starts from rest at the cathode it will have zero velocity when it reaches the first emitter. Under these circumstances, the electrons from the cathode could cause no secondary emission at this emitter. For this reason, an additional potential gradient must exist from the cathode to the first emitter and from the first to the second emitter, etc. The addition of this field distorts the original field, making an exact determination of the paths of the particles very difficult. Further, the effect of the initial velocities is to cause a slight defocusing of the beam in passing from one emitter to the next. Owing to these conditions, plus the fact that space charge also begins to play a part, a practical limitation is imposed upon the number of emitters that may be used and so upon the subsequent gain of the unit.

Because of the need for a magnetic field as well as an electric field, a great deal of attention has been given to the development of electrostatic secondary-emission multipliers. These differ from the magnetic type only in the shape and arrangement of the electrodes. The shapes and the orientations of the electrodes are such that the electrons pass progressively from one electrode to the next. The development of these units has been guided to a marked extent by the rubber-model method (see Sec. 2-17). Two different types of electrostatic multipliers are illustrated in Fig. 3-19. The RCA type 931-A tube with nine emitters (called *dynodes*) has a cur-
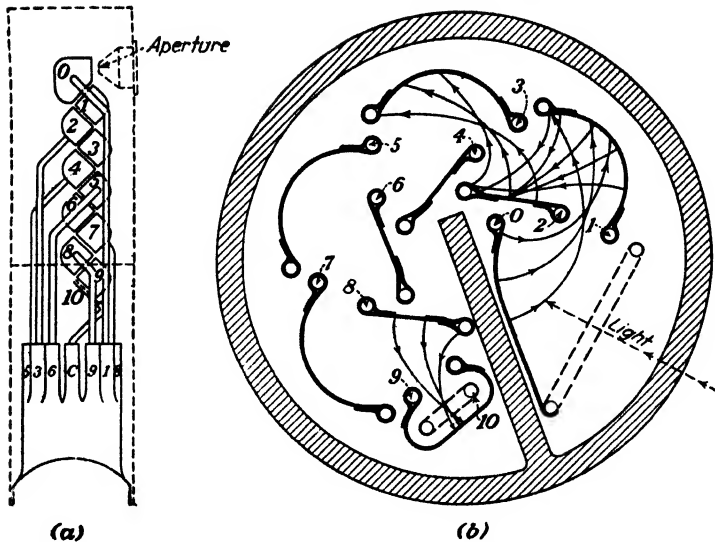
Fig. 3-19. (a) A linear electrostatic secondary-emission multiplier. (*Courtesy of Farnsworth Television and Radio Corp.*) (b) A circular electrostatic secondary-emission multiplier. (*Courtesy of Radio Corporation of America.*)

rent amplification of 200,000 and an output current of 2 amp/lumen. This tube is about the size of a small radio receiving tube.

**3-16. The Magnetron.** Consider a long, straight filament of radius $r_k$ and a coaxial cylindrical plate of radius $r_a$ which is maintained at a positive potential $E_b$ with respect to the cathode. Assume that the electrons that leave the filament do so with zero initial velocities. A magnetic field parallel to the axis of the filament is superposed upon the electric field. This is the so-called *magnetron* arrangement, first introduced by Hull.[24] This longitudinal magnetic field may be obtained by placing a long solenoid directly over the tube, the strength of the magnetic field being varied by controlling the current through the solenoid.
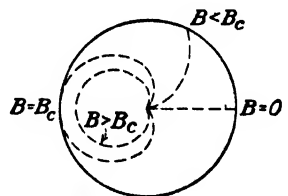


Fig. 3-20. The electronic path in a magnetron for several values of magnetic field.

The electronic paths will now be investigated qualitatively. Suppose the magnetic field is reduced to zero. The electron starts from rest at the cathode and is accelerated radially outward toward the plate by the electric field. If a weak magnetic field is present, then an application of the direction rule shows that the electronic path will be curved, although the electron will strike the plate. These conditions are illustrated in Fig. 3-20.

As the field is further increased, the path becomes more and more curved until a critical value $B_c$ is reached, when the electronic path becomes tangent to the plate. If the field is increased beyond this value, the electron does not strike the plate at all but instead returns to the filament.

A quantitative analysis of this problem leads to the following expression involving the critical field:

$$E_b = \frac{B_c^2 e r_a^2}{8m} \left(1 - \frac{r_k^2}{r_a^2}\right)^2 \tag{3-17}$$

Assume that a fixed potential $E_b$ is applied between the cathode and the anode and that the plate current is read as a function of the applied magnetic field. If $B$ is less than the critical value $B_c$ given by Eq. (3-17), the plate current should be unaffected, whereas for values of $B$ greater than
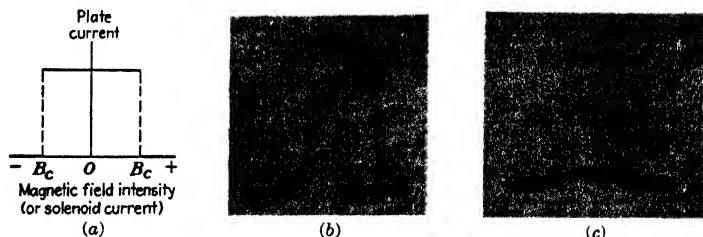


Fig. 3-21. Magnetron characteristics. (a) Ideal plate-current curve. (b) Oscillogram of plate current for two values of plate voltage. (c) Oscillogram of plate current for three values of screen voltage and a fixed plate voltage.

$B_c$ the current should suddenly drop to zero, as illustrated by the curve of Fig. 3-21a. Since a reversal of the direction of the magnetic field simply reverses the direction of travel of the electrons about the tube axis, the same current should be obtained for a given positive value of $B$ as for an equal negative value of $B$. This is observed in Fig. 3-21.

Actually, the oscillogram indicates more nearly what may be expected experimentally. Instead of a sharp cutoff, the current variation will be that shown in Fig. 3-21b. This curve results mainly from the fact that the commercial tubes do not possess the ideal geometrical arrangement assumed in the mathematical discussion; that is, the cathode may not be coaxial with the anode. Furthermore, the plate usually is not a perfect cylinder. Also, the theory assumes an infinitely long emitter and collector, a condition never satisfied in a commercial tube, although it can be approximated by using special tubes with "guard rings."

It has been found experimentally that the types 41 and 42 tubes yield reasonably satisfactory results. These tubes are pentodes.* In the fore-

* A *pentode* is a thermionic tube containing five electrodes—a cathode, a plate, and three grids, one of which is known as the *screen grid*.

going development no mention was made of any electrodes other than the cathode and anode. However, it is to be noted that in the final expression for the cutoff only the total potential difference $E_b$ between the cathode and anode appears. Since the field intensity is not present in this equation, *the cutoff value is independent of the potential distribution within the tube.* Direct experimental proof of this statement is given in the oscillogram (Fig. 3-21c). This oscillogram represents the plate current vs. magnetic-field intensity, with screen voltage as the parameter. It is seen that the cutoff value is approximately independent of the screen-grid potential. These curves justify the use of a multielectrode tube instead of a simple diode to illustrate the theory herein developed.
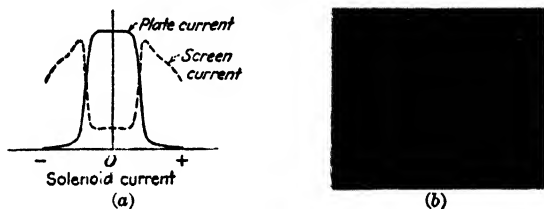
Fig. 3-22. Sketch and oscillogram of plate and screen current vs. solenoid current in a pentode used as a magnetron.

Direct experimental proof that the magnetic field causes the electrons originally directed toward the plate to be returned toward the cathode can be easily found with a multielectrode tube. Thus, if a screen is present near the plate, then, as the electrons miss the plate, they return toward the screen and may be collected by this electrode. Therefore, if a meter is inserted in the screen circuit, it will indicate an increasing current $I_s$ when the plate meter $I_b$ shows a decreasing value. However, the total space current $I_b + I_s$ will decrease since some of the electrons may pass through the screen openings and so will not contribute to the screen current. As the magnetic field is increased still further, the screen current will stop rising and will then start to decrease toward zero. The explanation is, of course, that with $B$ sufficiently large the electrons miss not only the plate but also the screen. This result is illustrated in Fig. 3-22.

If the tube contains an appreciable amount of gas, the repelling action of the electron cloud may be offset by the production of positive ions. When the electrons just miss the plate, they are traveling in their longest paths, so that the probability of collision with and consequent ionization of the gas molecules is greatest. This means that positive ions are produced which will partly neutralize the existing space charge. This explains why, in certain tubes, the plate current just before cutoff may actually be higher than its zero-field value. If the quantity of gas present in the tube

is too large, sufficient ionization may take place so that the tube behaves as a gas-filled tube. The foregoing magnetron theory will no longer be valid under these circumstances.

Another consequence of gas in the tube may be an extremely long tail to the $I_b$ versus $B$ curve. The reason is that some of the electrons that collide elastically with the gas molecules will suffer deflections in the direction of the plate and a stronger magnetic field will be required to prevent them from reaching the anode. A factor, thus far neglected, that also contributes to the tail of the curve is the effect of the initial velocities of the electrons. The electrons emitted with appreciable velocities in the radial direction will require a stronger magnetic field for deflection.

Another effect that contributes to the rounded shoulders of the curves in Fig. 3-21 is the following: For the values of $B$ slightly less than $B_c$ the electrons are moving in their longest paths. These electrons form an effective space charge that repels some of the electrons that would otherwise leave the cathode. In other words, the current from the cathode decreases with an increase in the magnetic field even though none of the electrons is missing the plate. This space-charge effect was not taken into account in the foregoing theory, since only the motion of an isolated electron was considered.

Very little practical use has been made of the static magnetron discussed here. However, it has been found that, if the anode is formed into a series of resonant cavities, the tube will perform efficiently as a very high power oscillator of centimeter waves (microwaves).[25] One most important application of such tubes is in radar systems.

**3-17. The Klystron.**[10, 26] All electrons leaving an electron gun have approximately the same velocity, that corresponding to the anode voltage. If they then pass through a region known as the *buncher* in which the potential varies with time, they will emerge from this second region with a speed that will depend upon the energy that they have acquired. Each electron will have a slightly different velocity from the preceding one, and the beam is said to be *velocity modulated.*

If the modulated beam is allowed to drift in a field-free space, then it is possible for a fast-moving electron to catch up with a slower moving one that left the modulating region at some earlier time. The density of the electron stream will then no longer be uniform with distance along the drift space, and the beam is said to be *density modulated.* It is possible for this modulation to be intense enough for the electrons to form in bunches along the drift space. This process is called *bunching.*

These principles are made use of in the klystron, a microwave oscillator. The general characteristics of the modulating portion of such a tube are illustrated in Fig. 3-23. A quantitative discussion of the processes mentioned above follows.

The effect of a superimposed a-c buncher potential on the d-c gun voltage is to give the electrons a speed $v$ given by

$$\tfrac{1}{2}mv^2 = e(E_a + E_b \sin \omega t) \qquad (3\text{-}18)$$

where $\omega$ is the angular frequency of the a-c buncher potential. It is assumed that the buncher grids are so close together that the electrons pass
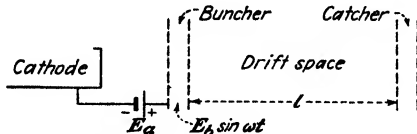


FIG. 3-23. A schematic of a klystron, showing the electron gun, the buncher, the drift space, and the catcher.

through this region in a time very short compared with one cycle of the modulating voltage. The resulting velocity of the particle is

$$v = \sqrt{\frac{2eE_a}{m}} \sqrt{1 + \alpha \sin \omega t} \qquad (3\text{-}19)$$

where $\alpha$ is the ratio $E_b/E_a$. It is possible in principle for $\alpha$ to be greater than unity, although velocity-modulation tubes ordinarily operate with $\alpha$ much less than 1.

Subject to the limitations that $\alpha$ is small, the expression under the second square-root sign may be expanded and only the lower order terms retained. Since, by the binomial expansion,

$$(1 + x)^{\tfrac{1}{2}} \doteq 1 + \tfrac{1}{2}x \qquad \text{for } x \ll 1$$

this leads to

$$v = \sqrt{\frac{2eE_a}{m}} \left(1 + \frac{\alpha}{2} \sin \omega t\right) \qquad (3\text{-}20)$$

The velocity of the electron at the buncher entrance is

$$v_0 \equiv \sqrt{\frac{2eE_a}{m}} \qquad (3\text{-}21)$$

and the expression for the velocity beyond the buncher is

$$v = v_0 \left(1 + \frac{\alpha}{2} \sin \omega t\right) \qquad (3\text{-}22)$$

This shows clearly that the electrons are velocity modulated at the frequency of the buncher.

At a distance $l$ from the buncher along the drift tube there is a second pair of grids, called the *catcher*. The time of arrival $t_2$ at the catcher of an electron that passed the buncher at the time $t_1$ is

$$t_2 = t_1 + \frac{l}{v} \tag{3-23}$$

This becomes, by virtue of Eq. (3-22),

$$t_2 = t_1 + \frac{l}{v_0 \left(1 + \dfrac{\alpha}{2} \sin \omega t_1\right)}$$

Upon expanding the denominator there results, to the same approximation as before,

$$t_2 = t_1 + \left(\frac{l}{v_0}\right)\left(1 - \frac{\alpha}{2} \sin \omega t_1\right) \tag{3-24}$$

It is convenient to multiply this equation by $\omega$, and then all transit times are converted into transit angles. Thus

$$\omega t_2 = \omega t_1 + \frac{\omega l}{v_0} - \frac{\omega \alpha l}{2v_0} \sin \omega t_1 \qquad \text{rad} \tag{3-25}$$

Introducing

$$\tau_1 \equiv \omega t_1 \qquad \tau_2 \equiv \omega t_2 \qquad \tau_0 \equiv \frac{\omega l}{v_0} \qquad \text{rad} \tag{3-26}$$

where $\tau_1$ is called the *departure angle*, $\tau_2$ the *arrival angle*, and $\tau_0$ the *transit angle of an electron passing through the buncher when the modulating voltage is zero*, and

$$k \equiv \frac{\omega \alpha l}{2v_0} = \frac{\omega E_b l}{2E_a v_0} \equiv \text{bunching parameter} \tag{3-27}$$

the above equation takes the form

$$\tau_2 = \tau_1 + \tau_0 - k \sin \tau_1 \qquad \text{rad} \tag{3-28}$$

This is a most important equation in klystron theory. Its significance can be seen by plotting $\tau_2$ as a function of $\tau_1$ for various values of the parameter $k$. In Fig. 3-24a, $k = 0$, which means that there is no modulation. It is seen that a group of electrons arriving in the angular interval $d\tau_2$ came from an angular interval of equal size, namely, $d\tau_1 = d\tau_2$. This means that the catcher current equals the buncher current and there is no bunching, an obviously correct conclusion since the modulation is zero.

In Fig. 3-24b, $k = 0.5$, the same size arrival interval $d\tau_2$ is taken, but it is seen that the interval from which the electrons departed depends upon

the time $\tau_1$. In the vicinity of the zero reference angle (region $A$) electrons are collected from a longer interval, $d\tau_1 > d\tau_2$ (as a matter of fact, $d\tau_1 = 2 d\tau_2$). This means that the catcher current is greater than the buncher current, or that a *bunch* is being formed. Physically, this comes about because the faster electrons are catching up with the slower ones which left at an earlier time. In region $B$, however, $d\tau_1 < d\tau_2$, which means a smaller catcher current than buncher current. It should be clear that a plot of catcher current vs. time has a peak (greater than the electron-
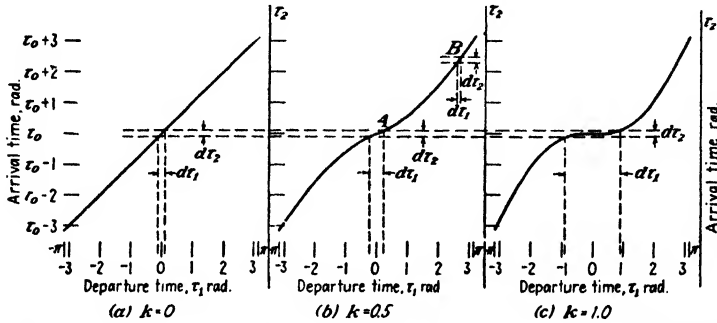


Fig. 3-24. Arrival time as a function of departure time (in radians). This illustrates the formation of a bunch.

gun current) near the zero reference time and tapers off on both sides to a value below the beam current as shown in Fig. 3-25.

In Fig. 3-24c, $k = 1.0$, the same size arrival interval $d\tau_2$ is again taken. It is now seen that, near the origin, the electrons are collected from a very large interval $d\tau_1$, that is $d\tau_1 \gg d\tau_2$. This represents a very large catcher current. As a matter of fact, if the interval $d\tau_2$ were taken very, very small (approaching zero), electrons would still be collected from a finite interval $d\tau_1$ or a finite number of electrons would be collected in an infinitesimal time. Since current is charge per unit time, this corresponds to an infinite current as indicated in Fig. 3-25.

The bunching parameter $k$ may be greater than unity. An analysis of this situation shows that electrons collected in a time $d\tau_2$ may have left the buncher in *several* discrete time intervals and that the current may contain more than one infinite peak.

The above considerations can be made quantitative as follows: If $I_1$ is the buncher current, then $I_1 dt_1$ is the charge leaving the buncher in the interval $dt_1$. If $I_2$ is the catcher current, then $I_2 dt_2$ is the charge entering the catcher in the interval $dt_2$. However, all electrons departing in the interval $dt_1$ arrive in the time $dt_2$, and hence, since there must be conservation of charge,

$$I_1 dt_1 = I_2 dt_2$$

or

$$I_2 = I_1 \frac{dt_1}{dt_2} \qquad (3\text{-}29)$$

The buncher current is constant and equal to the d-c beam current $I_0$. Hence

$$I_2 = I_0 \frac{d\tau_1}{d\tau_2}$$

From Eq. (3-28) this becomes

$$\frac{I_2}{I_0} = \frac{1}{1 - k \cos \tau_1} \qquad (3\text{-}30)$$

This equation clearly shows the infinite value of $I_2$ at $\tau_1 = 0$ for $k = 1$. The value of $I_2$ as a function of $\tau_1$ has no great significance, but $I_2$ versus
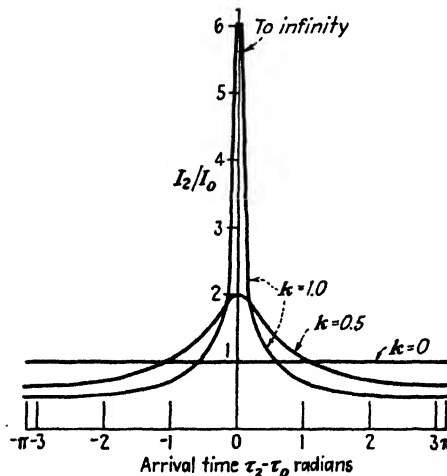


Fig. 3-25. Catcher current as a function of arrival time.

$\tau_2$ is desired. This cannot be obtained analytically since it requires solving the transcendental equation (3-28) for $\tau_1$ in terms of $\tau_2$. The graphs of Fig. 3-25 are obtained by assuming a value of $\tau_1$ and then solving for $\tau_2$ from Eq. (3-28) and for $I_2/I_0$ from Eq. (3-30).

A diagram of a klystron amplifier is shown in Fig. 3-26 (imagine the feedback line missing). The input is applied to the buncher resonator through a coaxial cable. A cavity resonator is the high-frequency (centi-meter-wave) equivalent of a parallel-resonant $LC$ circuit. The electrons are velocity modulated in passing through the input grids and form bunches

in the drift space.  The catcher grids are located at the drift distance cor-
responding to maximum bunching, and these bunches force the catcher
resonator into oscillations, giving an amplified output.  The electrons
after passing through the catcher grids hit the collector (shown with cool-
ing fins in Fig. 3-26) and are returned to the cathode.  The phase of the
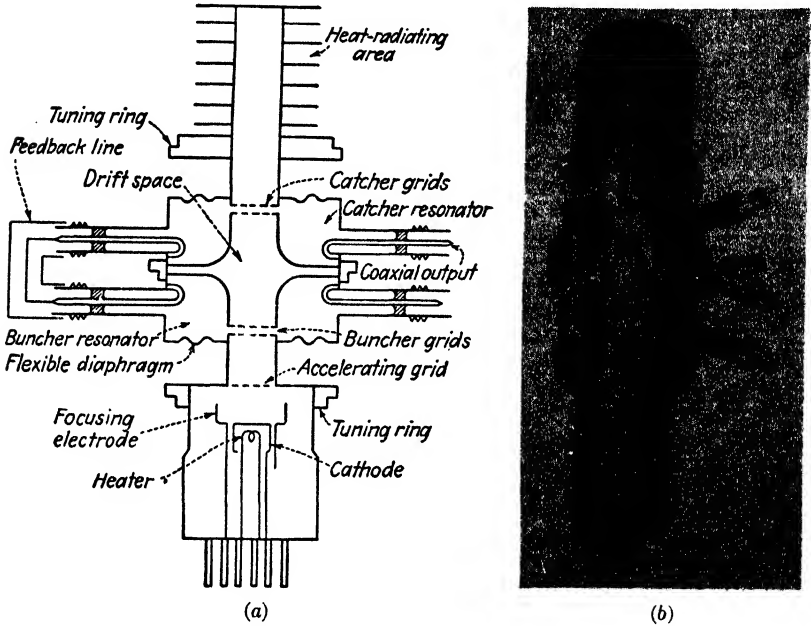


FIG. 3-26.  Double-resonator klystron.  (*From A. B. Bronwell and R. E. Beam, "Theory
and Application of Microwaves," McGraw-Hill Book Company, Inc., New York.*)

catcher voltage must be such as to slow down the electrons in order to
extract energy from the beam.

The principal use of the klystron is as an oscillator of centimeter waves.
For this application there is no r-f input, but the input and output leads
of Fig. 3-26 are connected together through a feedback line as shown.

It is possible to use one resonator as both buncher and catcher.  In
Fig. 3-26 the collector is replaced by an electrode, called the *reflector*, held
at a negative voltage with respect to the gun voltage.  The electrons are
bunched as in the two-cavity tube but are repelled by the reflector elec-
trode and hence returned to the resonator.  If the bunch arrives here when
the phase is such as to slow down these electrons, enough energy may be
extracted so as to maintain oscillations.  This is a most useful tube and
is known as a *reflex klystron*.

## PROBLEMS

**3-1.** The electrons that are emitted from the thermionic cathode of a cathode-ray-tube gun are accelerated by a potential of 400 volts. The essential dimensions of the tube are

$$L = 19.4 \text{ cm} \qquad l = 1.27 \text{ cm} \qquad d = 0.475 \text{ cm}$$

a. Compare the electrostatic sensitivity of this tube obtained from the theoretical expression with the experimental value of 0.89 mm/volt.

b. What must be the magnitude of a transverse magnetic field acting over the whole length of the tube in order to produce the same deflection as that produced by a deflecting potential of 30 volts? The distance from the anode to the screen is 23.9 cm.

c. Repeat part b for the case where the transverse magnetic field exists only in the region between the deflecting plates instead of over the entire length of the tube.

**3-2.** The cathode-ray tube of the previous problem has a transverse magnetic field of 1 milliweber/m² acting in the region between the plates and into the plane of the paper.

a. If there is no electrostatic field acting, what will be the deflection on the screen caused by the magnetic field?

b. Describe the motion and draw a sketch of the path of an electron from filament to screen, assuming that the electron leaves the filament with zero initial velocity.

c. What is the change in energy of the electrons caused by the magnetic field?

d. A voltage of what magnitude and sign must be applied to the plates in order that the electrons shall move down the tube undeflected?

**3-3.** A cathode-ray tube has the following dimensions:

Length of plates, 2.0 cm.
Separation of plates, 1.0 cm.
Distance from electron gun to center of plates, 5.0 cm.
Distance from center of plates to the screen, 20.0 cm.

Assume that there is only one set of plates in the tube. The accelerating voltage is 1,000 volts, and the beam, leaving the gun, is well focused. An a-c voltage applied to the plates produces a straight line 4.0 cm in length on the screen, if no magnetic field is present.

A uniform axial magnetic field is now applied over the entire length of the cathode-ray tube.

a. Neglecting the effect of this magnetic field in the region between the plates, calculate the minimum magnetic field that will reduce the line to a point on the screen.

b. If the magnetic field is reduced to half the value found in part a, a line is observed on the screen. Why? Calculate the length of this line and the angle that this line makes with the direction of the 4.0-cm line that was observed for zero magnetic field.
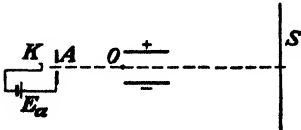


PROB. 3-4.

**3-4.** Consider the cathode-ray tube shown. A d-c potential is applied to the plates of this cathode-ray tube. In addition, a solenoid is placed over the tube giving a uniform magnetic field parallel to the axis of the tube. Describe in words the exact motion of an electron starting at rest at the cathode $K$ in the following sections of the tube:

a. Between cathode $K$ and anode $A$. Assume that the field is uniform in this region.

b. Between anode $A$ and the edge of the plates $O$.

c. In the region between the plates.

d. In the region beyond the plates.

**3-5.** In Prob. 3-4 the separation of the plates is 0.5 cm, the length of the plates is 2.0 cm, and the distance from the center of the plates to the screen is 15 cm. $E_a = 400$ volts, $E_d = 100$ volts, and $B = 10$ milliwebers/m². The electron leaves the cathode with zero initial velocity.

*a.* Find the coordinates of the electron just as it emerges from the plates.

*b.* Find the coordinates of the electron when it strikes the screen.

**3-6.** The accelerating voltage of a cathode-ray tube is 1,000 volts. A sinusoidal voltage is applied to a set of deflecting plates. The axial length of the plates is 2 cm.

*a.* What is the maximum frequency of this voltage if the electrons are not to remain in the region between the plates for more than one-half cycle?

*b.* For what fraction of a cycle does the electron remain in the region between the plates if the frequency is 60 cycles?

**3-7.** The electric field in the region between the plates of a cathode-ray tube is produced by the application of a deflecting potential given by

$$E_d = 60 \sin 2\pi \times 10^8 t$$

The important tube dimensions are

$$L = 19.4 \text{ cm} \qquad l = 1.27 \text{ cm} \qquad d = 0.475 \text{ cm}$$

The accelerating voltage is 200 volts. Where will an electron strike the screen if it enters the region between the plates at an instant when the phase of the deflecting voltage is zero?

**3-8.** *a.* A sinusoidal voltage of frequency $\omega$ is applied to the deflecting plates of a cathode-ray tube. The transit time between the plates is $\tau$. The length of the line on the screen is $A$. If $A_0$ is the line length when the transit time is negligible compared with the period of the applied voltage, show that

$$A = A_0 \frac{\sin (\omega\tau/2)}{\omega\tau/2}$$

*b.* If $E_a = 1,000$ volts and $l = 1$ cm, at what frequency will $A/A_0 = 0.9$?

**3-9.** Consider in detail the movement of an electron through a cathode-ray tube when electric fields are applied to both pairs of deflecting plates. Show that the vector displacement on the screen is approximately proportional to the vector sum of the electric-field intensities.

**3-10.** Prove that, if the accelerating voltage $E_a$ is high enough so that relativistic corrections must be made, the deflection of an electron in a cathode-ray tube will be

$$D = D_0 \frac{1 + \Psi}{1 + \frac{1}{2}\Psi}$$

where $D_0$ is the nonrelativistic deflection, and

$$\Psi = \frac{eE_a}{m_0c^2} = \frac{v_N^2}{2c^2}$$

where the symbols have the meanings given in the text. Show that, if $\Psi \ll 1$, then $D$ reduces to

$$D = D_0(1 + \frac{1}{2}\Psi)$$

How much error is made if the nonrelativistic instead of the correct formula is used for an 80,000-volt cathode-ray tube?
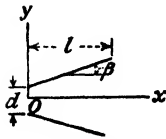
3-11. A cathode-ray tube is equipped with a pair of plates that are inclined at an angle $2\beta$ with each other as shown in the diagram. An electron is accelerated by the anode voltage $E_a$ so that it enters the region between the plates with a velocity $v_{0x}$. Neglect the effects of fringing, and assume that the lines of flux are vertical.

a. Prove that the vertical component of velocity is given by

$$v_y = \frac{eE_d}{2mv_{0x}\tan\beta} \ln\left(1 + \frac{2x\tan\beta}{d}\right)$$

PROB. 3-11.

where $E_d$ is the deflecting voltage.

HINT: Make use of the relationship

$$\frac{dv_y}{dt} = \frac{dv_y}{dx}\frac{dx}{dt} = v_x\frac{dv_y}{dx}$$

b. What is the slope ($\tan\theta$) of the path at the point where the electron leaves the region of the plates?

c. The distance from the center of the plates to the screen is $L$. If $L \gg l$, then little error will be made in assuming that the deflection on the screen is $D = L\tan\theta$. Under these conditions show that

$$D = \frac{E_d L}{4E_a \tan\beta} \ln\left(1 + \frac{2l}{d}\tan\beta\right)$$

3-12. Consider two cathode-ray tubes, one containing parallel deflecting plates and the other inclined plates as shown in the diagram. If the tubes are identical in all other respects prove that

a. The maximum possible deflection on the screen is the same for both tubes.

b. The sensitivity of the tube with the inclined plates is 38 per cent greater than that of the tube with the parallel plates.
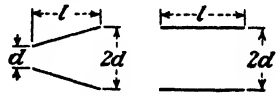
PROB. 3-12.

NOTE: Use the result of the previous problem.

3-13. A cathode-ray tube that is provided with a postaccelerating anode has the following dimensions:

$$l = 1.6 \text{ cm} \qquad L = 18.2 \text{ cm} \qquad d = 0.50 \text{ cm}$$

The distance from the center of the plates to the ring $R$ (Fig. 3-4) is 4.0 cm. The intensifier is 1.6 cm from the screen.

a. When used as an ordinary cathode-ray tube, the postaccelerating anode is connected to the accelerating anode. Calculate the deflection on the screen under these circumstances if $E_d = 50$ volts and $E_a = 1,200$ volts.

b. If one-half the accelerating voltage is applied after deflection between the anode and the intensifier electrode, what is the new deflection? See Fig. 3-4 where $E_{a1} = E_{a2} = 600$ volts. Assume that the intensifier field is axial and exists only in the region between $R$ and $I$.

3-14. The following experiment is performed on a cathode-ray tube that has all its plates brought out separately: An alternating voltage is applied to the ($Y$) plates nearest the gun. The other two ($X$) plates are tied together. Describe what happens to the vertical line as the potential between the $X$ set of plates and the anode is varied positively and negatively.

**3-15.** Linear sweep circuits are applied simultaneously to the horizontal and vertical plates of a cathode-ray tube. One plate of each set is tied to the second anode. What pattern will appear on the screen if

*a.* The frequency applied to the vertical plates is five times the frequency applied to the horizontal plates?

*b.* The frequency applied to the horizontal plates is five times the frequency applied to the vertical plates?

**3-16.** A mixture of $Li^6$ and $Li^7$ singly ionized atoms are produced in an ion source. These ions are accelerated by a potential difference of 1,000 volts between the source and an exit probe. The ions pass through a hole in the probe into a uniform transverse magnetic field of 1,000 gauss. A photographic plate is placed normal to the direction of the ions at the point where they enter the region of the magnetic field (see Fig. 3-13). What is the separation of the two lines on the photographic plate?

NOTE: The superscripts give the atomic weights of the isotopes.

**3-17.** A mixture of $K^{39}$, $K^{40}$, and $K^{41}$ singly ionized atoms are produced in an ion source. These ions are accelerated by a potential difference of 1,500 volts between the source and an exit probe. The ions pass through a hole in the probe into a uniform transverse magnetic field. If the $K^{39}$ line formed on a photographic plate that is oriented perpendicular to the original direction of the ions is 38.20 cm from the source, calculate the separation of the three lines on the photographic plate (see Fig. 3-13).

NOTE: The superscripts give the atomic weights of the isotopes.

**3-18.** A helium leak-detector mass spectrometer accelerates the ions through 500 volts and then bends them in a circle of radius 4.0 cm. What is the value of the magnetic field used in the instrument?

**3-19.** The position of a certain strong $\beta$-line of a sample of radium B which is placed at the origin of a magnetic spectrograph corresponds to a radius of curvature of 16.0 cm when the magnetic-field intensity is 120.5 gauss.

*a.* What would the speed be if the mass of the electrons did not change with velocity?

*b.* What is the true speed of the particles?

*c.* What is the ratio $m/m_0$?

(See M. M. Rogers, A. W. McReynolds, and F. T. Rogers, Jr., *Phys. Rev.*, **57**, 379, 1940.)

**3-20.** Given uniform electric and magnetic fields parallel to one another. Ions enter the region where these fields exist with velocities at right angles to the fields. *For small values of magnetic field*, prove that a photographic plate placed normal to the initial direction of the ions will, on development, show a series of parabolas, each parabola indicating a different value of the ratio $e/m$ of the ion. Show that each point on a given parabola corresponds to a different value of initial velocity. This is the principle of the positive-ray spectrograph.

**3-21.** The magnetic-field strength is 0.9 weber/$m^2$ in a certain cyclotron. Light hydrogen ions (protons) are used.

*a.* What must be the frequency of the oscillator supplying the power to the dees?

*b.* If each passage of the ions across the accelerating gap increases the energy of the ion by 60,000 volts, how long does it take for the ion introduced at the center of the spiral to emerge at the rim of the dee with an energy of 6 Mev?

*c.* Calculate the radius of the last semicircle before emergence.

**3-22.** Protons are accelerated in the MIT cyclotron. The magnetic-field strength is 1.3 webers/$m^2$, and the radius of the last semicircle is 0.5 m.

*a.* What must be the frequency of the oscillator supplying the power to the dees?

*b.* What is the final energy acquired by the proton?

*c.* If the total transit time of the proton is 3.3 $\mu$sec, how much energy is imparted to the particle in each passage from one dee to the other?

**3-23.** Explain what would happen if protons were used instead of electrons in the General Electric Co. 100-Mev betatron. In particular, calculate

  a. The maximum energy that the proton would acquire.

  b. The average energy acquired per trip around the doughnut.

  c. The number of revolutions that the proton would make.

**3-24.** Prove that the 1:2 stability condition for the betatron is satisfied at any radius if the magnetic-field intensity decreases as $1/r$ from the origin.

**3-25.** A betatron at the University of Illinois operates with a magnetic-field strength of 0.12 weber/m$^2$ at the orbit and accelerates electrons to a final energy of 2.2 Mev. The average energy acquired per revolution is 10 ev.

  a. What is the ratio $m/m_0$?

  b. What is the orbital radius?

  c. At what rate does the magnetic field vary?

  d. What is the average transit time?

  e. What is the total path length?

**3-26.** Calculate the number of stages required in a secondary-emission multiplier to give an amplification of $10^6$ if the secondary-emission ratio is 3.5.

**3-27.** In a nine-stage secondary-emission phototube multiplier, the incident photo-current is $10^{-8}$ amp and the output current from the multiplier is 0.1 amp. What is the secondary-emission ratio of the target material?

**3-28.** In the secondary-emission multiplier of Fig. 3-18, the distance between a target and its plate is 1.0 cm. The potential between these two elements is 100 volts. Assume that there is no field between targets and that the electrons leave each target with zero velocity so that the resultant motion is truly cycloidal.

  a. Find the minimum magnetic field required in order that this tube operate properly.

  b. If the tube were designed to operate with a field of 5 milliwebers/m$^2$, what would be the distance between centers of adjacent targets? Assume that the path remains cycloidal.

**3-29.** A cylindrical diode consists of a long, straight filament of radius $r_k$ and a concentric anode of radius $r_a$ between which is applied a potential $E_b$. Owing to the filament current $I$, a magnetic field surrounds the cathode, the lines of flux being circular and concentric with the filament.

If the electron starts from rest at the filament, describe its motion qualitatively. Show that there is a critical cutoff filament current above which no plate current exists, for a given plate voltage.

**3-30.** Sinusoidal voltages of the same amplitude and frequency are applied to both the horizontal and the vertical deflecting plates of a cathode-ray tube. If the phase angle between these two voltages is $\theta$, plot the pattern that will appear on the screen, if

  a. $\theta = 0$.

  b. $\theta = \pi/4$.

  c. $\theta = \pi/2$.

  d. $\theta = \pi$.

This problem illustrates the use of the cathode-ray tube as a phase-measuring device.

**3-31.** A voltage $E_x = A \sin \omega_1 t$ is applied to the horizontal plates, and a voltage $E_y = A \sin (\omega_2 t + \theta)$ is applied to the vertical plates of a cathode-ray tube. Plot the pattern that will be observed on the screen, if

  a. $\omega_2/\omega_1 = 2$, $\theta \neq 0$.

  b. $\omega_2/\omega_1 = 2$, $\theta = \pi/2$.

  c. $\omega_2/\omega_1 = 3$, $\theta = 0$.

  d. $\omega_2/\omega_1 = 3$, $\theta = \pi/2$.

  e. $\omega_2/\omega_1 = \frac{3}{2}$, $\theta = 0$.

  f. $\omega_2/\omega_1 = \frac{3}{2}$, $\theta = \pi/2$.

These patterns are called Lissajous figures.  This problem illustrates the use of the cathode-ray tube as an instrument for comparing an unknown frequency with a given standard.  (See *RCA Pamphlet* TS-2, "Cathode-ray Tubes and Allied Types," 1935.)

## REFERENCES

1. MACGREGOR-MORRIS, J. T., and V. A. HUGHES, *IEEJ*, **79**, 454, 1936.
2. BENHAM, W., *Wireless Eng.*, **13**, 10, 1936.
3. *Du Mont Lab. Comm.* 3, **19**, 33, 1939.
   BIGALKE, A., *Z. tech. Physik*, **19**, 163, 1938.
   ROGOWSKI, W., and H. THIELEN, *Arch. Elektrotech.*, **33**, 411, 1939.
4. LEMPERT, I. E., and R. FELDT, *IRE Proc.*, **34**, 432, 1946.
5. NUTALL, A. K., *IEEJ*, **78**, 229, 1936.
   MILLER, J. L., and J. E. L. ROBINSON, *ibid.*, **74**, 511, 1934.
6. MALOFF, I. G., and D. W. EPSTEIN, "Electron Optics in Television," McGraw-Hill Book Company, Inc., New York, 1938.
   ZWORYKIN, V. K., and G. A. MORTON, "Television," John Wiley & Sons, Inc., New York, 1940.
7. COSELETT, V. E., "Introduction to Electron Optics," Oxford University Press, New York, 1946.
8. LIEBMANN, G., *Electronic Eng.*, **18**, 289, 1946.
9. SHARPE, J., *ibid.*, **18**, 385, 1946.
   BOWIE, R. M., *IRE Proc.*, 1948.
10. SPANGENBERG, K. R., "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.
11. DU MONT, A. B., *Electronics*, **8**, 16, January, 1935.
    FLEMING-WILLIAMS, B. C., *Wireless Eng.*, **17**, 61, 1940.
    MILLMAN, J., and S. SEELY, "Electronics," 1st ed., McGraw-Hill Book Company, Inc., New York, 1941.
12. ASTON, F. W., "Mass Spectra and Isotopes," Longmans, Green & Co., Inc., New York, 1933.
    DEMPSTER, A. J., *Am. Phil. Soc. Proc.*, **75**, 762, 1935.
    BAINBRIDGE, K. T., and E. B. JORDON, *Phys. Rev.*, **50**, 282, 1936.
    SAMPSON, M. B., and W. BLEAKNEY, *ibid.*, **50**, 456, 1936.
13. HIPPLE, J. A., D. J. GROVE, and W. M. HICKMAN, *Trans. AIEE*, **64**, 141, 1945.
14. WORCESTER, W. G., and E. G. DOUGHTY, *ibid.*, **65**, 946, 1946.
    THOMAS, H. A., T. W. WILLIAMS, and J. A. HIPPLE, *Rev. Sci. Instruments*, **17**, 368, 1946.
    NIER, A. O., C. M. STEVENS, A. HUSTRULID, and T. A. ABBOTT, *J. Applied Phys.*, **18**, 30, 1947.
15. LAWRENCE, E. O., and M. S. LIVINGSTON, *Phys. Rev.*, **40**, 19, 1932; **45**, 608, 1934.
    LIVINGOOD, J. J., *Electronics*, **8**, 421, November, 1935.
    MANN, W. B., "The Cyclotron," Chemical Publishing Company, Inc., Brooklyn, 1940.
    WILSON, R. R., *J. Applied Phys.*, **11**, 781, 1940.
    LIVINGSTON, M. S., *ibid.*, **15**, 2, 129, 1944.
    LIVINGSTON, M. S., article in "Advances in Electronics," p. 269, Academic Press, New York, 1948.
16. BROBECK, W. M., E. O. LAWRENCE, K. R. MACKENZIE, E. M. MCMILLAN, R. SERBER, D. C. SEWELL, K. M. SIMPSON, and R. L. THORNTON, *Phys. Rev.*, **71**, 449, 1947.

17. ELDER, F. R., A. M. GUREWITSCH, R. V. LANGMUIR, and H. C. POLLACK, *J. Applied Phys.*, **18**, 810, 1947.
18. LIVINGSTON, M. S., J. P. BLEWETT, G. K. GREEN, and L. J. HAWORTH, *Rev. Sci. Instruments*, **21**, 7, 1950.
19. KERST, D. W., *Phys. Rev.*, **60**, 47, 1941.
    WANG, T. J., *Electronics*, **18**, 128, 1945.
20. WESTENDORP, W. F., and E. E. CHARLTON, *J. Applied Phys.*, **16**, 581, 1945.
21. KERST, D. W., and R. SERBER, *Phys. Rev.*, **60**, 53, 1941.
    WANG, *op. cit.*
22. SCHIFF, L. I., *Rev. Sci. Instruments*, **17**, 6, 1946.
23. ZWORYKIN, V. K., G. A. MORTON, and L. MALTER, *IRE Proc.*, **24**, 351, 1936.
24. HULL, A. W., *Phys. Rev.*, **18**, 31, 1921.
25. FISK, J. B., H. D. HAGSTRUM, and P. L. HARTMAN, *Bell System Tech. J.*, **25**, 167, 1946.
    COLLINS, G. B., "Microwave Magnetrons," Radiation Laboratory Series, Vol. 6, McGraw-Hill Book Company, Inc., New York, 1948.
26. HAMILTON, D. R., J. K. KNIPP, and J. B. H. KUPER, "Klystrons and Microwave Triodes," Radiation Laboratory Series, Vol. 7, McGraw-Hill Book Company, Inc., New York, 1948.
    BRONWELL, A. B., and R. E. BEAM, "Theory and Application of Microwaves," McGraw-Hill Book Company, Inc., New York, 1947.

### General References

"The Cathode-Ray Tube and Typical Applications," Allen B. Du Mont Laboratories, Clifton, N.J., 1948.
SOLLER, J. T., M. A. STARR, and G. E. VALLEY, JR.: "Cathode Ray Tube Displays," Radiation Laboratory Series, Vol. 22, McGraw-Hill Book Company, Inc., New York, 1948.
ZWORYKIN, V. K., G. A. MORTON, E. G. RAMBERG, J. HILLIER, and A. W. VANCE: "Electron Optics and the Electron Microscope," John Wiley & Sons, Inc., New York, 1945.

# CHAPTER 4

# ELECTRONIC PHENOMENA IN METALS

THE behavior of electrons in various configurations of electric and magnetic fields has been considered in some detail in Chaps. 2 and 3. The origin of these electrons did not enter the discussion. However, the important known electronic sources were outlined briefly in Chap. 1. Of these, matter in the metallic state is the most important, and the characteristics of this source will be considered in some detail.

**4-1. Introduction.** Before treating the intriguing but perhaps somewhat involved quantitative aspects of this topic, it may be well to summarize briefly what may be learned from such a study. A physical "picture" of the nature of the region inside a metal will be presented, which should assist us to visualize the types of electronic behavior that are possible in this region. It will be found that the classical notion of a metal as a simple equipotential volume will have to be modified considerably. Instead, the picture will present rather intense local variations of potential. Electrons of low energy will not be able to move about to any great extent within the metal. The more energetic electrons will be able to surmount potential-energy variations within the metal; but when they reach the surface, they find here a potential-energy barrier that they cannot surmount under ordinary conditions. These electrons are therefore confined to the interior of the metal, where they may wander about more or less freely, although they will never leave this region unless additional energy is given to them. This model will be found useful in studying the quantitative aspects of electronic emission: thermionic, photoelectric, and others.

Since the electrons may flow out of the metal only when an added energy has been given to them, the question which must first be considered is how much energy the electrons possess under normal conditions when they are not acted upon by outside agents. This will lead to the concept and discussion of *distribution functions*. For example, the energy distribution specifies the number of electrons that have large, medium, or small energies. More precisely, it will define the density of electrons in various energy ranges. From this distribution in energy there can be deduced a velocity distribution function that will give information with regard to the electrons traveling within specified velocity ranges. It is with the aid of such distribution functions that "body" phenomena, such as the con-

ductivity and specific heat of metals, can be most intelligently discussed.

From the electronic point of view the surface phenomena are even more important than the conditions within the metal. Many puzzling questions immediately suggest themselves: How many electrons per second will escape through the surface of a metal that is maintained at a given temperature? What will be the distribution in energy and velocity of the escaping electrons? How will the emission be affected by accelerating or retarding fields at the surface of the metal? What will happen to the potential-energy barriers at the surfaces of two metals when they are joined together, perhaps as the component parts of an electronic device? All of these and related questions will be considered. The fundamental ideas and some of the most important practical results will be given in this chapter, and the more theoretical concepts will be considered in the next chapter.

**4-2. Free Electrons in Metals.** X-ray and other studies reveal that most metals are crystalline in structure. This means that they consist of



a space array of atoms or molecules (strictly speaking, ions) built up by regular repetition in three dimensions of some fundamental structural unit. Under these conditions, the atoms of the elements are so close together that the outer electrons of the atom are as much associated with one atom as with another, so that the electron attachment to any individual atom is practically zero. Depending upon the metal, at least one, sometimes two, and in a few cases three electrons per atom are free to move throughout the interior of the metal under the action of applied forces.

Figure 4-1 shows the charge distribution within a metal, specifically, sodium. The plus signs represent the heavy positive sodium nuclei of the individual atoms. The heavily shaded regions represent the electrons in the sodium atom that are tightly bound to the nucleus. These are inappreciably disturbed as the atoms come together to form the metal. The light shading represents the outer, or valence, electrons in the atom; and it is these electrons that cannot be said to belong to any particular atom.

Fig. 4-1. Arrangement of the sodium atoms in the 110 plane of the metal. (*W. Shockley, J. Applied Phys.*, **10**, 543, 1939.)

Instead, they have completely lost their individuality and can wander freely about from atom to atom in the metal. Thus a metal is visualized as a region containing a periodic three-dimensional array of heavy, tightly bound ions permeated with a swarm of electrons that may move about quite freely. This is known as the "electron-gas" concept of a metal.

**4–3. The Energy Method of Analyzing the Motion of a Particle.** This method employs the principle of the conservation of energy, use being made of the potential-energy curve corresponding to the field of force. The principles involved may best be understood by considering specific examples of the method.

*Example.* An idealized diode consists of plane-parallel electrodes, 5 cm apart. The anode $A$ is maintained 10 volts negative with respect to the cathode $K$. An electron leaves the cathode with an initial energy of 2 ev. What is the maximum distance it can travel from the cathode?

*Solution.* This problem will be analyzed by the energy method. Figure 4-2a is a linear plot of potential vs. distance, and in Fig. 4-2b is indicated the corresponding poten-
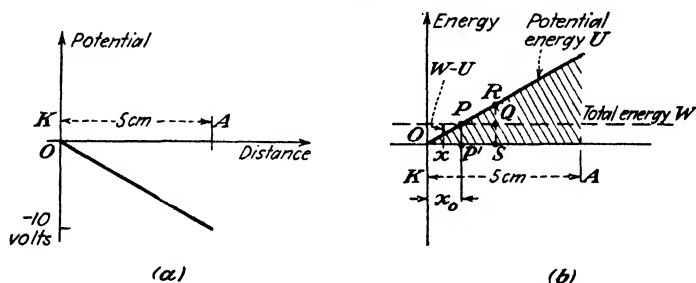


FIG. 4-2. Illustrating the potential-energy barrier encountered by an electron in a retarding field.

tial energy vs. distance. Since potential is the potential energy per unit charge (see Sec. 2-4), curve $b$ is obtained from curve $a$ by multiplying each ordinate by the charge on the electron (a negative number). Since the total energy $W$ of the electron remains constant, it is represented as a horizontal line. The kinetic energy at any distance $x$ equals the difference between the total energy $W$ and the potential energy $U$ at this point. This difference is greatest at $O$, indicating that the kinetic energy is a maximum when the electron leaves the cathode. At the point $P$ this difference is zero. This means that no kinetic energy exists, so that the particle is at rest at this point. This distance, $x_0$, is the maximum that the electron can travel from the cathode. At point $P$ it comes momentarily to rest and then reverses its motion and returns to the cathode. From geometry it is seen that $x_0/5 = \frac{2}{10}$ or $x_0 = 1$ cm.

Consider a point such as $S$ which is at a greater distance than 1 cm from the cathode. Here the total energy $QS$ is less than the potential energy $RS$, so that the difference, which represents the kinetic energy, is negative. This is an impossible physical condition, however, since negative kinetic energy ($\frac{1}{2}mv^2 < 0$) implies an imaginary velocity. This can be interpreted to mean that the particle can never advance a distance greater than $OP'$ from the cathode.

The foregoing analysis leads to the very important conclusion that the shaded portion

of Fig. 4-2 can never be penetrated by the electron. Thus, at point $P$ the particle acts *as if* it had collided with a solid wall, hill, or barrier and the direction of its flight had been altered. *Potential barriers* of this sort will play important roles in the analyses to follow.

It must be emphasized that the words "collides with" or "rebounds from" a potential "hill" are convenient descriptive phrases and that an actual encounter between two material bodies is not implied.

As a second illustration, consider a mathematical pendulum of length $l$, consisting of a "point" bob of mass $m$ that is free to swing in the earth's gravitational field. If the lowest point of the swing (point $O$, Fig. 4-3) is chosen as the origin, then the potential energy of the mass at any point $P$ corresponding to any angle $\theta$ of the swing is given by

$$U = mgy = mgl(1 - \cos \theta) \qquad (4\text{-}1)$$

where $g$ is the acceleration of gravity. This potential-energy function is illustrated graphically in Fig. 4-4.
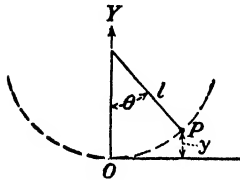
Fig. 4-3. Point $P$ represents the mass $m$ of a mathematical pendulum swinging in the earth's gravitational fields.

Consider the resultant motion of the bob if it is given a potential energy $U_1$ by raising it through an angle $\theta_0$ and releasing it with zero initial velocity. If dissipation is neglected, the particle will swing back and forth through the angle $2\theta_0$, going from $\theta_0$ on one side to $\theta_0$ on the other side of the vertical axis. How might one proceed to analyze the motion of the physical system if only the potential-energy field of Fig. 4-4 were given without specifying the physical character of the system?

The procedure is the same as that followed in the simple diode problem considered above. A horizontal line *aebc* is drawn at a height equal to the total energy $W_1$ of the particle. At any point, such as $e$, the total energy is represented by $eg = W_1$, and the potential energy is represented by $fg$. The difference between these two, namely, $ef$, represents the kinetic energy of the particle when the angle of swing, given by the intercept of $eg$ on the axis, corresponds to $Og$. In other words, the difference between the total-energy line and the potential-energy curve at any angle represents the kinetic energy of the particle under these conditions. This difference is greatest at $O$, indicating that the kinetic energy is a maximum at the bottom of the swing, an almost evident result. At the points $a$ and $b$ this difference is zero. This means that no kinetic energy exists, or that the particle is at rest at these points. This is readily evident, since corresponding to the points $a(\theta = \theta_0)$ and $b(\theta = -\theta_0)$, the particle is about to reverse its motion.

Consider a point in the shaded region outside the range $-\theta_0$ to $+\theta_0$, such as $h$. Here the total energy $ch$ is less than the potential energy $dh$. This impossible condition is interpreted by our previous reasoning to

mean that the particle whose total energy is $W_1$ can never swing to the angle $Oh$ corresponding to the point $c$, so that the motion must be confined to the region $ab$. The shaded portions of Fig. 4-4 represent the potential barrier which can never be penetrated by the bob, if its total energy is no greater than $W_1$. This type of constrained motion about a point $O$ is closely analogous to that of the so-called "bound" electrons in a metal, as will be seen later.

Now consider the case when the bob has a total energy equal to $W_2$, which is greater than the maximum of the potential-energy curve. Clearly from Fig. 4-4 the horizontal line corresponding to this energy cannot intersect the curve at any point. Consequently, the particle does not "collide" with the potential barrier, and its course is never altered, so that it moves through an ever-increasing angle. Of course, its kinetic energy varies over wide limits, being maximum for $\theta = 0$, $2\pi$, $4\pi$, ... and minimum for $\theta = \pi$, $3\pi$, $5\pi$, .... Physically, this type of motion results when the bob
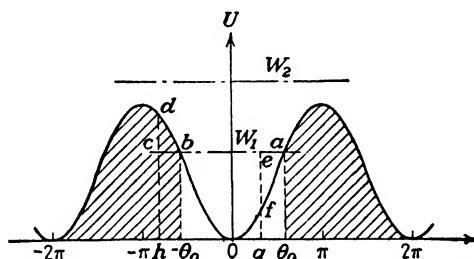


FIG. 4-4. The potential energy of the bob in Fig. 4-3 plotted as a function of the angle of swing.

has enough energy to set it spinning completely around in a circular path. This is somewhat analogous to the type of motion experienced by the so-called "free" electrons in a metal.

A method has already been considered in some detail in Chap. 2 by which the motion of charged particles in electric and magnetic fields may be analyzed. It consists of the solution of Newton's second law of motion in which the forces of electric and magnetic origin are equated to the product of the mass and the acceleration of the particle. Obviously, this method is not applicable when the forces are as complicated as they must be in a metal. Futhermore, it is neither possible nor desirable to consider what happens to each individual electron.

It was necessary, therefore, to consider the alternative method illustrated above. This simple but powerful method facilitates the discussion of the motion of a particle in a conservative field of force, such as that found in the body of a metal. It will also be applied to many other types of problem. For example, the method of analysis just considered

is extremely useful in determining whether electrons will possess sufficient energy to pass through grids and reach the various electrodes in a vacuum tube and whether or not electrons or ions will be able to penetrate electron clouds in a vacuum tube or ion sheaths in a gaseous-discharge tube. This method will now be applied to the analysis of the motion of electrons in metals.

**4-4. The Potential-energy Field in a Metal.** It is desired to set up the potential-energy field for the three-dimensional array of atoms that exists in the interior of a metal and to discuss the motion of electrons in this field. The resultant potential energy at any point in the metal is simply the sum of the potential energies produced at this point by all of the ions of the lattice. To determine the potential energy due to one ion, it is noted that an atom of atomic number $Z$ has a net positive charge $Ze$ on its nucleus. Surrounding this nucleus is an approximately spherical cloud, or shell, of $Z$ electrons. By Gauss's law the potential at a point at a distance $r$ from the nucleus, when $r$ is smaller than the inside diameter of this electron shell, is that resulting from the charge on the nucleus. In the mks rationalized system of units this potential is given by [1]

$$V = \frac{Ze}{4\pi\epsilon_0 r} + C_1 \tag{4-2}$$

where $C_1$ is a constant and $\epsilon_0$ is the permittivity of free space (see Appendix IV).

Since the potential $V$ equals the potential energy $U$ per unit charge (see Sec. 2-4), then $U = -eV$. The minus sign is introduced because the charge on the electron is negative. Hence

$$U = -\frac{Ze^2}{4\pi\epsilon_0 r} - eC_1 \tag{4-3}$$

This specifies the form of the variation of the potential energy in the immediate neighborhood of the nucleus.

A similar expression exists for the potential at a distance from the nucleus that is so large that the electron under consideration is completely outside of the cloud of the remaining $Z - 1$ electrons. The net charge of the nucleus and the $Z - 1$ electrons is simply $Ze - (Z - 1)e = +e$. The potential is

$$V = \frac{e}{4\pi\epsilon_0 r} + C_2$$

where $C_2$ is a constant, and the corresponding potential energy is

$$U = \frac{-e^2}{4\pi\epsilon_0 r} - eC_2 \tag{4-4}$$

The potential of any point may be chosen as the zero reference of potential because it is only differences of potential that have any physical significance. For the present discussion it is convenient to choose zero potential at infinity and then $C_2 = 0$. Thus, $U$ equals $-Ze^2/4\pi\epsilon_0 r +$ constant in the neighborhood of the nucleus and equals $-e^2/4\pi\epsilon_0 r$ at large distances from the nucleus. The determination of the exact variation for intermediate distances is unnecessary for the present discussion. Enough has been said, however, to make plausible the potential-energy curve illustrated in Fig. 4-5. Here $\alpha$ represents a nucleus, the potential energy of which is represented by the curves $\alpha_1\alpha_2$. The vertical scale represents
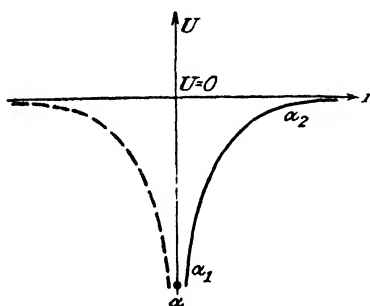


FIG. 4-5. The potential energy as a function of radial distance from an isolated nucleus.
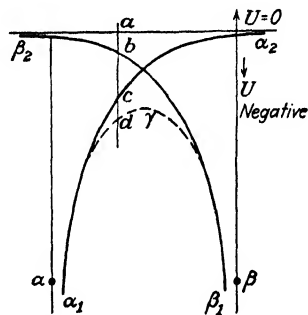
FIG. 4-6. The potential energy resulting from two nuclei $\alpha$ and $\beta$.

$U$, and the horizontal scale gives the distance $r$ from the nucleus. It must be emphasized that $r$ represents a radial distance from the nucleus and hence can be taken in any direction. If the direction is horizontal but to the left of the nucleus, then the dashed curve represents the potential energy.

To represent the potential energy at every point in space requires a four-dimensional picture, three dimensions for the three space coordinates and a fourth for the potential-energy axis. This difficulty is avoided by plotting $U$ along some chosen line through the crystal, say through a row of ions. From this graph and the method by which it is constructed it is easy to visualize what the potential energy at any other point might be. In order to build up this picture, consider first two adjacent ions, and neglect all others. The construction is shown in Fig. 4-6. $\alpha_1\alpha_2$ is the $U$ curve for nucleus $\alpha$, and $\beta_1\beta_2$ is the corresponding $U$ curve for the adjacent nucleus $\beta$. If these were the only nuclei present in the metal, the resultant $U$ curve in the region between $\alpha$ and $\beta$ would be the sum of these two curves, as shown by the dotted curve $\alpha_1\gamma\beta_1$ (since $ad = ab + ac$). It is seen that the resultant curve is very nearly the same as the original curves

in the immediate vicinity of the nuclei, but it is lower and flatter than either individual curve in the region between the nuclei.

Let us now single out an entire row of nuclei $\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$, ... from the metallic lattice (see Figs. 4-1 and 4-7) and sketch the potential energy as we proceed along this line from one nucleus to the other, until the surface of the metal is reached. Following the same type of construction as above, but considering the influence of other near-by nuclei, an energy distribution somewhat as illustrated in Fig. 4-7 is obtained.

According to classical electrostatics, which does not take the atomic structure into account, the interior of a metal is an equipotential region.
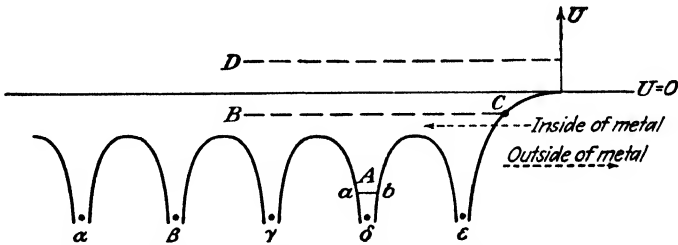


FIG. 4-7. The potential-energy distribution within and at the surface of a metal.

The present, more accurate, picture shows that the potential energy varies appreciably in the immediate neighborhoods of the nuclei and actually tends to $-\infty$ in these regions. These are the intense local variations referred to in Sec. 4-1. However, the potential is approximately constant for the greatest volume of the metal, as indicated by the slowly varying portions of the diagram.

Consider the conditions that exist near the surface of the metal. It is evident, according to the present point of view, that the exact position of the "surface" cannot be defined. It is located at a small distance from the last nucleus $\epsilon$ in the row. It is to be noted that, since no nuclei exist to the right of $\epsilon$, there can be no lowering and flattening of the potential-energy curve such as prevails in the region between the nuclei. This leads to a most important conclusion, *viz.: a potential energy "hill," or "barrier," exists at the surface of the metal*

**4-5. Bound and Free Electrons.** The motion of an electron in the potential energy field of Fig. 4-7 will now be discussed by the method given in Sec. 4-3. Consider an electron in the metal that possesses a total energy corresponding to the point $A$ in Fig. 4-7. This electron collides with, and rebounds from, the potential walls at $a$ and $b$. It cannot drift very far from the nucleus but can move about only in the neighborhood $ab$ of the nucleus. Obviously this electron is strongly bound to the nucleus and so is called a *bound electron*. It is evident that these bound electrons con-

tribute very little to the conductivity of the metal since they cannot wander about in the metal, even under the stimulus of an externally applied electric field. These are the electrons that are responsible for the heavy shading in the neighborhood of the nuclei of Fig. 4-1.

Our present interest is in the *free*, or *conduction*, electrons in the metal rather than in the bound ones. A free electron is one having an energy corresponding to the point $B$ of the figure. At no point *within* the metal is its total energy entirely converted into potential energy. Hence, at no point is its velocity zero, and the electron travels more or less freely throughout the body of the metal. However, when the electron reaches the surface of the metal, it collides with the potential-energy barrier

there. At the point $C$, its kinetic energy is reduced to zero, and the electron is turned back into the body of the metal. An electron having an energy corresponding to the point $D$ collides with no potential walls, not even the one at the surface, and so it is capable of leaving the metal.

In our subsequent discussions the bound electrons will be neglected completely since they in no way contribute to the phenomena to be studied. Attention will be focused on the free electrons. In so far as the free elec-
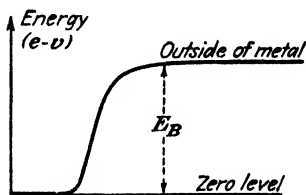
Fig. 4-8. For the free electrons, the interior of a metal may be considered an equipotential volume, but there is a potential barrier at the surface.

trons are concerned, the region in which they find themselves is essentially a potential plateau, or equipotential region. It is only for distances close to a metallic ion that there are any appreciable variations in potential. Since the regions of rapidly varying potential represent but a very small portion of the total volume of the metal, it will be supposed that the field distribution within the metal is equipotential and the free electrons are subject to no forces whatsoever. The present viewpoint is therefore essentially that of classical electrostatics.

Figure 4-7 is redrawn in Fig. 4-8, all potential * variations within the metal being omitted, with the exception of the potential barrier at the surface. The zero of energy is chosen at the level of the plateau of this diagram. This choice of the zero-energy reference level is valid since, as has already been emphasized, only difference of potential has physical significance. This means that for the present discussion the constant $C_2$

---

* This figure really represents potential energy and not potential. However, the phrase "potential barrier" is much more common in the literature than the phrase "potential-energy barrier." Where no confusion is likely to arise, these two expressions will be used interchangeably. These barriers will be measured in electron volts, and hence the symbol $E$ will replace the $U$ of the preceding sections. It must be emphasized that one unit of $E$ represents $1.60 \times 10^{-19}$ joule of energy.

in Eq. (4-4) is no longer chosen equal to zero but is given a value such as to make the potential at infinity equal to $E_B$, the height of the potential-energy barrier in electron volts.

**4-6. Energy Distribution of Electrons.** In order to be able to escape, an electron inside of the metal must possess an amount of energy at least as great as that represented by the surface barrier $E_B$. A sensible question is then: What energies are possessed by the electrons in a metal? The quantitative answer is given in the next chapter. It is found that there is a variation in energy among the electrons (they do not all have the same amount of energy) and that this relationship depends upon tem-
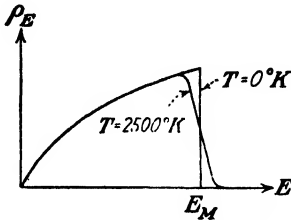
Fig. 4-9.  Energy distribution in metallic tungsten at 0° and 2500°K.
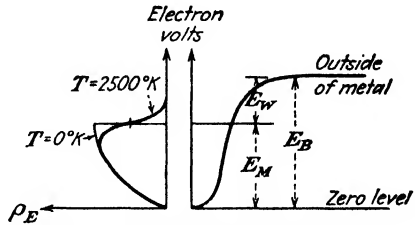
Fig. 4-10.  Energy diagram of free electrons.

perature. This is known as the *Fermi-Dirac-Sommerfeld* (FDS) *energy distribution function*. The form of the function is indicated in Fig. 4-9 for two temperatures, absolute zero and 2500°K. The ordinate $\rho_E$ represents the energy density, and so the number of electrons having energies between $E$ and $E + dE$ is $\rho_E \, dE$. It is seen that at 0°K there are very few low-energy electrons and that the density increases with energy up to a definite value $E_M$ and then abruptly drops to zero. Thus $E_M$ *represents the maximum energy an electron can have at absolute zero.* Classically all electrons should possess zero energy at 0°K. The distribution indicated in Fig. 4-9 can be understood only with the aid of quantum theory.

It should be noted that the distribution at a very high temperature (2500°K) differs not too radically from that at 0°K. The low-energy electrons are hardly disturbed at all. The high-energy particles are given still greater energies. This follows from the fact that the curve just to the left of $E_M$ is lowered (indicating that there are now fewer electrons having energies in this neighborhood), whereas the curve extends to the right of $E_M$ (indicating that some electrons now have energies greater than $E_M$). The curve approaches the axis asymptotically, indicating that a few electrons have very high energies.

**4-7. Work Function.** In Fig. 4-10, Fig. 4-9 has been rotated 90 deg counterclockwise and combined with Fig. 4-8 so that the vertical axis

represents energy for both sets of curves.[2]  At 0°K it is impossible for an
electron to escape from the metal because this requires an amount of
energy equal to $E_B$ and the maximum energy possessed by any electron is
only $E_M$.  It is necessary to supply an additional amount of energy equal
to the difference between $E_B$ and $E_M$ in order to make this escape possible.
This difference, written $E_W$, is known as the *work function* of the metal.
Thus

$$E_W \equiv E_B - E_M \qquad (4\text{-}5)$$

The term $E_B$ has been called the *outer work function* by Sommerfeld, and
the quantity $E_M$ is referred to as the *inner work function*, the *Fermi char-
acteristic energy*, or the *Fermi level*.

The experiments of Davisson and Germer [3] and of Rupp [4] on the dif-
fraction of electrons in passing through matter have verified the existence
of the potential-energy barrier at the surface of the metal.  In fact, based
on the results of these experiments together with
experimentally determined values of $E_W$, it is pos-
sible to calculate the values of $E_M$ for the metals
used.  These values of $E_M$ can be compared with
the theoretical values given in the next chapter.
The data given by Rupp show fair agree-
ment between the experimental and theoretical
values.



FIG. 4-11.  The force on
a charge $-e$ located at a
distance $x$ from the sur-
face of a metal is equal to
the force between $-e$ and
its image force $+e$ located
at the same distance $x$
within the metal.

A second physical meaning of the term work
function may be obtained by considering what
happens to an electron as it escapes from a metal,
without particular regard to the conditions within
the interior of the metal.  A negative electron
will induce a positive charge on a metal from
which it escapes.  There will then be a force of attraction between the
induced charge and the electron.  Unless the electron possesses sufficient
energy to carry it out of the region of influence of this image force of attrac-
tion, it will be returned to the metal.  This was the qualitative explanation
given in Sec. 1-3 to explain the fact that an electron did not escape from
a metal unless some external force was applied.  Some quantitative fea-
tures of this force will be investigated.  The method of electrical images in
classical electrostatics dictates that the force $f$ (newtons), acting on an
electron $-e$ (coulombs), at a distance $x$ (meters) from a plane metallic
conductor, is given by the attraction between the electron and its image
charge $+e$ located at a distance $x$ from the face of the conductor and sit-
uated within it, as illustrated in Fig. 4-11.  The force on the electron is
then

$$f = \frac{(+e)(-e)}{4\pi\epsilon_0(2x)^2} = \frac{-e^2}{16\pi\epsilon_0 x^2} \qquad \text{newtons} \qquad (4\text{-}6)$$

where $\epsilon_0$ is the permittivity of free space. The negative sign indicates that the force is one of attraction toward the surface of the metal.

Within the range in which the electron reacts to the metal as it would to a plane surface, the force varies according to Eq. (4-6). As the distance from the electron to the metal is decreased, the surface may no longer be considered as a plane surface and Eq. (4-6) is no longer valid. In fact, as the "surface" of the metal is approached, the force must change from the attractive force given by this equation to a zero force. This is so because
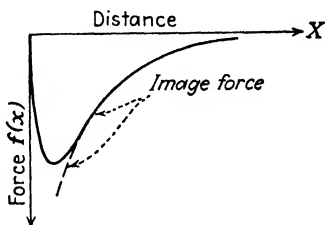
FIG. 4-12. The force on an electron escaping from a metal.
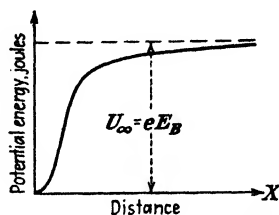
FIG. 4-13. The potential-energy curve corresponding to the force curve of Fig. 4-12.

the metal is an equipotential region and the average force on an electron inside the metal is zero.

From the microscopic point of view, it is extremely difficult to specify the position of the "surface," or for that matter to specify the regions "internal" or "external" to the metal. One can be no more specific than to say that the force on an electron which is, say, a few atomic diameters from the plane of ions nearest the exterior of the metal is given by Eq. (4-6). For distances of closer approach the force must change and reduce to zero. The force function is illustrated in Fig. 4-12. The exact nature of the force in the neighborhood of $x = 0$ depends upon the type of metal under consideration and upon the point of the surface at which the $X$ axis is erected (*i.e.*, upon how close to a surface ion the escaping electron comes). Fortunately, the exact shape of the force function is not necessary for the analysis that follows.

The potential-energy function that gives rise to the force $f$ is

$$U = -\int_0^x f(x)\, dx \qquad \text{joules} \qquad (4\text{-}7)$$

This expresses the work necessary to remove an electron from the surface of the metal to a distance $x$ from it. This potential energy $U$ (joules) is plotted as a function of $x$ in Fig. 4-13. Each ordinate of this curve is simply the negative of the area under the force curve of Fig. 4-12 from the point $O$ to the point $x$. The shape of this curve is the same as that of the

potential barrier of Fig. 4-8.  This must be so, since both of these curves represent the same surface characteristic.  It is only the method of development employed in obtaining the two curves that is different.

The total area under the curve of Fig. 4-12 represents the total amount of work that is necessary in order to remove an electron from the metal.  This is evidently equal to the height of the potential barrier.  Hence

$$U_\infty \equiv eE_B = -\int_0^\infty f(x) \, dx \qquad \text{joules} \qquad (4\text{-}8)$$

In view of the present considerations the position of the "surface" of a metal is taken as that distance where the extrapolated image force goes to minus infinity.

*Example.*  The experimental value of the work function $E_W$ of tungsten is 4.52 ev, and the theoretical value of the Fermi energy $E_M$ (see Sec. 5-3) is 8.95 ev.

*a.*  If an electron in metallic tungsten has 13.43 ev energy, how far from the surface can it wander?

*b.*  What is the image force when the electron is at its maximum distance from the metal?

*c.*  What electric field will cause the same force on the electron as that in part *b*?

*Solution.*  *a.*  The height of the barrier is

$$E_B = E_M + E_W = 8.95 + 4.52 = 13.47 \text{ ev}$$

After wandering a maximum distance $x_0$ from the surface, the electron collides with the potential-energy barrier and its kinetic energy is reduced to zero.  Hence the height of the barrier at this point is 13.43 ev.  The work required to remove this particle from $x_0$ to infinity is $13.47 - 13.43 = 0.04$ ev.  Assuming that the image force is valid for distances greater than $x_0$, then

$$0.04 \times 1.60 \times 10^{-19} = +\int_{x_0}^\infty \frac{e^2}{16\pi\epsilon_0} \frac{dx}{x^2} = \frac{e^2}{16\pi\epsilon_0 x_0} \qquad \text{joules}$$

Using $\epsilon_0 = 1/36\pi \times 10^9$ (Appendix IV), this yields $x_0 = 9.00 \times 10^{-9}$ m.

Since the interionic spacing is of the order of magnitude of an angstrom unit ($10^{-10}$ m), this is of the order of 90 ionic separations.  The image force undoubtedly is valid at such large distances from the surface.

*b.*  The image force is

$$f = \frac{e^2}{16\pi\epsilon_0 x_0^2} = \frac{(1.60 \times 10^{-19})^2}{16\pi \times (10^{-9}/36\pi)(9.00 \times 10^{-9})^2}$$

$$= 7.12 \times 10^{-13} \text{ newton}$$

*c.*  The electric field which will yield the same force is given by

$$e\mathcal{E} = 7.12 \times 10^{-13}$$

or

$$\mathcal{E} = \frac{7.12 \times 10^{-13}}{1.60 \times 10^{-19}} = 4.44 \times 10^6 \text{ volts/m}$$

The reason that the very small force in part *b* corresponds to the very high field intensity of part *c* is that the charge on the electron is so small.

**4-8. Thermionic Emission.** The curves of Fig. 4-10 show that the electrons in a metal at absolute zero are distributed among energies which range in value from zero to the maximum energy $E_M$. Since an electron must possess an amount of energy at least as great as $E_B$ in order to be able to escape, no electrons can leave the metal. Suppose now that the metal, in the form of a filament, is heated by sending a current through it. Thermal energy is then supplied to the electrons from the lattice of the heated metal crystal. The distribution of the electrons changes, owing to the increased temperature, as indicated in Fig. 4-10. If the temperature is raised sufficiently, some of the electrons represented by the tail of the curve of Fig. 4-10 will have energies greater than $E_B$ and so may be able to escape from the metal.

Using the analytical expression from the distribution functions, it is possible to calculate the number of electrons which strike the surface of the metal per second with sufficient energy to be able to surmount the surface barrier and hence escape. The derivation is given in Sec. 5-13, where it is shown that the thermionic current is given by

$$I_{th} = SA_0 T^2 \epsilon^{-E_W/E_T} \qquad \text{or} \qquad I_{th} = SA_0 T^2 \epsilon^{-b_0/T} \qquad \text{amp} \qquad (4\text{-}9)$$

where $S$ = area of filament, m$^2$

$A_0$ = constant whose dimensions are amp/(m$^2$)($°$K$^2$)

$T$ = temperature, $°$K

$E_T \equiv T/11{,}600$ is called the *electron-volt equivalent of temperature*

$E_W$ = work function, ev

$b_0 \equiv 11{,}600 E_W$, $°$K

Equations (4-9) are two forms of the equation of thermionic emission. They are sometimes referred to as the "Dushman equations" and sometimes as the "Richardson equations," since both workers developed equations of this form theoretically. The constant $E_W$, which has been termed the work function, is known by a variety of other names, the most common of which are the "effective" or "net work function"; the "electron affinity" of the metal; the "latent heat of evaporation of electrons" from the metal. The last term arose from the analogy of electron emission with the evaporation of molecules from a liquid, as already discussed in Sec. 1-3. A discussion of the experimental verification of this equation is given in Chap 6. Table 5-2 lists values of the thermionic-emission constants for some metals. The values for the most commonly used emitters (tungsten, thoriated-tungsten, and oxide-coated cathodes) are given in Sec. 6-4.

It must be emphasized that Eqs. (4-9) give the electron emission from a metal at a given temperature provided that there are no external fields present. If there are either accelerating or retarding fields at the surface, then the actual current collected will be greater or less than the emission current, respectively. The effect of such surface fields is discussed later.

**4-9. Contact Potential.** Consider two metals in contact with each other, as at the junction $C$ in Fig. 4-14. The contact difference of potential between these two metals is defined as the potential difference $E_{AB}$, between a point $A$ just outside metal 1 and a point $B$ just outside metal 2. The reason for the existence of the difference of potential is easily understood. When the two metals are joined at the boundary $C$, electrons will flow from the lower work-function metal, say 1, to the other metal, 2. This will continue until metal 2 has acquired so much negative charge that a retarding field has built up which repels any further electrons. It is shown in Sec. 5-17 that this equilibrium condition is attained when the
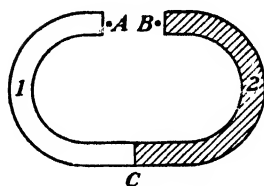


Fig. 4-14. Two metals in contact at the junction $C$.
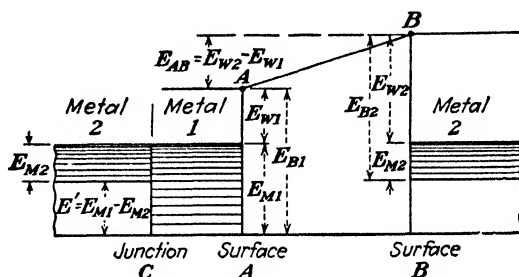
Fig. 4-15. The potential-energy system of two metals in contact.

Fermi energies $E_M$ of the two metals are located at the same height on the energy-level diagram. To satisfy this condition the potential-energy diagram for the two metals must be drawn as in Fig. 4-15. The barriers at the two surfaces $A$ and $B$ are indicated as vertical lines instead of curves as in Fig. 4-8 because the distance between the surfaces $A$ and $B$ is very large in comparison with atomic dimensions. The spacing of the horizontal lines is intended to indicate, in a crude way at least, the energy density. The progressively decreasing spacing between horizontal lines at the higher energies illustrates pictorially that the density of electrons is greatest in the neighborhood of the energy $E_M$.

The diagram should be clear if it is recalled that $E_W = E_B - E_M$. From this figure it is seen that

$$E_{AB} = E_{W2} - E_{W1} \qquad (4\text{-}10)$$

This means that *the contact difference of potential between two metals equals the difference between their work functions.* This result has been verified experimentally by numerous investigators.

If metals 1 and 2 are similar, the contact potential between them is evidently zero. If they are dissimilar metals, the metal having the lower work function becomes charged positively and the higher work-function

metal becomes charged negatively. In a vacuum tube the cathode is usually the lowest work-function metal. If it is connected to any other electrode externally by means of a wire, then the effective voltage between the two electrodes is not zero but equals the difference in the work functions. This potential difference is in such a direction as to *repel* the electrons being emitted from the cathode. If a battery is connected between the two electrodes, then the effective potential is the algebraic sum of the applied voltage and the contact potential. Further discussion of this topic is given in Sec. 5-17.

**4-10. Energies of Emitted Electrons.** Since the electrons inside a metal have a distribution of velocities, then those which escape from the metal will also have an energy distribution. It is easy to demonstrate this experimentally. Thus consider a plane emitter and a plane-parallel collector. The current is measured as a function of the retarding voltage $E_r$ (the emitter positive with respect to the collector). If all the electrons left the cathode with the same energy, then the current would remain constant until a definite voltage was reached and then it would fall abruptly to zero. For example, if they all had 2 ev energy, then when the retarding voltage was greater than 2 volts the electrons could not surmount the potential barrier between cathode and anode and no particles would be collected. Experimentally no such sudden falling off of current is found, but instead there is an exponential decrease of current with voltage according to the equation

$$I = I_{th}\epsilon^{-E_r/E_T} \tag{4-11}$$

This equation is derived in Sec. 5-18. It is based on the assumption that the effect of space charge (a matter considered in some detail in Chap. 7) may be neglected. Consequently, the total current that will be collected by the anode with zero retarding potential $E_r = 0$ is $I_{th}$, the thermionic current corresponding to the temperature $T$. With retarding fields, the current reduces to small values. If the potential is made accelerating, the collected current should remain constant at the saturation value. A plot of the term $\log_{10} I$ vs. $E_b$, the potential applied between the cathode and the collecting anode, should then be of the form shown in Fig. 4-16. The slope of the straight-line portion of this curve is $(11,600 \log_{10} \epsilon)/T = 5,030/T$.
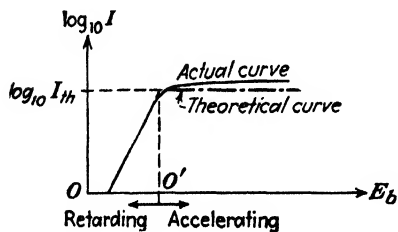


FIG. 4-16. To verify the retarding-potential equation, $\log_{10} I$ is plotted vs. $E_b$.

The point $O'$ is the true zero of potential, since from the foregoing considerations the thermionic saturation current will be collected at this

point. This means that the potential $OO'$ must represent the contact difference of potential $E_{AB}$ as indicated in Fig. 4-15. Metal 2 is the anode, and metal 1 is the cathode. Thus, with zero applied voltage a retarding potential actually exists which prevents the electrons emitted from metal 1 from getting to metal 2. A battery of magnitude $E_{AB} = E_{0'0}$ must be applied with the positive terminal at the collector in order to reduce the barrier between the two metals to zero so that an electron escaping from 1 can travel unimpeded to 2. Note that the retarding-potential method is an indirect way of measuring contact potentials.
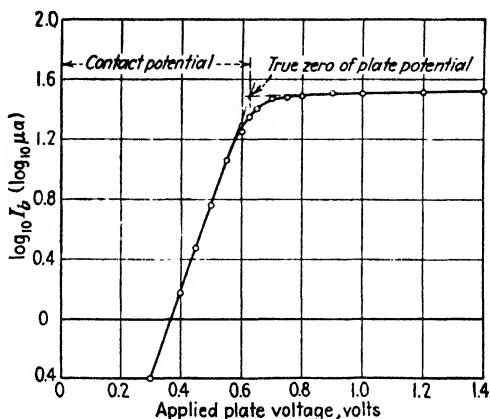


FIG. 4-17. Verification of the retarding-potential equation using a 12Z3 tube operating at a very low filament voltage. The temperature (calculated from the slope of the line) is $860°K$.

The "theoretical" curve of Fig. 4-16 represents an idealized condition that can never be realized practically, because some space charge is always present. Practically, the results do not show the sharp change in slope at the point $O'$, and the current rises continuously for increasing accelerating potentials. The greatest portion of the curve obtained for retarding potentials does give a straight-line variation, as shown. It is thus possible to determine the electron temperature with some accuracy from the straight-line portion of this curve. The value of the contact potential determined from this curve is only approximate.

A more serious shortcoming of this method of determining electron temperatures and contact potentials arises from the fact that it is extremely difficult to build a system which even approximates that of a plane cathode with a parallel collecting anode. However, certain of the available commercial tubes which possess cylindrical symmetry and in which the anode is but slightly larger than the cathode may be considered as fulfilling these ideal conditions with a fair approximation. The results obtained experi-

mentally using a 12Z3 diode having an indirectly heated oxide-coated cathode are shown in Fig. 4-17. In order to reduce the effects of space charge, the saturation current was set very low (about 20 $\mu$a).

*Example.* What percentage of the electrons leaving a tungsten filament at 2700°K have normal-to-the-surface-directed energies * in excess of 1 ev?

*Solution.* Those electrons whose surface-directed emission energies are in excess of 1 ev can surmount a 1-volt retarding potential. Using Eq. (4-11), with $E_r = 1$, and remembering that $E_T = T/11,600$ yields

$$\frac{I}{I_{th}} = \epsilon^{-(11,600 \times 1)/2,700} = \epsilon^{-4.28} \doteq 0.014$$

Hence, only about 1.4 per cent of the electrons have energies in excess of 1 ev.

If the emitter is an oxide-coated cathode operating at 1000°K, then a calculation similar to the above gives the result that only about 0.001 per cent of the electrons have a surface-directed energy in excess of 1 ev!

In Sec. 5-15 it is shown that the average energy of the escaping electrons is given by the expression

$$\bar{E} = 2E_T \qquad \text{ev} \tag{4-12}$$

For operating temperatures of 2700° and 1000°K the average energies of the emitted electrons are 0.47 and 0.17 ev, respectively.

These calculations demonstrate the validity of the assumption made in Chap. 2 in the discussion of the motion of electrons in electric and magnetic fields, *viz.*, that the electrons begin their motions with very small initial velocities. In most applications the initial velocities are of no consequence, but they are of significance in tubes which are operated at low electrode voltages.

**4-11. Accelerating Fields.** Under normal operating conditions, the field applied between the cathode and the collecting anode is accelerating rather than retarding, and so the field aids the electrons in overcoming the image force at the surface of the metal. This accelerating field tends, therefore, to lower the work function of the metal and so results in an increased thermionic emission from the metal. This matter is considered in detail in Sec. 5-19, where it is shown that the current $I$ under the condition of an accelerating field of $\varepsilon$ volts per meter at the surface of the emitter is

$$I = I_{th}\epsilon^{+0.440\varepsilon^{\frac{1}{2}}/T} \tag{4-13}$$

where $I_{th}$ is the zero-field thermionic current and $T$ is the cathode temperature in degrees Kelvin. The fact that the measured thermionic currents continue to increase as the applied potential between the cathode and the anode is increased is often referred to as the *Schottky effect*, after the man

---

* This phrase refers to the energy $E_N$ corresponding to the velocity component normal to the surface $v_N$. These are related by $eE_N = \frac{1}{2}mv_N^2$.

who first predicted this effect. Some idea of the order of magnitude of this increase can be obtained from the following illustration.

*Example.* Consider a cylindrical cathode of radius 0.01 cm and a coaxial cylindrical anode of radius 1.0 cm. The temperature of the cathode is 2500°K. If an accelerating potential of 500 volts is applied between the cathode and the anode, calculate the percentage increase in the zero-external-field thermionic-emission current because of the Schottky effect.

*Solution.* The electric-field intensity at any point $r$ (meters) in the region between the electrodes of a cylindrical capacitor, according to classical electrostatics, is given by the formula

$$\varepsilon = \frac{E_b}{\ln (r_a/r_k)} \frac{1}{r} \qquad \text{volts/m} \qquad (4\text{-}14)$$

where ln denotes the logarithm to the natural base $\epsilon$, $E_b$ is the plate voltage, $r_a$ is the anode radius, and $r_k$ is the cathode radius. Thus the electric-field intensity at the surface of the cathode is

$$\varepsilon = \frac{500}{2.303 \log_{10} 100} \frac{1}{10^{-4}} = 1.085 \times 10^6 \text{ volts/m}$$

It follows from Eq. (4-13) that

$$\log_{10} \frac{I}{I_{th}} = \frac{(0.434)(0.44)(1.085 \times 10^6)^{\frac{1}{2}}}{2,500} = 0.0795$$

Hence, $I/I_{th} = 1.20$, which shows that the Schottky theory predicts a 20 per cent increase over the zero-field emission current.

**4-12. High Field Emission.** It has been stated that the application of an accelerating field at the surface of a thermionic cathode results in an
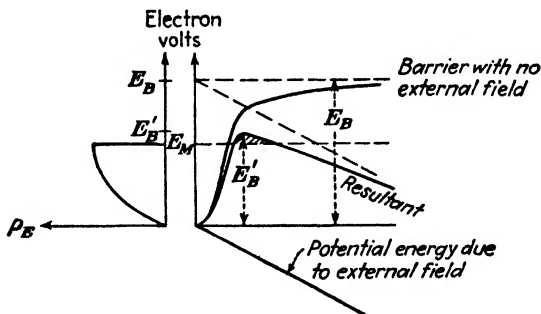


Fɪɢ. 4-18. The potential-energy curves with no field and with an intense electric field at the surface of a cold metal.

effective reduction of the height of the potential barrier at the surface. Suppose that the accelerating field at the surface of a "cold" cathode (one for which the thermionic-emission current is negligible) is very intense. The potential diagram under these conditions is shown in Fig. 4-18. Each

ordinate of the curve labeled "Resultant" is the algebraic sum of the corresponding ordinates of the potential-energy curves for zero field and for the external accelerating field. If it is assumed that the completely degenerate distribution function applies, then according to classical mechanics it will be impossible for any of the electrons in the metal to be emitted. This follows from the fact that there is a potential barrier at the surface whose magnitude $E_B'$ is greater than the maximum energy $E_M$ of the electrons.

According to the modern concept of the wave nature of the electron, however, it is possible for some electrons to penetrate *through* the potential barrier that exists at the surface of the metal and thereby escape, even though their energies may be less than the height of the barrier. The probability of emission under these conditions is very small and will depend upon the height, the width or depth, and the shape of the potential barrier through which the electrons must pass in order to escape. The variation of the emission-current density with the strength of the electric-field intensity at the surface of the metal has been calculated by several investigators.[5] The result obtained by Fowler and Nordheim is

$$J = C\mathcal{E}^2 \epsilon^{-D/\mathcal{E}} \qquad \text{amp/m}^2 \tag{4-15}$$

where

$$\left.\begin{array}{l} C = \dfrac{6.2 \times 10^{-6}}{E_B}\left(\dfrac{E_M}{E_W}\right)^{\frac{1}{2}} \qquad \text{amp/volt}^2 \\[12pt] D = 6.8 \times 10^9 E_W^{\frac{3}{2}} \qquad \text{volts/m} \end{array}\right\} \tag{4-16}$$

This equation has received direct experimental verification.[6] This effect is also called *cold-cathode emission* or *autoelectronic emission*.

Several thousand amperes have been obtained by the application of high field gradients to cold metallic electrodes in an X-ray tube used for high-speed radiography.[7]

**4-13. Secondary Emission.**[8] The number of secondary electrons that are emitted from a material, either a metal or a dielectric, when subjected to electron bombardment has been found experimentally to depend upon a number of factors. Among these are the number of primary electrons, the velocity of the primary electrons, the angle of incidence of the electrons on the material, the type of material, and the physical condition of the surface. The secondary-emission ratio, defined as the ratio of the number of secondary electrons per primary electron, is small for pure metals, the maximum value being between 1.5 and 2. It is increased markedly by the presence of a contaminating layer of gas or by the presence of an electropositive or alkali metal on the surface. For such composite surfaces, secondary-emission ratios as high as 10 or 15 have been detected. This

ratio as a function of the energy of the impinging primary electrons on a Cs-CsO-Ag surface used in a secondary-emission multiplier of the type described in Sec. 3-15 is shown in Fig. 4-19.

The exact mechanism of secondary emission, particularly from composite surfaces, is not clearly understood.[9] Qualitatively, the process is described as follows: The incident primary electron gives up its energy to several electrons within the matter. The depth of penetration of the primary electron depends upon its energy, the more energetic particles penetrating to greater depths. For example, Hastings[10] found that the depth of origin of secondary electrons from silver on platinum is about 15 atomic layers or less for primary electrons having energies of 20 ev. An appreciable emission occurs from a depth of 150 atomic layers when the incident electrons have energies greater than 50 ev. If the inner electrons do not give up all their energy by collision as they move toward the surface, they are able to overcome the barrier at the surface of the metal and hence can escape as secondary electrons.



FIG. 4-19. Ratio of secondary-emission current to primary current for a Cs-CsO-Ag surface as a function of the energy of the primary beam. (*V. K. Zworykin, G. A. Morton,* and *L. Malter, Proc. IRE,* **24,** 351, 1936.)

The maximum in the secondary-emission ratio curve (see Fig. 4-19) can be explained qualitatively. For low-energy primaries, the number of secondaries that are able to overcome the surface attraction is small. As the energy of the impinging electrons increases, more energetic secondaries are produced and the secondary-emission ratio increases. Since, however, the depth of penetration increases with the energy of the incident electron, the secondaries must travel a greater distance in the metal before they reach the surface. This increases the probability of collision in the metal, with a consequent loss of energy of these secondaries. Thus, if the primary energy is increased too much, the secondary-emission ratio must pass through a maximum.

Most secondary electrons are emitted with small energies. There is evidence[11] that more than 85 per cent of the secondary electrons emitted from a cesiated-silver surface have energies of less than 3 ev. This condition is to be expected since a rapidly moving inner electron should be able to induce the same type of phenomenon as a fast-moving primary electron. The small percentage of high-energy secondary electrons that is present in the energy spectrum of the emitted electrons is attributed to those primary electrons which have been scattered from the surface.

It is possible to induce electron emission by bombarding a surface with positive ions instead of with electrons.[12] This process is much less efficient than electron bombardment. As a result the energies of the impinging ions must be much greater than those of electrons in order to yield a comparable secondary-emission ratio. Nevertheless, this process plays a fundamental role in some types of discharge to be discussed later.

## PROBLEMS

**4-1.** A diode consists of a plane emitter and a plane-parallel anode separated by a distance of 0.5 cm. The anode is maintained at a potential of 10 volts negative with respect to the cathode.

*a.* If an electron leaves the emitter with a speed of $10^6$ m/sec and is directed toward the anode, at what distance from the cathode will it intersect the potential-energy barrier?

*b.* With what speed must the electron leave the emitter in order to be able to reach the anode?

**4-2.** A particle when displaced from its equilibrium position is subject to a linear restoring force $f = -kx$, where $x$ is the displacement measured from the equilibrium position. Show, by the energy method, that the particle will execute periodic vibrations with a maximum displacement which is proportional to the square root of the total energy of the particle.

**4-3.** A particle of mass $m$ kilograms is projected vertically upward in the earth's gravitational field with a speed $v_0$ meters per second.

*a.* Show by the energy method that this particle will reverse its direction at the height of $v_0^2/2g$ meters, where $g$ is the acceleration of gravity in meters per second per second.

*b.* Show that the point of reversal corresponds to a "collision" with the potential-energy barrier.

**4-4.** A triode consists of plane-parallel elements. The grid is located 0.2 cm, and the anode is 1.0 cm from the cathode. The grid is maintained at a potential of $-1.0$ volt and the plate at a potential of 100 volts with respect to the cathode. Assume that the potential varies linearly from the cathode to the grid and also linearly from the grid to the plate. Assume that the grid offers no mechanical hindrance to the flow of electrons.

*a.* If the electron leaving the cathode surface in the perpendicular direction collides with the potential-energy barrier after it has traveled a distance of 0.05 cm, with what energy was it emitted?

*b.* With what energy must it leave the emitter in order to be able to reach the anode?

The foregoing assumptions are not strictly valid in a practical triode.

**4-5.** *a.* If the cathode and plate of the previous problem are maintained at zero potential and if the potential of the grid is 4 volts (positive), will the electron collide with a potential-energy barrier at any point of its path, if its initial velocity is zero?

*b.* How long will it take the particle to reach the anode?

*c.* With what velocity will the electron strike the plate?

**4-6.** Consider the following model of an atom: The nucleus consists of a positive point charge $Ze$, where $Z$ is the atomic number and $e$ is the numerical value of the charge of the electron. This is surrounded by $Z$ electrons of which $Z - 1$ may be considered to be located on the surface of an imaginary sphere of radius $r_0$.

*a.* If the potential at infinity is taken as zero, show that the potential-energy function of the remaining (valence) electron is given by

$$4\pi\epsilon_0 U = -\frac{e^2}{r} \qquad \text{if } r > r_0$$

$$4\pi\epsilon_0 U = -\frac{Ze^2}{r} + (Z-1)\frac{e^2}{r_0} \qquad \text{if } r < r_0$$

In the equations above $r$ is expressed in meters, $e$ in coulombs, $U$ in joules, and $\epsilon_0$ is the permittivity of free space in the mks rationalized system.

*b.* Consider three such atoms in a row. The first is separated from the second by a distance of $4r_0$, and the second is separated from the third by the same amount. Assuming that sodium atoms ($Z = 11$) are under consideration, plot to scale the potential energy of the valence electron. Make the transformations

$$y = \frac{4\pi\epsilon_0 U r_0}{e^2} \qquad \text{and} \qquad x = \frac{r}{r_0}$$

and plot $y$ vs. $x$ instead of $U$ vs. $r$.

**4-7.** Consider seven sodium ions arranged in a plane lattice as indicated in the figure. Assuming the atomic model discussed in the preceding problem to be correct, plot the potential energy of the valence electron as a function of distance along row $ABC$.
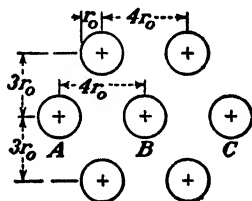
Compare the results of this problem with the preceding problem. Discuss qualitatively how the potential-energy function along the row $ABC$ will be altered as more ions are added to the figure in accordance with the lattice structure shown in Fig. 4-1.

If the lattice structure is extended into a three-dimensional array, discuss the effect of the additional ions on the potential-energy function along the row $ABC$.



PROB. 4-7.

**4-8.** In order to remove an electron from metallic tungsten an energy of 13.47 ev is required.

*a.* If an electron possesses 99.99 per cent of this escape energy, how far from the surface can it wander?

*b.* Repeat *a* for energies of 90 per cent and 50 per cent of the escape energy.

*c.* What energy is required for the electron to move 10 ionic spacings (say, $10^{-9}$ m) from the surface?

**4-9.** On the basis of the image-force theory plot a curve (to scale) of potential energy vs. distance for tungsten.

**4-10.** What fraction of the electrons emitted from a metal have normal-to-the-surface-directed energies greater than the average energy of the emitted electrons?

**4-11.** If 10 per cent of the thermionic-emission current is collected (under space-charge-free conditions), what must be the retarding voltage at the surface of the metal? The filament temperature is 2000°K.

**4-12.** What fraction of the thermionic current will be obtained with zero applied voltage between the cathode and anode of a diode? The work function of the cathode is 4.50 volts, and the work function of the anode is 4.75 volts. The cathode temperature is 2000°K.

**4-13.** A plane cathode having a work function of 3.00 volts is connected directly to a parallel plane anode whose work function is 5.00 volts. The distance between anode and cathode is 2.00 cm. If an electron leaves the cathode with a normal-to-the-surface velocity of $5.93 \times 10^5$ m/sec, how close to the anode will it come?

**4-14.** A diode has an oxide-coated cathode operating at a temperature of 1000°K. With zero plate voltage the anode current is essentially zero, indicating that the contact potential is high enough to keep most of the electrons from reaching the plate. The applied voltage is increased so that a small current is drawn. Show that there is a tenfold increase in current for every 0.2-volt increase in voltage.

**4-15.** A diode, with plane-parallel electrodes, is operated at a temperature of 1500°K. The filament is made of tungsten, the area being such that a saturation current of 10 $\mu$a is obtained. The contact difference of potential between cathode and anode is 0.5 volt with the cathode at the higher potential.

*a.* What current is obtained with zero applied voltage?

*b.* What applied voltage will yield a current of 1 $\mu$a?

*c.* What fraction of the electrons emitted from this filament can move against an *applied retarding* field of 1 volt?

**4-16.** What accelerating field must be applied to the surface of a tungsten emitter operating at 2500°K in order to increase the zero-field thermionic emission by 1 per cent?

**4-17.** Calculate the electron-emission current density from the surface of a cold tungsten metal if the electric-field intensity is

*a.* $10^8$ volts/m.

*b.* $10^9$ volts/m.

*c.* $3 \times 10^9$ volts/m.

*d.* $10^{10}$ volts/m.

This problem illustrates how large an electric field is needed to obtain appreciable cold-cathode emission.

**4-18.** Indicate by letter which of the following statements are true:

*a.* The work function of a metal is always less than the potential barrier at the surface of a metal.

*b.* The potential barrier at the surface of a metal is a solid hill made up of the material of the metal.

*c.* The ionic structure of a metal shows that the inside of the metal is not an equipotential volume.

*d.* At absolute zero the electrons in a metal all have zero energy.

*e.* The energy method of analyzing the motion of a particle can be applied to uncharged as well as to charged particles.

*f.* The ionic structure of a metal shows that the surface of a metal is not a specific quantity.

*g.* For an electron to escape from a metal, the potential barrier at the surface of the metal must first be broken down.

*h.* The FDS distribution function for the electrons in a metal shows how many electrons are close to a nucleus and how many are far away.

*i.* The number of secondary electrons which leave a metal is always greater than the number of primary electrons striking the metal surface.

**4-19.** Indicate by letter which of the following statements are true:

*a.* The potential energy as a function of distance along a row of ions *inside* a metal varies very rapidly in the immediate neighborhood of an ion but is almost constant everywhere else inside the metal.

*b.* The potential-energy barrier at the surface of a metal *cannot* be explained on the basis of the modern crystal-structure picture of a metal, but it can be explained on the basis of classical electrostatics (image forces).

*c.* In order to remove any one of the free electrons from a metal, it is necessary only to give this electron an amount of energy equal to the work function of the metal.

*d.* The symbol $E_M$ used in the energy distribution function represents the maximum number of free electrons per cubic meter of metal at absolute zero.

*e.* The area under the energy distribution curve represents the total number of free electrons per cubic meter of metal at any temperature.

*f.* The Dushman equation of thermionic emission gives the current that is obtained from a heated cathode as a function of applied plate voltage.

## REFERENCES

1. SEARS, F. W., "Principles of Physics," Vol. II, Addison-Wesley Press, Inc., 1947.
2. SEITZ, F., and R. P. JOHNSON, *J. Applied Phys.*, **8**, 246, 1937.
3. DAVISSON, C., and L. H. GERMER, *Proc. Natl. Acad. Sci. U.S.*, **14**, 619, 1928.
4. RUPP, E., in "The Interference of Electrons," edited by P. DEBYE, Blackie & Son, Ltd., Glasgow, 1931.
5. FOWLER, R. H., and L. NORDHEIM, *Proc. Roy. Soc. (London)*, **119**, 173, 1928.
   OPPENHEIMER, J. R., *Phys. Rev.*, **31**, 66, 1928; *Proc. Natl. Acad. Sci. U.S.*, **14**, 363, 1928.
6. MILLIKAN, R. A., and C. F. EYRING, *Phys. Rev.*, **27**, 51, 1926.
   MILLIKAN, R. A., and C. C. LAURITSEN, *Proc. Natl. Acad. Sci. U.S.*, **14**, 45, 1928.
   EYRING, C. F., S. S. MACKEOWN, R. A. MILLIKAN, *Phys. Rev.*, **31**, 900, 1928.
   STERN, T. E., B. S. GOSSLING, and R. H. FOWLER, *Proc. Roy. Soc. (London)*, **124**, 699, 1929.
7. SLACK, C. M., and L. F. EHRKE, *J. Applied Phys.*, **12**, 165, 1941.
8. SPANGENBERG, K. R., "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.
   McKAY, K. G., an extensive review in "Advances in Electronics," pp. 65–130, Academic Press, 1948.
   "Industrial Electronics Reference Book," by engineers of the Westinghouse Electric Corp., p. 26, John Wiley & Sons, Inc., New York, 1948.
9. WOOLDRIDGE, D. E., *Phys. Rev.*, **56**, 562, 1939.
   KHLEBNIKOV, H., *Tech. Phys. U.S.S.R.*, **5**, 592, 1938.
   ZWORYKIN, V. K., and G. A. MORTON, "Television," John Wiley & Sons, Inc., New York, 1940.
   BRUINING, H., and J. H. DEBOER, *Physica*, **5**, 17, 901, 913, 1938.
10. HASTINGS, A. E., *Phys. Rev.*, **57**, 695, 1940.
    COPELAND, P. L., *ibid.*, **58**, 604, 1940.
11. ZWORYKIN and MORTON, *op. cit.*
12. OLIPHANT, M. L. E., *Proc. Roy. Soc. (London)*, **127**, 373, 1930.
    OLIPHANT, M. L. E., and P. B. MOON, *ibid.*, **127**, 388, 1930.
    MOON, P. B., *Proc. Cambridge Phil. Soc.*, **27**, 573, 1931.

### General Reference

DOW, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

# CHAPTER 5

# STATISTICAL ELECTRON THEORY OF METALS

IN THE preceding chapter the concept of the distribution in energy of the free electrons in a metal was introduced. This statistical behavior of the particles will now be examined in some detail. The end results given in the preceding chapter will be justified. Several new concepts will be introduced, and some of the phenomena already studied will be reexamined from a more theoretical point of view.

**5-1. The Distribution Functions.** In order to give an accurate account of the behavior of the electrons in a metal, it would be necessary to set up and solve the equations of motion of the particles present. If it is noted that there are approximately $10^{29}$ atoms per cubic meter of a metal, it becomes evident that this is an impossible task. It is obvious, therefore, that any attempt to predict the exact behavior of each individual electron (the microscopic behavior) must be abandoned. However, it is possible to determine by statistical methods the behavior of a great number or an ensemble of electrons (the macroscopic behavior).

For example, if one is interested in the ages of the residents of the United States, a detailed list of the exact age of each person is scarcely desired. Rather, a curve showing the number of persons whose ages are between certain limits, say, between 0 and 5 years, 5 and 10 years, 10 and 15 years, etc., is wanted. This information, tabulated in exactly this way, is to be found in the *World Almanac*. These data are reproduced in Table 5-1, where, for convenience, the numbers have been listed to the nearest 100,000.

How can these data be plotted so that the plot will yield useful information at a glance? Several possibilities suggest themselves. With the axis of abscissa representing *age*, an ordinate at the middle of each age interval equal in magnitude to the number of persons whose ages lie in this interval might be erected. Then, to obtain a graph, a smooth curve might be drawn through these points, as illustrated in Fig. 5-1.

This procedure, though appearing quite natural and logical, is actually misleading and quite meaningless. Thus the graph seems to indicate that there are 10.6 million persons 6 years old, 10.7 million persons 7 years old, 10.8 million persons 8 years old, and 10.9 million persons 9 years old. This gives a total of 43.0 million persons for these four ages, whereas the table

gives only 10.7 million as the total number of persons whose ages lie in the age interval from 5 to 10 years. Thus, by selecting four particular ages, about four times as many persons appear to be in a given age interval

TABLE 5-1 *

AGE GROUPS IN THE UNITED STATES IN 1940
(Total population = 131.7 million)

| Age range | Number, millions | Age range | Number, millions |
|-----------|------------------|-----------|------------------|
| 0–5 | 10.5 | 50–55 | 7.3 |
| 5–10 | 10.7 | 55–60 | 5.9 |
| 10–15 | 11.7 | 60–65 | 4.7 |
| 15–20 | 12.3 | 65–70 | 3.8 |
| 20–25 | 11.6 | 70–75 | 2.6 |
| 25–30 | 11.1 | 75–80 | 1.5 |
| 30–35 | 10.2 | 80–85 | 0.8 |
| 35–40 | 9.5 | 85–90 | 0.3 |
| 40–45 | 8.8 | 90–95 | 0.1 |
| 45–50 | 8.3 | 95–100 | 0.0 |

* From the *World Almanac.*

as there actually are. In fact, by choosing different ages, different totals are obtained for the entire population. Thus, in the limit, by choosing an infinite number of ages, instead of four as above, and since the ordinate
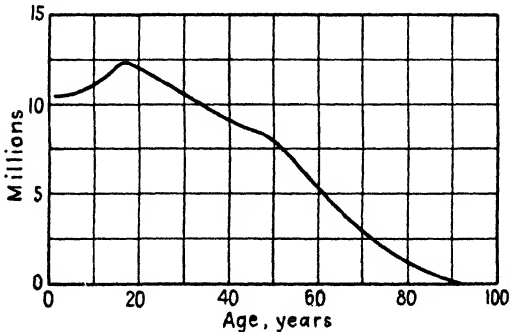


FIG. 5-1. An incorrect way to plot the distribution function in age of the people in the United States.

corresponding to each is finite, an infinite sum is obtained. This is, of course, inconsistent with our data.

If the ordinates are chosen as *density* rather than *number* of persons in a given age interval, then the apparent inconsistency mentioned above

vanishes. The density is obtained by dividing the number of persons in each age interval by the length of that interval. Since the intervals chosen in Table 5-1 are each equal and cover a period of 5 years, the numbers in this table need only be divided by 5 in order to convert them to a density. A plot of these densities vs. age gives a curve that is identical in shape with that of Fig. 5-1 except for the scale of the axis of ordinates, which differs
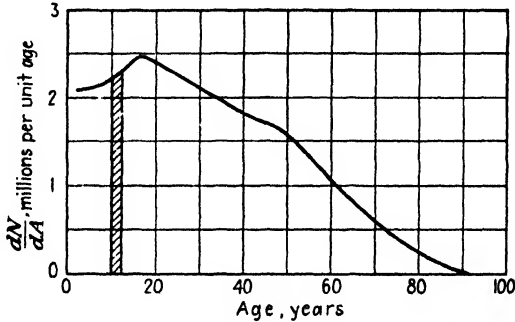


FIG. 5-2. The correct way to plot the distribution function in age of the people in the United States.

by a factor of 5. And yet how different these two curves are in physical meaning!

The vertical axis of Fig. 5-2 is labeled $dN/dA$, since this is the mathematical representation of a density. The density will be denoted by the symbol $\rho$. Thus

$$dN = \rho\, dA \qquad (5\text{-}1)$$

where $dN$ denotes the number of persons in the age interval $dA$. Since many different densities and distribution functions will be encountered in our work, distinctive subscripts will be used to distinguish one from another. For example, Eq. (5-1) will be written in the more informative form

$$dN_A = \rho_A\, dA \qquad (5\text{-}2)$$

where the subscript $A$ is to remind us that a distribution in age is being considered.

The number of persons having ages between 10 and 12 years is now represented by $dN_A$, with $\rho_A = 2.26$ millions per year chosen as the mean ordinate between 10 and 12 years, and $dA$ is taken as $12 - 10 = 2$ years. Thus $dN_A = \rho_A\, dA = 4.52$ millions. Geometrically, this is the shaded area of Fig. 5-2. Evidently, now, the total population is given by

$$N = \int dN_A = \int \rho_A\, dA \qquad (5\text{-}3)$$

or simply the total area under the curve.

The age distribution function has been considered in some detail because an understanding of it should facilitate the meaning and significance of the electronic distribution functions now to be introduced. We shall first be concerned with the distribution in energy of the free electrons in a metal. By analogy with Eq. (5-2), this will be given mathematically by the relation

$$dN_E = \rho_E \, dE \qquad (5\text{-}4)$$

where $dN_E$ represents the number of free electrons per cubic meter whose energies lie in the energy interval $dE$ electron volts and where $\rho_E$ gives the density of electrons in this interval.

The question that immediately presents itself is: What is the mathematical expression for the density function $\rho_E$? Unfortunately there is no almanac of the electronic world which might contain this information. Fermi,[1] and independently Dirac,[2] taking into account the quantum nature of the electron, other physical facts, and the laws of probability, deduced this most probable distribution function for electrons. The derivation of this distribution function [3] is beyond the scope of this text, but the significance of the distribution function and the inferences that can therefrom be drawn will be considered in some detail. The application of this function to the theory of metals is due primarily to Sommerfeld; it will be called the Fermi-Dirac-Sommerfeld function and abbreviated FDS.

Strictly speaking, a distribution function should depend upon the time. Because of the birth and death of various persons, the age distribution function will depend upon the instant at which the distribution is being considered, and it varies from instant to instant. It is because of this that a census is taken at regular intervals. Similarly, owing to collisions with ions and interactions among themselves, the electrons are continuously changing their energies. Although an infinite number of possible distribution functions exist theoretically, it is found that one of these distribution functions occurs much more frequently than all the others. This is the most probable distribution and is assumed to describe adequately the state of the electrons at all times.

The FDS energy density function is expressed in the form

$$\rho_E = \frac{\gamma E^{\frac{1}{2}}}{1 + \epsilon^{(E-E_M)/E_T}} \qquad (\text{electrons/m}^3)/\text{ev}$$

where $\gamma$ is a constant defined by

$$\gamma \equiv \frac{4\pi}{h^3} (2me)^{\frac{3}{2}}$$

$$= 6.82 \times 10^{27} \qquad (\text{electrons/m}^3)/(\text{ev})^{\frac{3}{2}}$$

(5-5)

and where $m$ is the mass of the electron in kilograms; $h$ is a constant (its dimensions are joule-seconds) first introduced by Planck; $\epsilon$ is the base of natural logarithms; $E$ is the energy of the electron in electron volts; $E_M$ is a parameter introduced in Sec. 4-6 and again to be discussed later; and $E_T$ is defined by the relationship

$$eE_T \equiv kT \qquad (5\text{-}6)$$

where $k$ is the Boltzmann gas constant in joules per degree Kelvin; $T$ is the temperature in degrees absolute or Kelvin; and $e$ is the electronic charge in coulombs. As before the quantity $E_T$ is the *electron-volt equivalent of temperature* [4] and is a convenient abbreviation. The numerical values of the physical constants introduced here are contained in Appendix I. Equation (5-6) becomes, upon substituting numerical values for the constants contained in the equation,

$$E_T = \frac{T}{11,600} \qquad (5\text{-}7)$$

This permits a rapid conversion from temperature to the electron-volt equivalent.

Several points must be emphasized before discussing Eq. (5-5). Since our interests are confined only to the free electrons, it will be assumed that there are no potential variations within the metal. Hence, there must be, a priori, the same number of electrons in each cubic meter of the metal. That is, the density in space (electrons per cubic meter) is a constant. However, within each unit volume of metal there will be electrons having all possible energies. It is this distribution in energy (per cubic meter of the metal) that is expressed by Eq. (5-5). To distinguish it from a simple density in coordinate space, Eq. (5-5) will be referred to as a density in "energy space." Similarly, the curve of Fig. 5-2 can be referred to as a density in "age space."

**5-2. The Completely Degenerate Function.** At a temperature of absolute zero, Eq. (5-5) attains a very striking form known as the completely degenerate function. When $T = 0°K$, then $E_T = 0$ and two possible conditions exist: (1) If $E > E_M$, then the exponential term becomes infinite, whence $\rho_E = 0$. Consequently, *there are no electrons with energies greater than $E_M$ at absolute zero of temperature.* That is, $E_M$ is the *maximum* energy that any electron may possess at absolute zero. This is the physical significance of the important quantity $E_M$, which is often referred to as the "Fermi characteristic energy" or the "Fermi level." An analytical expression for this quantity will be obtained in the next section. (2) If $E < E_M$, then the exponential in Eq. (5-5) becomes zero. Hence

$$\begin{aligned} \rho_E &= \gamma E^{\frac{1}{2}} \qquad \text{for } E < E_M \\ \rho_E &= 0 \qquad \text{for } E > E_M \end{aligned} \right\} \quad \text{when } E_T = 0 \qquad (5\text{-}8)$$

Evidently, a *band of energies* exists extending from zero to $E_M$ at absolute zero.   For temperatures other than 0°K the energy band extends beyond the $E_M$ level.

A plot of the distribution in energy given by Eqs. (5-5) and (5-8) for metallic tungsten at $T = 0°K$ and $T = 2500°K$ is given in Fig. 5-3.   The area under each curve is simply the total number of particles per cubic meter of the metal, whence the two areas must be equal.   Also, the curves for all temperatures must pass through the same ordinate, namely, $\rho_E = \gamma E_M^{\frac{1}{2}}/2$, at the point $E = E_M$, as is evident from Eq. (5-5).

A most important characteristic is to be noted, *viz.*, the distribution function changes only very slightly with temperature, even though the temperature change is as great as 2500°K. The effect of the high temperature is merely to give those electrons having the high energies at absolute zero (those in the neighborhood of $E_M$) still higher energies, whereas those having lower energies have been left practically undisturbed.   Since the curve for $T = 2500°K$ approaches the energy axis asymptotically, a few electrons will have large values of energy.   It is these relatively few high-energy electrons which play the important role in thermionic emission.
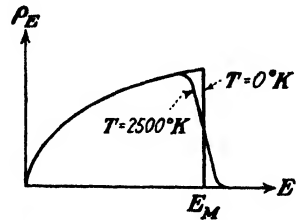


Fig. 5-3.   Energy distribution in metallic tungsten at 0° and 2500°K.

It may appear strange that the discussion of the energy function was begun by considering certain characteristics at the unattainable absolute zero of temperature.   The reasons for doing this were twofold.   (1) It made it possible to ascribe a physical interpretation to the constant $E_M$. (2) An inspection of Fig. 5-3 shows that the distribution function is only slightly disturbed even for temperatures as high as 2500°K.   This will permit the use of the extremely simple expression (5-8) instead of the complicated mathematical function (5-5) in certain phases of the work.   The results can then be extrapolated to any finite temperature.   Of course, when phenomena that depend explicitly upon the temperature are under survey, this method will not be used.

**5-3. The Expression for $E_M$.**   An approximate expression for $E_M$ may be obtained on the basis of the completely degenerate function.   The area under the curve of Fig. 5-3 represents the total number of free electrons (as always, per cubic meter of the metal).   Thus

$$N = \int_0^{E_M} \gamma E^{\frac{1}{2}} \, dE = \tfrac{2}{3}\gamma E_M^{\frac{3}{2}}$$

or

$$E_M = \left(\frac{3N}{2\gamma}\right)^{\frac{2}{3}} \qquad \text{ev} \qquad (5\text{-}9)$$

Inserting the numerical value of the constant $\gamma$ in this expression, there results

$$E_M = 3.64 \times 10^{-19} N^{\frac{1}{3}} \qquad \text{ev} \qquad (5\text{-}10)$$

Since the density $N$ (electrons per cubic meter) varies from metal to metal, then $E_M$ will also vary among metals. Knowing the specific gravity, the atomic weight, and the number of free electrons per atom, it is a simple matter to calculate $N$, and so $E_M$. For most metals the numerical value of $E_M$ is less than 10 ev.

*Example.* The specific gravity of tungsten is 18.8, and its atomic weight is 184.0.[*] Assume that there are two free electrons per atom. Calculate the numerical values of $N$ and $E_M$.

*Solution.* A quantity of any substance equal to its molecular weight in grams is a *mole* of that substance. Further, 1 mole of any substance contains the same number of molecules as 1 mole of any other substance. This number is *Avogadro's number* and equals $6.02 \times 10^{23}$ molecules per mole. Thus

$$N = 6.02 \times 10^{23} \frac{\text{molecules}}{\text{mole}} \times \frac{1 \text{ mole}}{184 \text{ g}} \times 18.8 \frac{\text{g}}{\text{cm}^3} \times \frac{2 \text{ electrons}}{\text{atom}} \times \frac{1 \text{ atom}}{\text{molecule}}$$

$$= 12.3 \times 10^{22} \frac{\text{electrons}}{\text{cm}^3} = 1.23 \times 10^{29} \frac{\text{electrons}}{\text{m}^3}$$

since for tungsten the atomic and the molecular weights are the same. Therefore, for tungsten,

$$E_M = 3.64 \times 10^{-19} (123 \times 10^{27})^{\frac{1}{3}}$$

$$= 8.95 \text{ ev}$$

It is to be remembered that the expression (5-9) is strictly true only at absolute zero. In order to calculate $E_M$ at any temperature other than zero, it is necessary to integrate Eq. (5-5) over all possible energies. Thus

$$N = \int_0^\infty \frac{\gamma E^{\frac{1}{2}}}{1 + \epsilon^{(E - E_M)/E_T}} \, dE \qquad (5\text{-}11)$$

This integration is a very difficult one to perform. It is carried out as a series expansion.[3,5] The result is

$$E_M = E_{M0} \left( 1 - \frac{\pi^2}{12} \frac{E_T^2}{E_{M0}^2} \right) \qquad (5\text{-}12)$$

where the term

$$E_{M0} = \left( \frac{3N}{2\gamma} \right)^{\frac{1}{3}}$$

is the first approximation value and is the value given by Eq. (5-9).

[*] The atomic weights of the elements are given in the Periodic Table (Appendix III).

At the temperature $T = 0°K$ the value of $E_M$ given by Eq. (5-12) reduces to $E_{M0}$, as it should. At the temperature $T = 3000°K$ the numerical value of the second term in this expression is found to be less than 0.1 per cent of the first term for tungsten. For all practical purposes, therefore, $E_M$ may be taken as a constant, independent of the temperature.

**5-4. The Speed Distribution Function.** Consider the query: How many electrons have a total speed between the values $v$ and $v + dv$? To use the terminology previously introduced, what is the density function in *speed space?* * Certainly this is a valid question; in fact, it is just as valid as it would be were we to ask for the distribution in heights instead of the distribution in age in our original consideration of distribution functions. One fundamental difference exists between electronic distribution functions and those pertaining to human beings, however. Whereas the height of a person is none too intimately connected with his age, the speed of an electron and its kinetic energy are directly dependent upon each other. If one is known, the other may readily be calculated from the relationship

$$eE = \tfrac{1}{2}mv^2$$

It is possible, with the aid of this relation, to eliminate $E$ and $dE$ from the expression $dN_E = \rho_E\,dE$. The result is

$$dN_E = \frac{8\pi m^3}{h^3} \frac{v^2}{1 + \epsilon^{(E-E_M)/E_T}}\,dv \qquad \text{electrons/m}^3$$

Since this equation represents a distribution in speed, the subscript of $dN$ will be changed to the appropriate form. Thus

where
$$\left.\begin{array}{c} dN_v = \rho_v\,dv \qquad \text{electrons/m}^3 \\[2mm] \rho_v = \dfrac{8\pi m^3}{h^3}\dfrac{v^2}{1 + \epsilon^{(E-E_M)/E_T}} \qquad \text{(electrons/m}^3)/(\text{m/sec}) \end{array}\right\} \qquad (5\text{-}13)$$

represents the density in speed space. $E$ has not been written in terms of $v$ in the exponent of this expression, since the change will merely serve to make the equation appear more complicated.

**5-5. The Velocity Distribution Function.** It will be found necessary to know not only the magnitudes but also the directions of the velocities of the electrons. For example, in the derivation of the thermionic-emission equation, it is necessary to calculate the number of electrons that escape through the surface of the metal per second. Evidently, it is only a certain portion of the electrons which are surface-directed that may escape from the metal.

---

* A geometrical interpretation will be given to *speed space* in the next section. It should also be noted that it is customary to use the terms "velocity" and "speed" synonymously in referring to distribution functions.

To specify completely the motion of an electron, its $X$, $Y$, and $Z$ velocity components must be known simultaneously. Hence, the velocity distribution now to be determined must furnish an answer to the following question: How many electrons have $X$-directed velocities between $v_x$ and $v_x + dv_x$ and simultaneously have $Y$-directed velocities between $v_y$ and $v_y + dv_y$ and $Z$-directed velocities between $v_z$ and $v_z + dv_z$? A somewhat analogous query concerning the peoples of the United States would be: How many persons have ages between $A$ and $A + dA$, heights between $H$ and $H + dH$, and weights between $W$ and $W + dW$? To be more specific, one might ask for the number of persons between 21 and 22 years
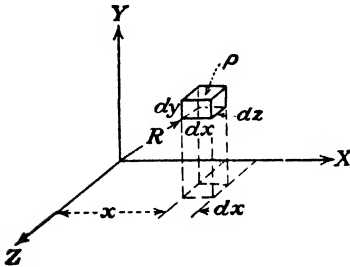


Fig. 5-4. A rectangular element of volume in coordinate space.

of age whose heights range from 5 ft 8 in. to 5 ft 10 in. and who weigh between 160 and 170 lb. A census analyst could supply the answer. How is the corresponding solution for the velocity distribution function to be obtained? Obviously, a density in a three-dimensional velocity space is being sought. With this end in mind, it is advisable to digress for the moment in order to introduce a concept that will permit a geometrical interpretation to be given to our problem.

If three mutually perpendicular axes $XYZ$ are erected, then the position of a point whose coordinates are $x$, $y$, $z$ is uniquely located with respect to this frame of reference. This is the well-known *coordinate space*. Similarly, the velocity of a particle can be represented geometrically by erecting three mutually perpendicular axes which are labeled $v_x$, $v_y$, $v_z$. This is called the *velocity space*, and any point in this space has associated with it the three components of velocity of the particle.

An element of volume in the coordinate space is given by $dx\,dy\,dz$, as illustrated in Fig. 5-4. In this space, the number of points contained in the volume element $dx\,dy\,dz$ is

$$dN = \rho\,dx\,dy\,dz \qquad (5\text{-}14)$$

where $\rho$ is the density of points in the region of the volume element. An inspection of Fig. 5-4 shows that this expression gives the number of electrons whose coordinates lie in the range from $x$ to $x + dx$, $y$ to $y + dy$, and $z$ to $z + dz$. This may be expressed by saying that Eq. (5-14) gives the number of points whose coordinates lie in the range $dx\,dy\,dz$. Each "point" specifies, of course, the position of one electron.

Similarly, an element of volume in velocity space is given by $dv_x\,dv_y\,dv_z$ and may be illustrated geometrically as in Fig. 5-5. Evidently, the num-

ber of points contained in this element of volume is given by

$$dN = \rho\, dv_x\, dv_y\, dv_z \tag{5-15}$$

where $\rho$ is an appropriate density function. For convenience the symbol $\tau$ will be used to indicate that velocity space is being considered. An element of volume in this space will be abbreviated to $d\tau$. That is, the symbol $d\tau \equiv dv_x\, dv_y\, dv_z$.

Now, in accordance with the previous practice, Eq. (5-15) can be written in the more informative form

$$dN_\tau = \rho_\tau\, d\tau \tag{5-16}$$

where $dN_\tau$ gives the number of particles per cubic meter whose velocities
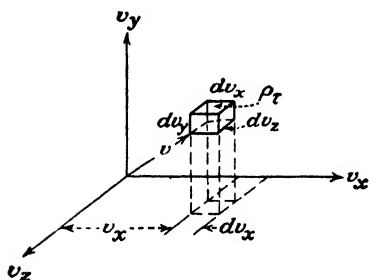


Fig. 5-5.   A rectangular element of volume in velocity space.
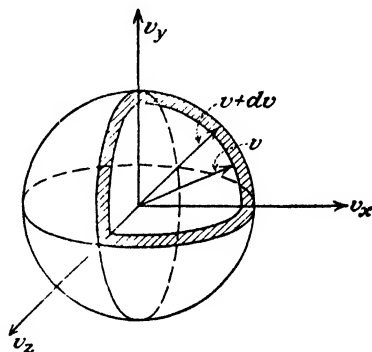


Fig. 5-6.   A spherical shell of inner radius $v$ and outer radius $v + dv$ in velocity space.

lie in the range $d\tau$ and where $\rho_\tau$ is the density of particles in this range. In other words, $\rho_\tau$ is the density of points in velocity space per unit volume of metal.

Now, consider the question set at the beginning of this section. This may be reworded to: What is the value of $\rho_\tau$ expressed as a function of $v_x$, $v_y$, and $v_z$? The answer must come from Eq. (5-13) since there is a simple connection between the speed and the velocity components, *viz.*,

$$v^2 = v_x{}^2 + v_y{}^2 + v_z{}^2 \tag{5-17}$$

Interpreted geometrically, $v$ is a radius vector in velocity space (in complete agreement with the notion of $R$, a radius vector in coordinate space, which is defined by the expression $R^2 = x^2 + y^2 + z^2$ and is shown in Fig. 5-4).

Consider now the spherical shell contained between the spherical surfaces of radii $v$ and $v + dv$, respectively, as illustrated in Fig. 5-6. The expression $dN_v = \rho_v\, dv$ gives the number of points contained within this

spherical shell because it represents the number of electrons having speeds between $v$ and $v + dv$. Since the volume of this shell is $4\pi v^2 \, dv$, the number of points per unit volume of velocity space is

$$\frac{dN_v}{4\pi v^2 \, dv} = \frac{\rho_v}{4\pi v^2} = \frac{2m^3}{h^3} \frac{1}{1 + \epsilon^{(E - E_M)/ET}}$$

where use has been made of Eq. (5-13). But this is exactly what is meant by the symbol $\rho_r$. Therefore,

$$\rho_r = \frac{2m^3}{h^3} \frac{1}{1 + \epsilon^{(E - E_M)/ET}} \qquad (\text{electrons/m}^3)/(\text{m/sec})^3 \qquad (5\text{-}18)$$

is the desired distribution function in velocity space. The energy $E$ in the exponent may be expressed in terms of $v_x$, $v_y$, and $v_z$ by the relation

$$eE = \tfrac{1}{2}m(v_x{}^2 + v_y{}^2 + v_z{}^2) \qquad \text{joules} \qquad (5\text{-}19)$$

**5-6. The Exclusion Principle.** Consider the distribution function Eq. (5-18) at the temperature of $0°K$. For values of $E$ such that $E > E_M$, the exponent in the denominator is infinite, whence $\rho_r = 0$. This is simply the result previously used to show that the maximum energy of the electrons at absolute zero of temperature is $E_M$. This implies the existence of a corresponding maximum speed $v_M$, which is given by the relation

$$\tfrac{1}{2}mv_M{}^2 = eE_M \qquad \text{joules} \qquad (5\text{-}20)$$

That is, if a sphere of radius $v_M$ were drawn about the origin in velocity space, *there would be no electrons outside this sphere, for* $T = 0°K$.

To determine the density of points within the sphere, $E$ is set less than $E_M$ in Eq. (5-18), and $T$ is made to approach zero. The exponential term approaches zero, and $\rho_r$ reduces to

$$\rho_r = \frac{2m^3}{h^3} \qquad (\text{electrons/m}^3)/(\text{m/sec})^3 \qquad (5\text{-}21)$$

It must be kept in mind, however, that this expression, as well as all similar expressions considered, implicitly involves the volume of the metal. Thus, for a volume $V$ cubic meters of the metal, the total density function (density in velocity space for all the electrons contained in the metal) is given by Eq. (5-21) multiplied by the volume of the metal. Thus, in general, corresponding to Eq. (5-21) per unit volume, will be the expression

$$\rho_r' = \frac{2m^3}{h^3} V \qquad \text{electrons/(m/sec)}^3 = \text{const} \qquad (5\text{-}22)$$

for the entire volume. Thus the density of points within a sphere of radius

$v_M$ drawn in velocity space, and representing the total volume of the metal, is a constant and is given by Eq. (5-22).

Since the total number of electrons per unit volume of velocity space is $(2m^3/h^3)V$, then "the volume occupied by one electron" is $h^3/2m^3V$. The expression in quotes is to be understood in the following light: If the volume of the sphere in velocity space is divided into elements of volume, each of magnitude $h^3/2m^3V$, then in each volume element there will be one, and only one, electron. This is a very remarkable result which has far-reaching consequences; it follows directly from a principle first introduced by Pauli [6] and which is known as the Pauli "exclusion principle."

This principle is not confined in its scope to phenomena of electrons in metals but applies equally well to many other physical systems. In fact, this principle was first introduced in order to account for the structure of the periodic table of chemical elements.

The reader should not gain the impression that the exclusion principle was derived from considerations of the distribution function; rather, the form of the distribution function is a consequence of this principle. As a



FIG. 5-7.    Illustrating the exclusion principle.

matter of fact, no "proof" of the exclusion principle can be given. Its validity is tested indirectly by comparing the consequences demanded by it with the results of experiment. The agreement that results is adequate proof of its basic nature and validity.

This principle is so important that certain of the more significant features of it will be considered. First, it is noted that, if an element of volume equal to $h^3/m^3V$ is chosen, then two electrons will occupy this volume. Suppose that the velocity space consists of cubical elements of volume, the linear dimensions of which are $h/mV^{\frac{1}{3}}$ on a side. Figure 5-7 indicates the projection on the $v_x v_y$ plane. Numbers have been put on each axis so that any particular block may be designated. These numbers will be denoted by $n_x$, $n_y$, $n_z$ along the $v_x$, $v_y$, and $v_z$ axes, respectively. For example, the shaded block is specified as $n_x = 2$, $n_y = 3$, $n_z = 0$. These numbers are known as "quantum numbers"; the entire process represented by Fig. 5-7 is known as "quantization." Since only two electrons may exist in any elementary block, then, of course, only two electrons may have the same three quantum numbers. This last statement is another form of the exclusion principle. It is easy to show that the dimensions of $h/mV^{\frac{1}{3}}$, the unit of velocity space, is meters per second, as it must be if Fig. 5-7 is to have any meaning.

By virtue of Eq. (5-19), each block or cell represents a different amount of energy. Those farther from the origin represent larger amounts of energy. Thus the blocks are referred to as "energy states" or "energy levels."

The reader may raise the legitimate objection that such a designation introduces a certain element of vagueness or uncertainty in the specification of the velocity of the particle, since the specification that an electron is in the (2, 3, 0) state (for example, the shaded block of the diagram) does not exactly specify the velocity components of the electron. This specification merely states that the electron has an $X$-directed velocity between $2h/mV^{\frac{1}{3}}$ and $3h/mV^{\frac{1}{3}}$. a $Y$-directed velocity between $3h/mV^{\frac{1}{3}}$ and $4h/mV^{\frac{1}{3}}$, and a $Z$-directed velocity between 0 and $h/mV^{\frac{1}{3}}$. This, as a matter of fact, is the only information that can be given. Certainly the statistical equations do not dictate more exact values for the components of velocity. In fact, it will be recalled that the entire discussion of distribution functions was based upon the supposition that the *exact* energy or speed or velocity could *not* be specified.

**5-7. The Uncertainty Principle.** In accordance with the foregoing discussion, the uncertainty in each component of velocity is $h/2mV^{\frac{1}{3}}$, which represents one-half the linear dimensions of a quantum box. This quantity increases as $V$ decreases. If, then, one attempts to specify the exact position of the electron (by considering smaller and smaller volumes of the metal), the more indeterminate becomes the specification of the velocity of the electron. Conversely, the larger the volume of the metal chosen (thereby making the position of the electron more indeterminate), the smaller will be the linear dimensions $h/mV^{\frac{1}{3}}$ of the quantum blocks, and the more certain or precise will the specification of the velocity of the electron be. These conclusions are summed up in the *uncertainty principle*, which was first introduced by Heisenberg.[7]

According to the uncertainty principle, the measurement of physical quantities is essentially characterized by a lack of precision. In terms of our quantities, it states that the momentum and position of a particle can never be known exactly. There will always be an uncertainty in its momentum and an uncertainty in its position such that the product of these uncertainties is of the order of Planck's constant, $h$. In the case of electrons in metals, the uncertainty in the position of the electron is measured by $\frac{1}{2}V^{\frac{1}{3}}$, since we cannot isolate or label any particular electron. The corresponding uncertainty in the momentum is $m$ times $h/2mV^{\frac{1}{3}}$, where $m$ is the mass of the electron. The product of these two uncertainties equals $h/4$.

The uncertainty principle has been applied to many systems other than to electrons in metals. It has been here introduced to show that the indeterminacy discussed above expresses a fundamental behavior of nature and is not inherent solely in the behavior of electrons in metals.

**5-8. Electron Spin.** It has been found necessary, in order to explain certain spectroscopic and magnetic phenomena, to attribute to the electron four degrees of freedom. In addition to the translatory motion along the $X$, $Y$, and $Z$ axes, the electron is supposed to possess a certain rotation in space, just as the earth, in addition to traversing its orbit around the sun, also rotates on its own axis. This phenomenon is called *electron angular momentum, moment of momentum,* or simply *electron spin.*

In order to fit theory to experimental data, it is found necessary to assume that when an electronic system is subjected to a magnetic field the spin axis will orient itself either parallel or antiparallel to the direction of the field. In other words, the direction of rotation of the electron is either clockwise or counterclockwise about the direction of the field. The spin of the electron is thus said to be "quantized." In accordance with this discussion only two possible motions exist. These two possible states are characterized by two spin quantum numbers $+1$ and $-1$. The terms "positive" and "negative" spin are frequently used in this connection.

It is noted, therefore, that whereas the energy of the electrons in metals may assume practically an infinite number of values, all differing from each other by discrete amounts, the spin of the electron can exist in only two states. If such restrictions appear startling, then the quantization of charge and mass imposed by nature on an electron should be equally startling. As has been experimentally verified, there is but one possible "charge level" and but one "mass level," since all electrons have exactly the same charge and the same mass.

It is not entirely necessary to introduce the concept of electron spin in order to explain those electronic phenomena in metals with which we are concerned. It has been included, however, for the sake of completeness. Furthermore, it permits further insight into the exclusion principle.

It may have appeared strange to choose as the basic element of volume in velocity space one that permits two electrons per elementary volume element. This choice was made in order to take the electron spin into account. Thus the Pauli principle allows only one electron of each spin to exist in a given volume element. The exclusion principle may be re-worded to the form: Only one electron of a given system may have all four of its quantum numbers alike. Then no two electrons exist in the same "quantum state." The quantum state of an electron is now understood to be specified by four quantum numbers, three for velocity or momentum and one for spin.

**5-9. Zero-point Energy.** The difference in energy between energy levels is small, so that the distribution in energy may be considered to be practically continuous. However, one consequence of the discreteness of the energy levels is a definite, nonvanishing energy at the temperature of absolute zero, known as the *zero-point energy.* The reason for the existence

of this zero-point energy is easy to understand. It should be recalled that only two electrons (one with positive spin and one with negative spin) can exist in the zero-energy state. The next few electrons must possess the small, but finite, energy of the adjacent cells illustrated in Fig. 5-7. The following electrons must possess still higher energies, corresponding to the next vacant levels, etc. Thus, even at absolute zero, when the electrons occupy all the lowest energy levels, they still possess a very appreciable amount of energy. The magnitude of this energy will be calculated in the succeeding section.

**5-10. Average Values Calculated with the Aid of Distribution Functions.** One of the very important uses of the distribution functions is in the calculation of average quantities. If one is interested in the average value of some function of energy, the energy distribution function is used. If the average value of some function of the velocity is under consideration, the velocity distribution function is used. Clearly, the average age of the population of the United States is obtained by using the age distribution function.

The procedure to be followed in order to obtain such an average value is best illustrated by means of a specific example.

*Example.* Calculate the average energy $\bar{E}$ of the electrons of a metal at the temperature $T = 0°K$.

*Solution.* Since $E \, dN_E$ is equal to the total amount of energy that $dN_E$ electrons possess, then $E \, dN_E$ integrated over all the electrons is equal to the total amount of energy possessed by all of them. If this result is divided by the total number of electrons, $\int dN_E$, the average energy per electron is given by the expression

$$\bar{E} = \frac{\int E \, dN_E}{\int dN_E} = \frac{\int \rho_E E \, dE}{\int \rho_E \, dE} \tag{5-23}$$

Use Eq. (5-8) for $\rho_E$, and integrate between the limits from zero to $E_M$, since no electrons have energies greater than $E_M$ at this temperature. Then

$$\bar{E} = \frac{\int_0^{E_M} \gamma E^{\frac{3}{2}} \, dE}{\int_0^{E_M} \gamma E^{\frac{1}{2}} \, dE} = \frac{3}{5} E_M \tag{5-24}$$

For tungsten, this average energy is 5.37 ev. It will be shown in Chap. 8 that the average energy of a free monatomic gas molecule is $\frac{3}{2}E_T$ electron volts, independent of the mass of the particle. At room temperature, this represents only about 0.038 ev. Hence, the completely degenerate electron gas in metallic tungsten has an average energy per particle that is about 140 times that of a nondegenerate gas molecule at room temperature. This large average energy at absolute zero appears as a direct consequence of the exclusion principle.

By analogy with Eq. (5-23), the average value of any function of energy $\overline{f(E)}$ is given by

$$\overline{f(E)} = \frac{\displaystyle\int_0^\infty f(E)\,dN_E}{\displaystyle\int_0^\infty dN_E} = \frac{\displaystyle\int_0^\infty f(E)\rho_E\,dE}{\displaystyle\int_0^\infty \rho_E\,dE} \qquad (5\text{-}25)$$

where the upper limit of integration $\infty$ can be replaced by $E_M$, if the average is to be evaluated at absolute zero.

In the same way, the average value of any function of speed, $\overline{f(v)}$, is deduced from the expression

$$\overline{f(v)} = \frac{\displaystyle\int_0^\infty f(v)\,dN_v}{\displaystyle\int_0^\infty dN_v} = \frac{\displaystyle\int_0^\infty f(v)\rho_v\,dv}{\displaystyle\int_0^\infty \rho_v\,dv} \qquad (5\text{-}26)$$

As above, if the conditions under consideration are those at absolute zero, the upper limit in these integrals is to be replaced by $v_M$.

Unfortunately, the integrals in Eqs. (5-25) and (5-26) for any temperature $T$ may frequently be extremely difficult to evaluate. Since, however, the distribution functions vary so little with temperature, the error made in most calculations by using the completely degenerate functions instead of the general distribution functions will be small. If greater precision is desired, the integrals can be evaluated [3,5] as a power series in the temperature $T$.

Equations (5-25) and (5-26) will be used in evaluating average values in Chap. 8. These average values will, however, be the corresponding values for molecules and ions in gaseous-discharge devices. Although the general method of averaging is the same as that encountered for electrons, one very significant difference exists between the two systems, as will later be shown. This difference lies in the mathematical expression for the density functions that exist for each system. The density functions that apply for gaseous systems will be found to be less complicated than the FDS functions that apply for electronic systems in metals.

It may seem strange that none of the average values defined above was obtained by finding the area under a curve and then dividing by the base, a common mathematical definition of an average value. This arises from the fact that two types of average value exist that are of physical importance. The first is illustrated by the reading of a d-c ammeter in an electrical circuit, the meter reading being a measure of the average value of the current in the circuit into which it is inserted. If an oscillogram of the

wave shape of the current wave is available, the meter reading can be obtained by taking the area under one cycle of the curve, assuming it to be periodic, and then dividing this value by the period. This is the full-cycle *time* average. The second important type of average value is that considered in this section. It might be called a *weighted average*, in contrast with the time average, and it is this type of averaging that always appears in statistical calculations.

**5-11. Specific Heat of Electrons.** If the electrons in a metal were considered as a gas obeying classical laws, then, in accordance with the theorem of the equipartition of energy, they must contribute as much to the specific heat of the metal as the metallic ions. This conclusion is contrary to experimental fact, which favors the assumption that the electrons contribute almost nothing to the specific heat. This follows from the fact that the contribution to the specific heat of a metal by the ions alone is in agreement with experiment. It was a triumph of the FDS statistics to offer an explanation of this difficulty.

The average energy per electron at any temperature $T$ is given by Eq. (5-23),

$$\bar{E} = \frac{1}{N} \int_0^\infty \frac{\gamma E^{\frac{1}{2}} \, dE}{1 + \epsilon^{(E - E_M)/E_T}} \tag{5-27}$$

The value of this integral is [3,5]

$$\bar{E} = \frac{3}{5} E_{M0} \left( 1 + \frac{5\pi^2}{12} \frac{E_T{}^2}{E_{M0}{}^2} \right) \tag{5-28}$$

The specific heat of a body is, by definition, the amount of heat absorbed by unit mass of the body per degree rise of temperature. In other words, the specific heat of a body is a measure of the rate of change of energy with temperature. To evaluate the specific heat of the free electrons per gram of solid, the average energy per electron is multiplied by the number of electrons per gram of the material, and the result is differentiated with respect to the temperature. This quantity is proportional to $d\bar{E}/dT$ or to $d\bar{E}/dE_T$, since the term $E_T$ is proportional to the temperature. Therefore, from Eq. (5-28),

$$\frac{d\bar{E}}{dE_T} = \frac{\pi^2}{2} \frac{E_T}{E_{M0}}$$

As already mentioned, it will be shown that the classical average energy per particle is $\frac{3}{2}E_T$. Classically, therefore,

$$\frac{d\bar{E}}{dE_T} = \frac{3}{2}$$

The ratio of these two specific heats is

$$\frac{C_v]_{\text{FDS}}}{C_v]_{\text{Cl}}} = \frac{\pi^2}{3}\frac{E_T}{E_{MO}} \tag{5-29}$$

For tungsten at room temperature, this ratio is 0.01, approximately. This means that the specific heat of free electrons, as calculated on the basis of the FDS statistics, is only about 1 per cent of that demanded by the classical theory. This factor is negligible compared with the contribution by the ions. These results are in accord with experiment.

This same result is evident from an inspection of Fig. 5-3. The average energy per electron is determined from a distribution function that varies but slightly with temperature, even for large values of temperature. Hence, the specific heat, which is measured by the rate of change of average energy with temperature, should be small.

**5-12. Random Current Density.** A problem of importance in gaseous-discharge devices, and one that will serve as an introduction to the problem of thermionic emission, is that concerning the subject of random current density. By definition, the random current density is the charge per second striking normally a unit area of surface. In a gaseous discharge, a probe or sounding electrode is inserted into the discharge in order to measure this quantity. In this case, the term "unit area" in the definition refers to unit surface area of this electrode. In a metal, it is the current passing through the surface of the metal into the region outside that is of importance, and in this case it is the surface area of the emitting metal that is of importance. In either



FIG. 5-8. The parallelepiped $dx\,dy\,dz$ is situated within the metal with its surface $dy\,dz$ in the surface of the metal.

case, the flow is unidirectional only (surface-directed). The direction normal to the surface is chosen as the $X$ axis.

A problem closely related to the foregoing is first considered. The problem is: Of all of the electrons whose $X$-directed velocities lie in the range from $v_x$ to $v_x + dv_x$, how many will strike unit area of surface in unit time? In order to answer this question, refer to Fig. 5-8. The parallelepiped is chosen with the sides $dy$ and $dz$ in the surface and with the side $dx = v_x\,dt$ normal to the surface. The volume of this parallelepiped is evidently given by $dx\,dy\,dz = v_x\,dy\,dz\,dt$. The number of particles in this volume element with velocities in the range $d\tau$ is, of course, $dN_\tau(v_x\,dy\,dz\,dt)$ since $dN_\tau$ gives the number of electrons per cubic meter with velocity components in the range $d\tau$.

In this expression, $dN_\tau$ gives a measure of the number of electrons per cubic meter whose $X$-directed velocities lie in the range $dv_x$. Hence,

in the time $dt$ they all will have moved a distance $v_x\,dt$. Since this is exactly the length of the parallelepiped normal to the surface, then all the electrons in this box whose $X$-directed velocities lie in the range $dv_x$ will strike the surface in the time $dt$. If this is divided by $dt$ seconds and also by $dx\,dy$ square meters, it then follows that

$$dN_\tau' \equiv v_x\,dN_\tau \tag{5-30}$$

gives the number of electrons with $X$-directed velocities in the range $dv_x$, with $Y$-directed velocities in the range $dv_y$, and with $Z$-directed velocities in the range $dv_z$ that strike unit area in unit time.

The prime on the term $dN_\tau'$ is used as a reminder that this new quantity expresses the number of electrons whose velocities lie in the range $d\tau$ and that strike unit area in unit time, in distinction to the "static" distribution function $dN_\tau$. The relation between the two forms is given, of course, by Eq. (5-30). This relation shows that the quantity $dN_\tau$ when multiplied by $v_x$ changes the function from the number of electrons per cubic meter of the metal with velocities in a given velocity range to the number of electrons with velocities in a given velocity range striking unit area of surface per second.

In order to answer the above query, it is necessary to carry the operations one step farther. This arises from the fact that the desired quantity is one which expresses the number of electrons which strike unit area in unit time with $X$-directed velocities in the range from $v_x$ to $v_x + dv_x$, regardless of the other velocity components. This result is obtained from Eq. (5-30) by integrating over $v_y$ and $v_z$ from $-\infty$ to $+\infty$. The logic of this step can be made evident by reference to the three-dimensional distribution in age, height, and weight of the residents of the United States, $\rho_{AHW}$. Thus, to find the number of persons of *all* heights and weights between the ages $A$ and $A + dA$, it is necessary to concentrate on this age range and count the persons having all heights and weights in this interval. This counting process, in mathematical language, means simply that $\rho_{AHW}$ is integrated over all values of $H$ and $W$, $A$ remaining fixed. The result will be a function of $A$ alone and is, in fact, represented by the curve of Fig. 5-2.

If $dN_x'$ denotes the total number of electrons striking unit area of surface per unit time with $X$-directed velocities between $v_x$ and $v_x + dv_x$ and with all possible $Y$- and $Z$-directed velocities, then, according to the foregoing discussion,

where
$$\left. \begin{aligned} dN_x' &= v_x\,dN_x \\ dN_x &\equiv \int_{v_y=-\infty}^{\infty} \int_{v_z=-\infty}^{\infty} dN_\tau \end{aligned} \right\} \tag{5-31}$$

is called *the X-directed distribution function*. It gives the number of electrons per cubic meter with $X$-directed velocities between $v_x$ and $v_x + dv_x$ and *any* $Y$- or $Z$-directed velocities whatsoever.

The evaluation of the integrals involved in Eq. (5-31) is left as an exercise for the student (see Prob. 5-12).  The result is

$$dN_x = \rho_x \, dv_x \qquad \text{electrons/m}^3$$

where

$$\rho_x \equiv \frac{4\pi m^2 e E_T}{h^3} \ln \left[ 1 + \epsilon^{(E_M - E_x)/E_T} \right] \qquad \text{(electrons/m}^3\text{)/(m/sec)}$$

(5-32)

where ln means the logarithm to the base $\epsilon$ and $E_x$ is defined by the equation

$$eE_x \equiv \tfrac{1}{2} m v_x^2$$

The symbol $E_x$ is called the "X-directed energy."  This does *not* mean that the energy is being treated as a vector quantity, but rather that the energy corresponding to a particular component of velocity is being specified.  Similarly, the Y-directed energy is defined by the equation $eE_y \equiv \tfrac{1}{2} m v_y^2$.

The total charge striking unit area in unit time, or, simply, the random current density, is, therefore

$$J_r \equiv e \int_{v_x=0}^{\infty} dN_x' \qquad \text{amp/m}^2$$

(5-33)

It is to be noted that the lower limit of integration is taken as zero and not $-\infty$ since only particles with positive X-directed velocities can strike the surface.  The factor $e$ (the electronic charge) is needed in order to convert from number to charge.

**5-13. Thermionic Emission.**  A metal is heated to a temperature $T°\text{K}$.  It is desired to calculate the number of electrons that pass through unit surface area in unit time.

The problem here being considered is precisely that involved in the calculation of the random current density with the exception that only those electrons striking the surface with X-directed velocities in excess of

$$v_{xB} \equiv \left( \frac{2eE_B}{m} \right)^{\tfrac{1}{2}}$$

(5-34)

can escape, since only these will possess sufficient energy to surmount the surface barrier $E_B$.  The solution to the present problem is given by Eq. (5-33) except that the lower limit involved in the integration must be changed from zero to $v_{xB}$.  The result is the thermionic-emission current density $J_{th}$, at the temperature $T$.  Thus,

$$
\begin{aligned}
J_{th} &= e \int_{v_{xB}}^{\infty} dN_x' \\
&= \frac{2em^3}{h^3} \int_{v_x = v_{xB}}^{\infty} \int_{v_y = -\infty}^{\infty} \int_{v_z = -\infty}^{\infty} \frac{1}{1 + \epsilon^{\frac{E - E_M}{E_T}}} v_x \, dv_x \, dv_y \, dv_z
\end{aligned}
$$

(5-35)

where Eqs. (5-33), (5-31), (5-16), and (5-18) have been used.

The integrations involved in this evaluation are extremely complicated. However, if the "unity" in the denominator of the integrand could be neglected in comparison with the exponential, then the integral would reduce to the product of three simple integrals. Fortunately, very little error is introduced by neglecting this term. This step will be justified later.

The exponential term that now appears in the numerator of the integrand is simply

$$\epsilon^{\frac{E_M - E}{E_T}} = \epsilon^{\frac{E_M}{E_T}} \epsilon^{-\frac{1}{2}\frac{m}{eE_T}(v_x{}^2 + v_y{}^2 + v_z{}^2)}$$

by virtue of Eq. (5-19). Equation (5-35) reduces, under these conditions, to the product of three integrals, *viz.*,

$$J_{th} = \left(\frac{2em^3}{h^3} \epsilon^{\frac{E_M}{E_T}}\right)\left(\int_{v_{xB}}^{\infty} v_x \epsilon^{-\frac{1}{2}\frac{mv_x{}^2}{eE_T}}\, dv_x\right)\left(\int_{-\infty}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv_y{}^2}{eE_T}}\, dv_y\right)\left(\int_{-\infty}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv_z{}^2}{eE_T}}\, dv_z\right)$$

(5-36)

The integrals in $v_y$ and $v_z$ are standard-form integrals, the solutions of which are given in Appendix V. Each has the value

$$\int_{-\infty}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv^2}{eE_T}}\, dv = 2\int_{0}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv^2}{eE_T}}\, dv = \left(\frac{2\pi eE_T}{m}\right)^{\frac{1}{2}}$$

(5-37)

The integral in $v_x$ is readily evaluated by introducing a new variable $\xi \equiv v_x{}^2$, whence $v_x\, dv_x = d\xi/2$. The result is

$$\int_{v_{xB}}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv_x{}^2}{eE_T}} v_x\, dv_x = \frac{eE_T}{m} \epsilon^{-\frac{1}{2}\frac{mv_{xB}{}^2}{eE_T}}$$

$$= \frac{eE_T}{m} \epsilon^{-\frac{E_B}{E_T}}$$

(5-38)

where Eq. (5-34) has been used. By combining Eqs. (5-38) and (5-37) with Eq. (5-36), there results

$$J_{th} = \left(\frac{2em^3}{h^3}\right)\left(\frac{2\pi eE_T}{m}\right)\left(\frac{eE_T}{m}\right)\epsilon^{(E_M - E_B)/E_T}$$

By writing $E_W = E_B - E_M$ and noting that $eE_T \equiv kT$, this equation reduces to

$$J_{th} = A_0 T^2 \epsilon^{-E_W/E_T} \qquad \text{or} \qquad J_{th} = A_0 T^2 \epsilon^{-b_0/T} \qquad \text{amp/m}^2 \quad (5\text{-}39)$$

where

and

$$A_0 = \frac{4\pi mek^2}{h^3} = 120 \times 10^4 \text{ amp/(m}^2)(^\circ\text{K}^2)$$

$$b_0 = \frac{eE_W}{k} = 11{,}600 E_W \qquad ^\circ\text{K}$$

(5-40)

If $S$ denotes the area of the filament in square meters, then the total thermionic-emission current is

$$I_{th} = SJ_{th} \qquad \text{amp}$$

Either form of Eqs. (5-39) is known as the "thermionic-emission equation," the "Dushman equation," or the "Richardson equation."[8]   The constant $E_W$, called the "work function," was introduced in Sec. 4-7 and discussed in connection with thermionic emission in Sec. 4-8.

In order to justify the approximation made in carrying out the integrals in Eq. (5-35), all that is necessary is to observe that the *smallest* value of the exponential term is obtained by setting $v_y = v_z = 0$ and $v_x = v_{xB}$. This is equivalent to letting $E = E_B$.  The minimum value of this term is then

$$\epsilon^{(E_B - E_M)/\epsilon T} = \epsilon^{E_W/E_T} \tag{5-41}$$

For tungsten, $E_W = 4.52$, and at $T = 3000°K$, $E_T = 0.259$.   Under these conditions, Eq. (5-41) becomes at least equal to

$$\epsilon^{17.5} = 40 \text{ million}$$

Certainly the error made by neglecting unity in comparison with 40 million cannot be very great!

**5-14. Thermionic-emission Constants.**   Table 5-2 contains the results of measurements of $A_0$, $b_0$, and $E_W$ for some common metals.   A substantial difference is noted in some cases between the corresponding values contained in columns 4 and 5.   These differences are an indication of the diffi-

TABLE 5-2

THERMIONIC-EMISSION CONSTANTS

| Element | $A_0 \times 10^{-4}$ amp/(m²) (°K²) | $b_0$, deg | $E_W$,* volts | $E_W$,† volts |
|---------|------|--------|-------|-------|
| C | 30 | 50,300 | 4.34 ‡ | 4.7 |
| Ca | 60.2 | 26,000 | 2.24 | 3.2 |
| Cs | 162 | 21,000 | 1.81 | 1.8 |
| Mo | 60.2 | 50,900 | 4.38 | 4.3 |
| Ni | 26.8 | 32,100 | 2.77 | 5.0 |
| Pt | 32 | 61,700 | 5.32 § | 6.0 |
| Ta | 60.2 | 47,200 | 4.07 | 4.1 |
| Th | 60.2 | 38,900 | 3.35 | 3.4 |
| W | 60.2 | 52,400 | 4.52 | 4.52 |

* S. Dushman, *Revs. Modern Phys.*, **2**, 381, 1930.

† J. A. Becker, *ibid.*, **7**, 95, 1935.

‡ A. L. Reimann, *Proc. Phys. Soc. (London)*, **50**, 496, 1938.

§ L. V. Whitney, *Phys. Rev.*, **50**, 1154, 1936.

culties in making such measurements, and the values in the table must be accepted with some reservations. References to the original literature reporting the data included in the table will be found in the reviews by Dushman and Becker, except for those values which are separately marked.

The parameter $A_0$ which appears in Eq. (5-40) contains only universal constants and ought to be the same for all pure metals. Table 5-2 reveals that this is not found to be true experimentally. It is observed that the value obtained most frequently is $60 \times 10^4$ amp/$(m^2)(\deg^2)$, which is one-half of that predicted theoretically. There are at least three reasons [9] for the differences between the experimental and theoretical values of $A_0$.

1. The work function must depend to a slight extent upon temperature. This follows from the fact that the ionic spacing in a metal changes with temperature. Problem 5-17 illustrates how the value of $A_0$ is modified by a temperature-dependent work function.

2. In the discussion of the potential energy within and at the surface of a metal in Sec. 4-7 it was pointed out that this function depended upon the position of the electron relative to the ions. Hence, the emission should be different from different parts of the same crystal face and should depend upon how close the electron comes to a surface ion. That different crystal planes have different values of work function has been demonstrated experimentally.[10] For tungsten the spread is 0.3 ev among the different faces.

The real surface of a metal consists of a large number of crystallographic planes which are oriented at various angles with respect to one another. The size and shape of these planes, or *patches* as they are called, depend upon the material, the crystal size, and the degree of heat-treatment. The effect of surface patches on thermionic emission, accelerating and retarding field measurements, contact potential, etc., has received detailed study in a review by Herring and Nichols.[10]

Some work has been done in an effort to determine the ratio of the true area to the apparent area of a gross surface for thermionic emission. Becker [11] believes that the ratio should lie in the range between 1 and 10 for rough surfaces, such as those on oxide-coated cathodes, and between 1 and 2 for relatively smooth surfaces such as tungsten, although the exact value will depend on the treatment of each surface. As a result of this factor the measured value of $A_0$ should be different from the theoretical value.

3. Classically all electrons with surface-directed energies greater than $E_B$ will escape from the metal. According to the wave nature of the electron, however, some of these electrons will be reflected back into the metal and will not escape. As a result of this reflection factor the measured value of $A_0$ must always be less than the theoretical value.

Owing to the foregoing factors, it is perhaps more surprising that there appears to be uniformity in the experimental value of $A_0$ for some metals than that differences occur.

**5-15. Energies of Escaping Electrons.** As an illustration of the modifications that must be made in Eqs. (5-25) and (5-26) for calculating average quantities so as to include only those particles which succeed in escaping from a metal, let us calculate the average energy of the electrons that leave the metal under the conditions of thermionic emission. It will be found more convenient to evaluate this term as a function of velocity, according to the expression

$$eE = \tfrac{1}{2}mv^2$$

The average energy of those electrons in a metal which are capable of escaping is

$$E = \frac{\displaystyle\int_{v_x=v_{xB}}^{\infty}\int_{v_y=-\infty}^{\infty}\int_{v_z=-\infty}^{\infty}\frac{1}{2}\frac{m}{e}(v_x{}^2+v_y{}^2+v_z{}^2)\,dN_\tau{}'}{\displaystyle\int_{v_x=v_{xB}}^{\infty}\int_{v_y=-\infty}^{\infty}\int_{v_z=-\infty}^{\infty}dN_\tau{}'} \tag{5-42}$$

The lower limit of integration for $v_x$ must be $v_{xB}$ since only those electrons with $X$-directed velocities greater than this value can escape.

From Eqs. (5-30) and (5-18),

$$dN_\tau{}' = v_x\,dN_\tau = \frac{2m^3}{h^3}\frac{v_x\,dv_x\,dv_y\,dv_z}{1+\epsilon^{(E-E_M)/E_T}}$$

Remembering that the minimum value of $E$ is $E_B$, the approximation made in Sec. 5-13 is again valid; *viz.*, the unity may be neglected in comparison with the exponential. Then

$$dN_\tau{}' = \frac{2m^3}{h^3}\,\epsilon^{(-E+E_M)/E_T}v_x\,dv_x\,dv_y\,dv_z$$

$$= \frac{2m^3}{h^3}\,\epsilon^{E_M/E_T}\epsilon^{-\lambda v_x{}^2}\epsilon^{-\lambda v_y{}^2}\epsilon^{-\lambda v_z{}^2}v_x\,dv_x\,dv_y\,dv_z$$

where $\lambda \equiv m/2eE_T$.

Equation (5-42) is the sum of three terms, the first of which is

$$\frac{\displaystyle\iiint\left(\frac{1}{2}\frac{m}{e}v_x{}^2\right)\left(\frac{2m^3}{h^3}\epsilon^{E_M/E_T}\right)\epsilon^{-\lambda v_x{}^2}\epsilon^{-\lambda v_y{}^2}\epsilon^{-\lambda v_z{}^2}v_x\,dv_x\,dv_y\,dv_z}{\displaystyle\iiint\frac{2m^3}{h^3}\epsilon^{E_M/E_T}\epsilon^{-\lambda v_x{}^2}\epsilon^{-\lambda v_y{}^2}\epsilon^{-\lambda v_z{}^2}v_x\,dv_x\,dv_y\,dv_z}$$

where the limits of integration are those given in Eq. (5-42).

The numerator (and also the denominator) can be written as the *product* of three integrals. If the common factors in the numerator and denominator are canceled, the preceding equation reduces to

$$\frac{\displaystyle\int_{v_{xB}}^{\infty} \frac{1}{2}\frac{m}{e} v_x{}^3 \epsilon^{-\lambda v_x{}^2}\, dv_x}{\displaystyle\int_{v_{xB}}^{\infty} v_x \epsilon^{-\lambda v_x{}^2}\, dv_x}$$

This is integrated in the manner indicated in the preceding section. The result of this integration may be shown to be $E_T + E_B$ electron volts. The second and third terms in Eq. (5-42) are carried out in a similar manner and yield $\frac{1}{2}E_T$ electron volts for either the $Y$- or the $Z$-directed average energy of those electrons which are capable of escaping.

In leaving the metal, the average $X$-directed energy of each electron is reduced by an amount $E_B$, so that the average $X$-directed energy of each electron *after it has left the metal* is $(E_B + E_T) - E_B = E_T$ electron volts. Hence the total average energy of each electron that *escapes from the metal* is $E_T + \frac{1}{2}E_T + \frac{1}{2}E_T = 2E_T$ electron volts.

**5-16. Distribution Function of Escaping Electrons.** Once the electrons have left the metal, they no longer obey the FDS statistics. In order to find the appropriate form of the distribution function for the electrons outside the metal, the procedure is the following: Let the subscript $i$ denote the conditions on the inside of the metal and the subscript $o$ denote the conditions in the region outside the metal. If the $X$ direction is chosen normal to the surface, then the following relations are evidently true:

so that

$$\left.\begin{aligned} \tfrac{1}{2}mv_{xi}{}^2 &= \tfrac{1}{2}mv_{xo}{}^2 + eE_B \\ v_{yi} &= v_{yo} \qquad v_{zi} = v_{zo} \\ E_i &= E_o + E_B \end{aligned}\right\} \tag{5-43}$$

These equations simply state that an electron, in escaping from a metal, has the same $Y$- and $Z$-directed velocities, whereas its $X$-directed energy decreases by an amount equal to the height of the potential-energy barrier at the surface.

By differentiation, the first of Eqs. (5-43) yields

$$v_{xi}\, dv_{xi} = v_{xo}\, dv_{xo} \tag{5-44}$$

However, the number of electrons in the range $d\tau$ striking unit area of the surface in unit time is, from Eqs. (5-30), (5-16) and (5-18),

$$dN_{\tau i}' = v_{xi}\, dN_{\tau i} = \frac{2m^3}{h^3}\frac{1}{1 + e^{\frac{E_i - E_M}{E_T}}} v_{xi}\, dv_{xi}\, dv_{yi}\, dv_{zi} \tag{5-45}$$

where the subscript $i$ has been included to specify that these conditions refer to the interior of the metal.

This expression becomes, by virtue of Eqs. (5-43) and (5-44),

$$dN_{ro}' = \frac{2m^3}{h^3} \frac{1}{1 + \epsilon^{\frac{E_o + E_B - E_M}{E_T}}} v_{xo}\, dv_{xo}\, dv_{yo}\, dv_{zo} \qquad (5\text{-}46)$$

where the subscripts $o$ denote that the distribution function refers to the region outside the metal. This expression may be simplified by following the same general procedure as that adopted in the development of the thermionic-emission equation. The exponential term becomes $\epsilon^{(E_o + E_W)/E_T}$, since $E_W = E_B - E_M$. The smallest value of this term, which results when $E_o = 0$, is again about 40 million for tungsten. It is again justifiable to neglect the unity in the denominator of Eq. (5-46). By doing this and dropping the subscripts $o$ for the sake of simplicity, there results

$$dN_r' = a'\epsilon^{-E/E_T} v_x\, dv_x\, dv_y\, dv_z \qquad (5\text{-}47)$$

where $a'$ is a constant given by $(2m^3/h^3)\epsilon^{-E_W/E_T}$ and where, as usual, $eE = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)$.

More will be said about this distribution function $dN_r'$ in Chap. 8 in connection with the discussion of the kinetic theory of gases. It is known as the *Maxwell-Boltzmann distribution function*. It thus appears that those electrons which have escaped from a metal obey the Maxwell-Boltzmann distribution laws in the region outside of the metal, although they obey the FDS distribution law within the metal.
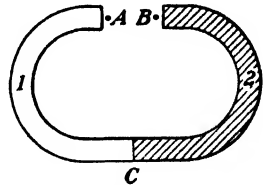
**5-17. Contact Potential.**[12] Consider two metals in contact with each other, as at the junction $C$ in Fig. 5-9. The contact difference of potential between these two metals is defined as the potential difference $E_{AB}$, between a point $A$ just outside metal 1 and a point $B$ just outside metal 2. In Sec. 4-9 it was stated that the



FIG. 5-9. Two metals in contact at the junction $C$.

Fermi characteristic energies of two metals in contact were at the same level. This will now be justified.

In all the equations that were previously under survey, only a single metal was involved, so that the zero reference of potential could be chosen arbitrarily. It was found convenient to choose this reference in such a way that $E$ represented the kinetic energy of the particle. Since two or more metals may be involved, then the choice of the zero level of only one of these is arbitrary. That is, once the zero level of metal 1 is fixed, as it will be in the present circumstances, then the zero level of metal 2 cannot like-

wise be arbitrary. If the difference between the zero levels of the two metals be denoted by $E'$, then the problem at hand is to determine $E'$.

Since the two metals are assumed to be in thermal equilibrium at the junction $C$, there must be as many electrons passing from metal 1 to metal 2 across this junction as there are in the reverse direction in each velocity range. According to Eq. (5-30) the number of electrons with velocities in the range $dv_x \, dv_y \, dv_z$ striking unit area in unit time is

$$dN_{r1}' = v_{x1} \, dN_{r1} = \frac{2m^3}{h^3} \frac{v_{x1} \, dv_{x1} \, dv_{y1} \, dv_{z1}}{1 + \epsilon^{(E_1 - E_{M1})/E_T}} \tag{5-48}$$

where the subscript 1 denotes conditions in metal 1. Similarly, the number of electrons in the velocity range $dv_x \, dv_y \, dv_z$ striking unit area in unit time of metal 2 is

$$dN_{r2}' = \frac{2m^3}{h^3} \frac{v_{x2} \, dv_{x2} \, dv_{y2} \, dv_{z2}}{1 + \epsilon^{(E_2 - E_{M2})/E_T}} \tag{5-49}$$

If the $X$ direction is chosen normal to the surface of contact $C$, then the following relations are evidently true:

so that
$$\left.\begin{array}{c} \frac{1}{2}mv_{x1}^2 = \frac{1}{2}mv_{x2}^2 + eE' \\ v_{y1} = v_{y2} \qquad v_{z1} = v_{z2} \\ E_1 = E_2 + E' \end{array}\right\} \tag{5-50}$$

These equations simply state that an electron in passing across the junction $C$ from metal 1 to metal 2 has its $X$-directed energy decreased by an amount $E'$, whereas its $Y$- and its $Z$-directed velocity components remain unchanged. The similarity between these equations and those used in the last section [Eqs. (5-43)] should be noted.

It follows from the first of Eqs. (5-50) that $v_{x1} \, dv_{x1} = v_{x2} \, dv_{x2}$. Hence Eq. (5-48) becomes

$$dN_{r1}' = \frac{2m^3}{h^3} \frac{v_{x2} \, dv_{x2} \, dv_{y2} \, dv_{z2}}{1 + \epsilon^{(E_2 + E' - E_{M1})/E_T}} \tag{5-51}$$

For equilibrium to exist there must be as many electrons (in each velocity range) passing across the junction $C$ from metal 1 to metal 2 as in the reverse direction in unit time, then $dN_{r1}' = dN_{r2}'$. This can be true only if

$$E_2 + E' - E_{M1} = E_2 - E_{M2}$$

This follows from Eqs. (5-49) and (5-51). Consequently

$$E' = E_{M1} - E_{M2} \tag{5-52}$$

This equation states that an electron will encounter a barrier $E_{M1} - E_{M2}$ in passing from metal 1 to metal 2 across the junction $C$.

In view of this result, the potential-energy diagram of two metals in contact at a junction may be drawn. This is done in Fig. 5-10, which is identical with Fig. 4-15 and is repeated here for convenience. Since the difference in the zero levels of the two metals is $E_{M1} - E_{M2}$ *the Fermi levels of the two metals are located at the same heights.* It is seen that a·potential barrier equal to

$$E_{AB} = E_{W2} - E_{W1} \tag{5-53}$$

exists between the two metals. In other words, *the contact difference of potential of two metals equals the difference between their work functions.* This result has been verified experimentally by numerous investigators.[13]
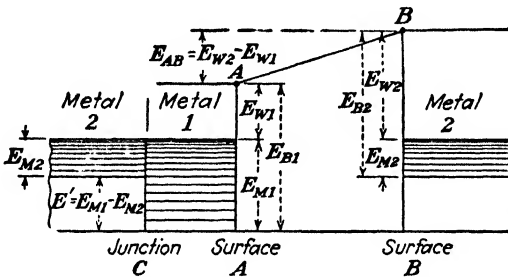


FIG. 5-10.  The potential-energy system of two metals in contact.

In Fig. 5-10, metal 2 (which has the greater work function) has a negative surface charge. That is, work must be done to take an electron from a point just outside of metal 1 (the point $A$) to a point just outside of metal 2 (the point $B$) against the repelling action of this charge. This is clearly indicated by the potential-energy barrier between points $A$ and $B$.

If metals 1 and 2 are joined together through a third metal instead of being in direct contact at the point $C$, then Fig. 5-10 will not be altered in so far as conditions in the interspace between points $A$ and $B$ are concerned. It is merely necessary to insert on the diagram the energy levels of metal 3 at the junction $C$ in such a manner that its Fermi energy is at the same vertical level as $E_{M1}$ and $E_{M2}$. This result can be generalized to include any number of metals. Hence it can be concluded that, if any number of metals are connected in series, the contact potential between the first and the last metal equals the difference of the work functions of these two metals and is independent of the interconnecting metals.

Suppose that metals 1 and 2 are joined together by a battery $E_b$, with its negative terminal connected to metal 2. The potential energy of the entire system will then be as shown in Fig. 5-11. It is obtained from Fig. 5-10 by raising all the energy levels of metal 2 by an amount equal to $E_b$. If the positive terminal of the battery were connected to metal 2,

then all its levels would be lowered by an amount corresponding to $E_b$. With the polarity as indicated in Fig. 5-11, the barrier between the two metals is increased by the amount $E_b$, indicating that now this additional work must be done in order to take an electron from a point just outside of metal 1 to a point just outside of metal 2. This is to be expected since metal 2 is made negative with respect to metal 1.

The contact difference of potential can be measured indirectly in a number of ways. One indirect method has already been discussed in Sec. 4-10. Another, known as the *Kelvin method*,[14] is a direct one. It is based upon
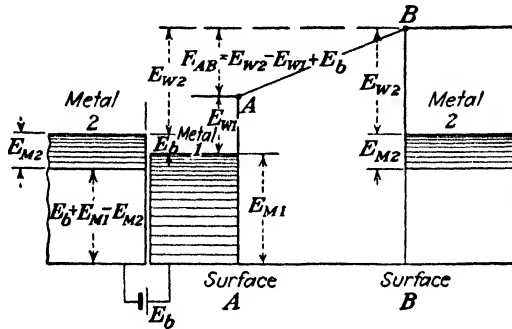


FIG. 5-11. The potential-energy system of two metals joined together through a battery $E_b$. This is obtained from Fig. 5-10 by raising all the levels in metal 2 by an amount $E_b$.

the concept of a capacitor as a system of two metals that are separated by a dielectric. If the distance between the two plates is varied, the capacitance of the system changes. Hence, if a contact difference of potential exists between the two plates, then a varying charge will appear on the surface of the plates as the distance is varied. If these metals are connected together through a high resistance, a current will pass through it and the voltage drop across the resistor can be detected with the aid of suitable indicating instruments.

One practical apparatus [15] for making contact-potential measurements consists of two parallel filaments mounted in a vacuum chamber, the material of the filaments being those whose contact difference of potential is to be measured. An electromagnet is provided to maintain a steady magnetic field perpendicular to the plane of the filaments. One filament is excited from an a-c source, the frequency of which is in the audible range. This results in setting the filament into vibration. An adjustable source of potential is applied in series with a resistor between the two filaments. The potential is adjusted until the output from the amplifier that measures the voltage across the resistor is zero. Under these conditions the applied voltage is equal to and of opposite polarity to the contact potential. This follows from the fact that, if the net voltage between the filaments is zero,

there will be no current through the resistor as the capacitance of the system is varied.

**5-18. Retarding Fields.** Suppose that a collecting electrode is placed parallel to a plane metallic surface which is maintained at a temperature of $T°K$. Suppose also that a potential $E_r$ is applied between these two metallic surfaces in such a direction as to repel the electrons that tend to reach the collecting electrode. How many electrons will reach the collector? Obviously, only those electrons will be collected whose initial energies are sufficient to overcome the effect of this retarding field.
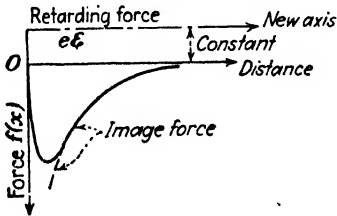


FIG. 5-12. The effect of a retarding electric-field intensity $\varepsilon$ at the surface of a metal is to increase the retarding force on an escaping electron by an amount $e\varepsilon$. (This figure is not drawn to scale.)
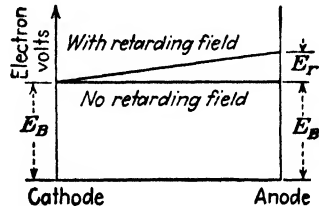
FIG. 5-13. The potential energy between plane-parallel electrodes with and without a retarding field. (The material of the cathode and anode is assumed to be the same so that the contact potential is zero.)

In order to obtain quantitative results, the problem is restated in a more formal manner: How many electrons per second are emitted from the metallic surface with $X$-directed energies greater than that given by $\frac{1}{2}mv_x^2 = eE_r$ (because it will be only these electrons which will be able to move against the retarding potential $E_r$ and reach the collector)? A little thought will show that the solution to this problem is obtained simply by replacing $E_W$ in the Dushman equation by $E_W + E_r$. Hence

$$I = A_0 S T^2 \epsilon^{-(E_W + E_r)/ET} \qquad \text{amp}$$

or, by Eq. (5-39),

$$I = I_{th}\epsilon^{-E_r/ET} = I_{th}\epsilon^{-11,600E_r/T} \qquad \text{amp} \qquad (5\text{-}54)$$

where $S$ is the area of the emitter, $I$ is the current collected for a retarding potential $E_r$, $I_{th}$ is the thermionic-emission current that would be collected when the retarding potential is zero, and $E_r$ is the magnitude of the retarding potential.

That is, the effect of the retarding field may be considered as an effective increase in height of the potential-energy barrier at the surface of the metal. This effect is illustrated graphically in Figs. 5-12 and 5-13. Figure 5-12 represents the force function in the immediate neighborhood of

the cathode, whereas Fig. 5-13 represents the potential energy in the region between the emitter and the collector. Although the potential energy is the integral of the force function, it must be remembered, in comparing the two diagrams, that Fig. 5-12 extends over a range of about $10^{-7}$ m, whereas Fig. 5-13 represents the entire distance from cathode to anode, which is usually of the order of $10^{-2}$ m. The sharp rise at the surface of the cathode represents the barrier energy $E_B$ and equals the area under the curve of Fig. 5-12 with respect to the horizontal axis through $O$. The linear rise in potential energy between the electrodes represents the integral of the constant retarding force.

For the case when the system possesses cylindrical symmetry, but in which the collecting plate radius $r_a$ is considerably larger than the cathode radius $r_k$, Eq. (5-54) is no longer valid and must be replaced, as was first shown by Schottky,[16] by the expression

$$ I = \frac{2}{\sqrt{\pi}} I_{th} \left( \sqrt{\frac{E_r}{E_T}} \, \epsilon^{-E_r/E_T} + \int_{\sqrt{E_r/E_T}}^{\infty} \epsilon^{-x^2} \, dx \right) \qquad (5\text{-}55) $$

where $I_{th}$ is the thermionic-emission current and $E_r$ is the magnitude of the true retarding potential.

The essential differences in the physical arguments that lead to Eq. (5-55) instead of to Eq. (5-54) lie in the fact that some electrons with escaping energies less than $E_r$ normal to the surface of the cathode may reach the plate. For plane electrodes, the components of velocity possessed by the electron not in the normal-to-the-surface direction remains unchanged. Hence, only electrons with energies greater than $E_r$ in the normal-to-the-surface direction will reach the plate. However, an electron moving in a tube possessing cylindrical symmetry has, in addition to its radial component of velocity, a tangential component of velocity. Both of these components depend upon the position of the electron between the cathode and the anode. The tangential velocity decreases as the electron moves from cathode to anode, and the normal velocity is increased over what it would be for plane-parallel electrodes. Hence, more electrons can surmount the retarding potential barrier for a cylindrical system than for the plane-parallel system. Thus, Eq. (5-55) represents a greater current than Eq. (5-54), at the same values of $E_r$ and $I_{th}$.



FIG. 5-14. Tangential and radial coordinates at the anode and at the cathode of a cylindrical system.

The above statements are justified by referring to Fig. 5-14. Let $v_x$ and $v_y$ represent the normal and tangential components of velocity, re-

spectively, of an electron *after leaving* the cathode. Similarly, let $v_x'$ and $v_y'$ represent the normal and tangential components of velocity, respectively, of the electron when it gets to the anode. The conservation of energy is expressed by

$$\tfrac{1}{2}mv_x^2 + \tfrac{1}{2}mv_y^2 = \tfrac{1}{2}m(v_x')^2 + \tfrac{1}{2}m(v_y')^2 + eE_r \qquad (5\text{-}56)$$

because the $Z$ component of velocity remains unchanged.

The conservation of momentum is expressed by

$$mv_y r_k = mv_y' r_a$$

or

$$v_y' = \frac{r_k}{r_a} v_y \qquad (5\text{-}57)$$

where $r_k$ and $r_a$ are the radii at the cathode and anode, respectively. This equation justifies the statement made in the preceding paragraph that the tangential component of velocity is less at the anode than at the cathode.

Substituting Eq. (5-57) into Eq. (5-56) yields

$$(v_x')^2 = v_x^2 + v_y^2\left(1 - \frac{r_k^2}{r_a^2}\right) - \frac{2eE_r}{m} \qquad (5\text{-}58)$$

or, assuming that $r_k \ll r_a$,

$$(v_x')^2 = v_x^2 + v_y^2 - \frac{2eE_r}{m} \qquad (5\text{-}59)$$

For the electron to be collected, $v_x'$ must be real, or $(v_x')^2 \geqslant 0$, or

$$v_x^2 + v_y^2 \geqslant \frac{2eE_r}{m} \qquad (5\text{-}60)$$

This condition is illustrated in Fig. 5-15. Electrons in the shaded portion of velocity space are collected. Note that $v_x$ must be positive since only these electrons leave the cathode. For plane-parallel electrodes only the electrons with velocity components to the right of the dashed vertical line are collected, and hence a smaller current is obtained than for the cylindrical case.

If the distribution function of the escaping electrons [Eq. (5-47)] is integrated between the limits indicated in Fig. 5-15 for $v_x$ and $v_y$ and from $-\infty$ to $+\infty$ for $v_z$, the result is that given in Eq. (5-55) (see Prob. 5-25).

The verification of Eq. (5-55) is somewhat more difficult than the simpler form (5-54). Germer [17] has performed experiments that show the validity of Eq. (5-55). This verification is accomplished with the aid of the data in Table 5-3. The data obtained in the laboratory are plotted in semilogarithmic fashion in the manner shown in Fig. 4-17. The resulting

curve is no longer linear, theoretically, although an examination of the data in Table 5-3 shows that for values of $E_r/E_T$ greater than about 3 the curve should be essentially linear. Moreover, Eq. (5-55) indicates that
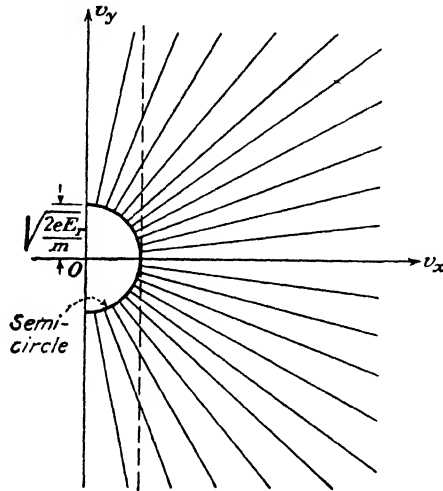


Fig. 5-15. The region of velocity space from which electrons are collected if a retarding voltage $E_r$ exists between cathode and anode of a cylindrical system.

the slope of the $\log_{10} I$ vs. $E_r/E_T$ curve approaches the constant value 0.434 ($= \log_{10} \epsilon$). Hence, if $\log_{10} I$ is plotted vs. $E_b$, the slope of the curve for large values of $E_r$ gives $0.434/E_T$, from which the temperature may be determined. It must be remembered that the retarding voltage

TABLE 5-3

RETARDING EQUATION VALUES

| $E_r/E_T$ | $\log I_{th}/I$ | $E_r/E_T$ | $\log I_{th}/I$ |
|-----------|------------------|-----------|------------------|
| 1 | 0.2423 | 10 | 3.7698 |
| 2 | 0.5827 | 11 | 4.1850 |
| 3 | 0.9523 | 12 | 4.6024 |
| 4 | 1.3371 | 14 | 5.4398 |
| 5 | 1.7312 | 16 | 6.2812 |
| 6 | 2.1318 | 18 | 7.1245 |
| 7 | 2.5369 | 20 | 7.7914 |
| 8 | 2.9455 | 25 | 10.0978 |
| 9 | 3.3567 | | |

$E_r$ is the applied voltage $E_b$ minus the contact potential. Germer found that the temperature obtained in this way was in agreement with the value determined by direct measurement, thereby showing that the emitted electrons were in temperature equilibrium with the cathode.

If the contact potential can be determined, then the experimental points can be compared with the theoretical ones contained in Table 5-3. If space charge exists, thereby masking the sharp break in the logarithmic plot from which the contact potential may be found, then the point at which $I = I_{th}$ is taken as the true zero of potential. This requires a knowledge of the dimensions and the temperature of the filament in order to calculate $I_{th}$ from the thermionic-emission equation.

The experimental technique necessary in order to carry out experiments of this type using directly heated filaments is rather delicate and special. A real appreciation of the difficulties involved may be had by referring to Germer's paper. The results of this experiment prove that the electrons emitted thermionically from a heated filament satisfy the Maxwell-Boltzmann distribution function [since Eq. (5-55) is derived on the assumption that the electrons satisfy this distribution function] and also that the emitted electrons are in temperature equilibrium with the cathode.

The investigations of Germer were made on pure tungsten filaments operating in the range from 1440° to 2475°K. Other investigators have considered the problem. Nottingham [18] investigated the matter from a study of the emission from thoriated tungsten filaments. Although the results he obtained are in general agreement with the theoretical expectations, the calculated electron temperatures are about 1.5 times the measured ones. Fan [19] has obtained good agreement with theory using an oxide-coated cathode.

**5-19. Accelerating Fields.** Assume that an accelerating rather than retarding field is applied between the cathode and the collecting anode. Because of the electric field $\mathcal{E}$ which acts in a direction to aid the electron to leave the surface of the metal, the shape of the force function and, consequently, the shape of the potential-energy function will be altered. It is noted that two forces act on the electron, the image force which tends to attract the electron to the cathode surface and the applied field which acts to draw the electron away from the surface of the metal. Consequently, a critical distance $x_c$ exists at which the total force on the electron is zero. The critical distance is the point where $f(x)$, the image force, just equals the applied force $e\mathcal{E}$. For all points for which $x > x_c$, the net force acting on the electron is in the escape direction. Thus, if the electron succeeds in reaching the point $x = x_c$, it will be free of the metal.

The net force on the electron, at any distance from the surface, is the sum of the surface-attraction force $f(x)$ and the force of the applied field $e\mathcal{E}$, as illustrated in Fig. 5-16. Since the entire diagram covers a range of

only $10^{-6}$ or $10^{-7}$ m, the external force $\mathcal{E}$ may be considered constant over this range, and the net force is that shown in Fig. 4-12, with the reference axis shifted downward by the constant amount $e\mathcal{E}$, as shown by the dotted axis of Fig. 5-16.

As before, the area under the force curve of Fig. 5-16 must give the potential energy possessed by an escaping electron at any distance $x$ from



FIG. 5-16. The effect of an accelerating electric-field intensity $\mathcal{E}$ at the surface of a metal is to decrease the retarding force on an escaping electron by an amount $e\mathcal{E}$. At larger distances than the "escape distance" $x_c$ the net force acting on the electron is accelerating rather than retarding.

the surface of the metal. The potential-energy curve corresponding to the force curve of Fig. 5-16 is illustrated in Fig. 5-17. Each ordinate of the curve labeled "Resultant" is the algebraic sum of the corresponding ordinates of the potential-energy curves for zero field and for the accelerating field due to $\mathcal{E}$.

The density function also illustrated in this diagram indicates the position of the normal maximum energy level $E_M$ at 0°K. The corresponding density function at a high temperature $T$ (considerably distorted for purposes of clarity)

is also shown. At the temperature $T$°K and a zero applied external field, only those electrons having energies greater than $E_B$ can escape from the metal. These are represented by the heavy shading of the appropriate portion of the density function $\rho_E$. Owing to the presence of an applied
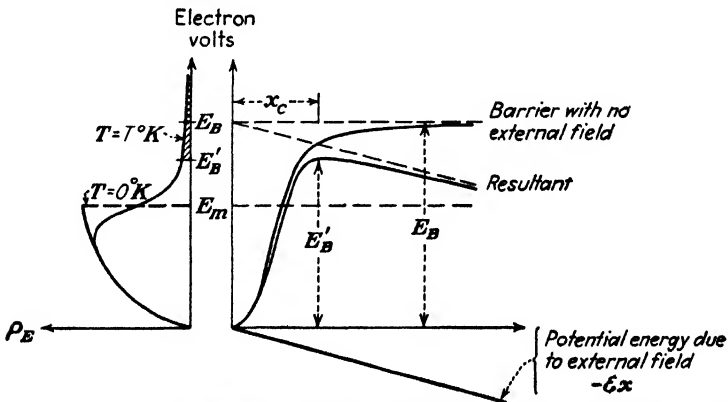


FIG. 5-17. The potential-energy curves with and without an external accelerating field corresponding to the force curves of Fig. 5-16. The peak of the potential barrier $E_B'$ (with an accelerating field) is reached at the "escape distance" $x_c$.

accelerating field, the potential-energy curve is lowered. Under these circumstances, the number of electrons that escape from the metal will be greater than before, since now an electron must possess an energy greater than $E_B{}'$, a quantity that is smaller than $E_B$, in order to escape. This effect is indicated by the lighter shading of the density function.

Under the present circumstances, the area under the force curve, which equals the barrier work function, is given by the expression

$$eE_B{}' = \int_0^{x_c} [-f(x) - e\mathcal{E}] \, dx \tag{5-61}$$

The integration does not extend to infinity but extends only to the point $x_c$. This follows from the fact that the only work that must be done in order to remove the electron from the influence of the attraction of the surface is that necessary to carry it to the point $x_c$, since for all distances beyond this point the image attraction is less than the externally applied force. This result is evident from Fig. 5-17. Thus it follows that

$$eE_B{}' = -\int_0^{x_c} f(x) \, dx - \int_0^{x_c} e\mathcal{E} \, dx$$

which may be written as

$$eE_B{}' = -\int_0^{\infty} f(x) \, dx - \int_{\infty}^{x_c} f(x) \, dx - \int_0^{x_c} e\mathcal{E} \, dx$$

since

$$\int_0^{x_c} f(x) \, dx = \int_0^{\infty} f(x) \, dx + \int_{\infty}^{x_c} f(x) \, dx$$

This reduces to

$$eE_B{}' = eE_B + \int_{x_c}^{\infty} f(x) \, dx - \int_0^{x_c} e\mathcal{E} \, dx \tag{5-62}$$

where use has been made of Eq. (4-8), which defines the term $E_B$. However, in the interval from $x_c$ to $\infty$ the function $f(x)$ is the image-force function $-e^2/16\pi\epsilon_0 x^2$ (see Sec. 4-7); and in the interval 0 to $x_c$ the applied accelerating force $\mathcal{E}$ is constant. By using these facts in Eq. (5-62), there results

$$E_B{}' = E_B - \frac{e}{16\pi\epsilon_0 x_c} - \mathcal{E}x_c \tag{5-63}$$

This equation indicates that the effect of the applied field is to reduce the work function at the surface of the metal, as expected.

Since the sum of the forces acting on the electron at the point $x_c$ is zero, then

$$e\mathcal{E} = \frac{e^2}{16\pi\epsilon_0 x_c{}^2} \quad \text{or} \quad \mathcal{E}x_c = \frac{e}{16\pi\epsilon_0 x_c} \tag{5-64}$$

This shows that the second term is equal to the third term in Eq. (5-63). Hence

$$E_B' = E_B - 2\mathcal{E}x_c \tag{5-65}$$

From Eq. (5-64)

$$x_c = \left(\frac{e}{16\pi\epsilon_0\mathcal{E}}\right)^{\frac{1}{2}} = 1.90 \times 10^{-5}\mathcal{E}^{-\frac{1}{2}} \qquad \text{m} \tag{5-66}$$

where $\mathcal{E}$ is in volts per meter.

Using this expression in Eq. (5-65), there results

$$\begin{aligned}
E_B' &= E_B - \left(\frac{e}{4\pi\epsilon_0}\right)^{\frac{1}{2}}\mathcal{E}^{\frac{1}{2}} \\
&= E_B - 9.47 \times 10^4 (e\mathcal{E})^{\frac{1}{2}} \\
&= E_B - 3.79 \times 10^{-5}\mathcal{E}^{\frac{1}{2}} \tag{5-67}
\end{aligned}$$

This theory may be verified experimentally by investigating the effects upon the thermionic-emission current of a high accelerating field. It is known from Eq. (5-39) that the saturation thermionic current of a cathode maintained at a temperature $T°K$ when the effect due to the accelerating field is negligible is

$$I_{th} = A_0 S T^2 \epsilon^{-(E_B - E_M)/E_T}$$

When the applied field is appreciable, $E_B$ in this expression must be replaced by $E_B'$ and this equation attains the form

$$I = A_0 S T^2 \epsilon^{-\frac{E_B - 9.47 \times 10^4 (e\mathcal{E})^{\frac{1}{2}} - E_M}{E_T}}$$

The ratio of the last two equations yields

$$I = I_{th}\epsilon^{\frac{9.47 \times 10^4 (e\mathcal{E})^{\frac{1}{2}}}{E_T}} = I_{th}\epsilon^{\frac{0.440\mathcal{E}^{\frac{1}{2}}}{T}} \tag{5-68}$$

The logarithm of this equation is

$$\log_{10} I = \log_{10} I_{th} + \frac{0.190\mathcal{E}^{\frac{1}{2}}}{T} \tag{5-69}$$

and predicts that a plot of $\log_{10} I$ vs. $\sqrt{\mathcal{E}}$ should give a straight line having the slope $0.190/T$. This result, which was first predicted by Schottky,[20] has been experimentally verified.[21] Such a plot is given in Fig. 5-18.

Note also that Eq. (5-68) can be rearranged to the following form:

$$\frac{\left(\log_{10}\dfrac{I}{I_{th}}\right)E_T}{\sqrt{\mathcal{E}}} = 0.434 \times 9.47 \times 10^4 e^{\frac{1}{2}} \tag{5-70}$$

DeBruyne,[21] using this equation, obtained a value of $1.61 \times 10^{-19}$ coulomb for $e$. These results appear to furnish ample proof of the validity of the Schottky equation.

Some experiments on tungsten and tantalum have revealed the existence of small periodic deviation from the Schottky line.[22] However, these have been shown [23] to be in agreement with theory if wave mechanics is taken into consideration.

Although the derivation given above is quite independent of the geometry of the electrodes, nevertheless, in order to calculate the electric-field
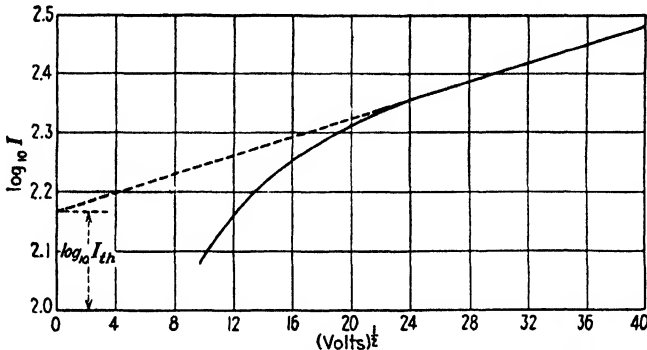


FIG. 5-18.  Experimental verification of the Schottky theory of the effect of an accelerating field at the surface of a metal.

intensity $\varepsilon$ at the surface of the cathode from a knowledge of the applied potential between the cathode and the anode, the shape and the dimensions of the electrodes must be known.

The experimental verification of this theory furnishes proof of the existence of image forces for points that are not too close to the surface of the metal. In fact, this theory actually offers a means of exploring the field at the surface of the metal. Thus, by combining the thermionic-emission equation with Eq. (5-61), there results

$$I = A_0 S T^2 \epsilon^{\frac{E_M}{E_T} + \frac{1}{E_T}\int_0^{x_c}\left[\frac{f(x)}{e} + \varepsilon\right] dx} \tag{5-71}$$

The derivative of the logarithm of this expression with respect to the field intensity $\varepsilon$ yields the expression

$$\frac{d(\log_{10} I)}{d\varepsilon} = \frac{0.434 x_c}{E_T} \tag{5-72}$$

This result is based on the assumption that the applied electric field $\varepsilon$ is constant over the range from 0 to $x_c$, a reasonable assumption when one recalls the minuteness of $x_c$ compared with the interelectrode spacing. If

the current $I$ is measured experimentally as a function of $\mathcal{E}$, then the critical distance $x_c$ corresponding to each value of $\mathcal{E}$ can be calculated from Eq. (5-72). Since, however, the critical distance is the point at which the applied accelerating force just equals the attractive surface force, *i.e.*, the point at which $f(x) = e\mathcal{E}$, then one may, in this way, determine $f(x)$ as a function of distance. As yet no experiments have been performed that make $x_c$ so small as to reveal any serious departures from the image-force law.

**5-20. Barrier Layer Rectification.**[24] A very interesting situation is contained in Sec. 5-17. Notice that the application of the potential $E_b$ (see Fig. 5-11) raises the potential energy of metal 2 upward by an amount corresponding to $E_b$. Moreover, if the polarity of $E_b$ were reversed, the
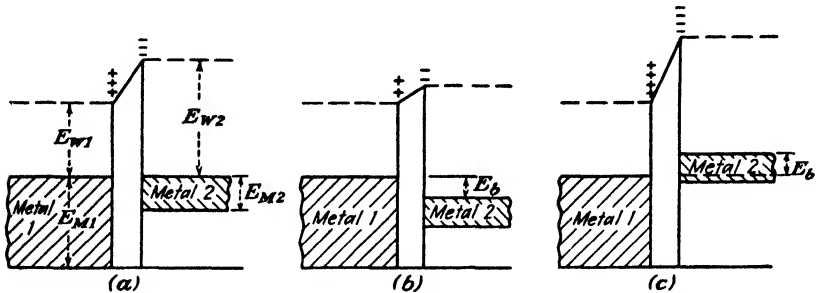


Fig. 5-19. The potential energy system of two metals in contact. (*a*) With no applied voltage (the Fermi levels are on the same line). (*b*) With voltage applied in the forward direction. (*c*) With voltage applied in the backward direction.

potential energy of metal 2 would move downward. The situation, for an applied voltage which is smaller in magnitude than the work function difference of the two metals, is illustrated in Fig. 5-19. In either case, there is no change in the height of the barrier as viewed from the metal of higher work function (metal 2 in this case), but the height of the barrier when viewed from the metal of lower work function varies linearly with applied voltage.

It should be observed from these diagrams that the application of a voltage $E_b$ does not change the current in the direction from metal 2 to metal 1 across the contact or barrier region, because of the constant barrier energy height in this direction. If this current is designated by $I_0$, then the current from metal 1 to metal 2 is also $I_0$ in Fig. 5-19a. The net current across the boundary is zero under these circumstances. It is recalled that it was this equilibrium condition that resulted in the alignment of the two Fermi levels in Sec. 5-17.

In Fig. 5-19b the current from metal 1 to metal 2 is increased and equals $I_0 \epsilon^{E_b/ET}$ because the effective barrier in this direction is decreased by the

height $E_b$. The net current from metal 1 to metal 2 is $I_0(\epsilon^{E_b/E_T} - 1)$. This makes metal 1 the effective cathode and metal 2 the anode. The applied voltage, in this case, is said to be in the "forward" direction. An opposite voltage is said to be in the "reverse" or "blocking" direction. This is depicted in Fig. 5-19c. Under these circumstances the current from metal 1 to metal 2 is lowered and equals $I_0\epsilon^{-E_b/E_T}$. Hence, the net current is in the reverse direction, from metal 2 to metal 1. The magnitude of this current is $I_0(1 - \epsilon^{-E_b/E_T})$. Clearly, the current does not vary linearly with voltage and is much greater in the forward than in the reverse direction. Consequently conduction at such a boundary is nonlinear, and rectification is possible.

Attempts to construct satisfactory metal-insulator-metal rectifiers of the type here being discussed have not been too successful. However, if the insulator and one metal is replaced by a semiconductor, a very important class of rectifier results. These are discussed in some detail in Secs. 5-22 to 5-24.

**5-21. Insulators.** Quantum mechanical theory proves that in a periodic potential field such as exists in a crystal there are bands of allowed energies which may be separated by forbidden energy regions. A metal is characterized by the fact that the band occupied by the "free" electrons is not completely filled and that there are no forbidden levels at higher energies. Recall that in a metal, at absolute zero, the energies of the electrons range from zero to a value $E_M$. If any amount of energy less than $E_W$ is given to an electron it will be raised to a higher level. If the energy is supplied by an externally applied voltage, conduction results and hence this energy band is called the "conduction band." As already discussed electron emission occurs when the energy of an electron is raised to a value in excess of the barrier height $E_B$.
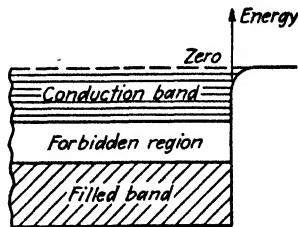


Fig. 5-20. The electronic energy bands of an insulator.

The band structure of an insulator is different from that of a metal. The "valence" electrons completely fill one band. The next empty band of allowed energy states is separated from the filled band by a forbidden region which may be several electron volts high as shown in Fig. 5-20. The energy that can be supplied to an electron from an applied field is generally too small to carry the electron from the filled band into the empty band and hence no conduction takes place.

**5-22. Semiconductors.**[34] It appears reasonable, on the basis of the foregoing, to expect a class of material in which the width of the forbidden region is relatively narrow (of the order of 1 ev). It would be expected that such a material would exhibit insulator properties at the low tem-

peratures, but owing to the changing distribution of electrons with temperature, a limited conduction might occur at the higher temperatures. Materials of this type are called semiconductors, and conduction of the type here discussed is known as *intrinsic* conduction.  Ultrapure germanium or silicon are semiconductors of this type.

Of considerably more importance than intrinsic semiconduction is a second class of semiconductors which are known as *extrinsic* semiconductors.  Such semiconductors also possess a band structure like that illustrated in Fig. 5-20, but they have the added feature that a number of discrete energy states exist in the nominally forbidden region.  In one type,
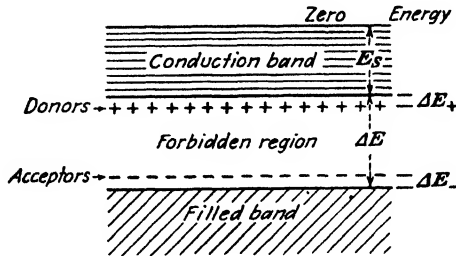


FIG. 5-21.  Energy level diagram of an extrinsic semiconductor with acceptor levels (marked − on the diagram) and with donator levels (marked + on the diagram).

the extra levels exist in the general neighborhood of the filled band.  In another type, the extra levels exist near the empty conduction band. These extra energy levels occur either because of the existence of lattice imperfections, or because of the presence of impurities in the material.

The extra levels may be filled at absolute zero temperature, in which case they will act as *donators* or *donors* at finite temperatures.  In such a case, the electron in the donor level may be donated to the conduction band.  Once the electron is in the conduction band, conduction is then possible in the same way as for the metal.

If the extra levels are empty at absolute zero, they will act as *acceptors* at finite temperatures.  In such a case, the acceptor level may accept an electron from the filled band at finite temperatures, thereby causing an empty level to occur in the band.  Conduction in this empty level is now possible.  It is of interest to note that both intrinsic and extrinsic semiconduction may occur in the same semiconductor.

The energy level diagram of an extrinsic semiconductor with both acceptor levels (marked −) and donator levels (marked +) is given in Fig. 5-21.  The donator levels are found to lie below and near the empty band, and the acceptor levels lie above and near the filled band, somewhat as illustrated in Fig. 5-21.  The energy separation of these levels $\Delta E_-$ and $\Delta E_+$ may be quite small in a given case.  For example, in the case of sili-

con and germanium, the extra levels are only about 0.1 ev from the appropriate band edge.

A semiconductor that conducts principally by electrons in the nearly empty conduction band is called an "*n*-type" (negative) semiconductor, and the process is called "excess" conduction. A semiconductor that conducts principally because of the presence of empty levels in the nearly filled band is called a "*p*-type" (positive) semiconductor and the process is called "defect" or "deficit" conduction.

The above discussion is consistent with the experimentally observed results that the conductivity of semiconductors increases with temperature and decreases with purification. Such devices have a negative temperature coefficient of resistance and are called "thermistors." They find applications in thermometry in measurement of R.F. power, in control devices, etc.

**5-23. Crystal Structure of Silicon and Germanium Semiconductors.** Each atom is situated at the corners of a regular tetrahedron. These atoms are tetravalent and hence possess, in addition to the inner tightly bound electrons, four outermost electrons. The crystal is held together by "electron-pair bonds." Since each atom is surrounded by four nearest neighbors, each of its four valence electrons is shared with one of its neighbors. A schematic diagram, drawn in two dimensions for simplicity, is shown in Fig. 5-22. Each circle marked +4 is intended to represent the nucleus surrounded by its cloud of tightly bound electrons and has a net charge of four positive units. The "electron-pair bonds" between atoms is clearly indicated.



Fig. 5-22. Crystal structure of silicon or germanium.

In this structure every electron is tightly bound. The width of the forbidden band in Fig. 5-20 corresponds to the amount of energy necessary to remove an electron from its bond so that it may move through the lattice as a conduction electron.

Consider that an impurity, such as arsenic, which has five valence electrons is added to the crystal. This atom will replace the tetravalent parent atom and four of its valence electrons will form electron pair bonds with its neighbors. This leaves the fifth valence electron which is free to wander about in the crystal. These electrons account for the donor states in Fig. 5-21. The addition of an impurity such as arsenic, phosphorus, or antimony which has five valence electrons therefore results in an *n*-type semiconductor.

A *p*-type semiconductor results from the addition of an impurity such as boron, or aluminum which is trivalent. The impurity atom replaces

one of the parent atoms in the crystal, but its three valence electrons cannot complete the four electron-pair bonds to its four neighbors. Hence, a "hole" is left in one of these bonds. This can then be filled by an electron from an adjacent atom. When this happens the "hole" moves over to the atom from which the electron came. When the electrons move in one direction the holes wander in the opposite direction and hence they behave as if they were mobile positive charges. These holes account for the acceptor states of Fig. 5-21. It has become customary to ignore completely the presence of the electrons in the filled band of a $p$-type semiconductor and to confine attention to the positively charged holes.

**5-24. The Semiconductor–Metal Barrier Layer.** It is desired to investigate the situation when an $n$-type semiconductor is placed in contact with a metal. Just as in the case of two metals, the equilibrium condition is
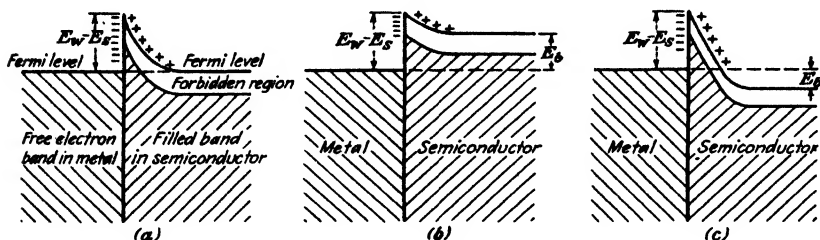


FIG. 5-23. Rectification by a metal-semiconductor barrier. (*a*) No voltage applied. (*b*) Voltage applied in forward direction. (*c*) Voltage applied in backward direction.

reached when the two Fermi levels line up in the body of the crystals. It can be shown that the Fermi level is approximately equal to the donor level for an $n$-type semiconductor and to the acceptor level for a $p$-type semiconductor. If the work function $E_S$ of the semiconductor is assumed to be less than the work function $E_W$ of the metal, electrons will spill over from the semiconductor into the metal. This establishes a negative surface-charge density on the metal surface, with an equal positive charge near the surface of the semiconductor. The approximate width of this positive space-charge region varies from about $10^{-8}$ to $10^{-6}$ m. The barrier at the junction is approximately equal to $E_W - E_S$, the difference in the work functions. The energy-level diagram is pictured in Fig. 5-23$a$.

Rectification in such a metal-semiconductor contact is accomplished in much the same manner as for the case of two metals in contact. Suppose that the system of Fig. 5-23 has a voltage $E_b$ applied in the forward direction, the semiconductor being negative with respect to the metal. Since the potential energy of the semiconductor has been *increased* by the amount $E_b$ electron volts all of the levels in the body of the semiconductor must be raised by this amount, as indicated in Fig. 5-23$b$. The height of the barrier viewed from the metal to semiconductor is a constant, independent

of the applied voltage. However, in the reverse direction the barrier height has been decreased by $\dot{E}_b$ and hence there will be a resultant electron flow in the direction from the semiconductor to the metal. This is analogous to the situation pictured in Fig. 5-19$b$. The limiting condition occurs when $E_b$ equals the contact barrier height, after which the current is determined by the bulk resistance of the semiconductor.

When the voltage is applied in the back or blocking direction, the barrier height as viewed from the metal does not change. However, as viewed from the semiconductor it has increased as indicated in Fig. 5-23$c$. This results in a decrease in the electron flow from semiconductor to metal. The situation is the same as that discussed in Sec. 5-20, where it was shown that the contact is nonohmic and rectification takes place.

Now examine the conditions which exist when $E_S$ of the $n$-type semiconductor exceeds $E_W$ of the metal. The situation under equilibrium is illustrated in Fig. 5-24. The Fermi levels again are lined up. Now, however, electrons have spilled over from the metal to the semiconductor, and a neg-

Fig. 5-24. An $n$-type semiconductor in contact with a metal which has a smaller work function than the semiconductor.

ative space charge layer results which lowers all the levels in the semiconductor by an amount approximately equal to $E_S - E_W$ as indicated. Note that in this case no barrier is formed on contact. No impediment to current flow results, and such a contact will obey Ohms' law.

The situation for a $p$-type semiconductor is somewhat similar to the foregoing. For barrier layer rectification to take place $E_S$ must be greater than $E_W$, and the forward direction is that which makes the semiconductor positive with respect to the metal.

### PROBLEMS

**5-1.** How many electrons per cubic meter in metallic tungsten have energies between 8.5 and 8.6 ev

    *a.* At 0°K?

    *b.* At 2500°K?

**5-2.** Verify that the second term in the Sommerfeld expansion for $E_M$ in terms of $T$ [Eq. (5-12)] is less than one-tenth of 1 per cent of the first term, even for temperatures of tungsten as high as 3000°K.

**5-3.** *a.* Show that the FDS speed distribution function at absolute zero is

$$\rho_v = \frac{8\pi m^3}{h^3} v^2 \qquad \text{for } v < v_M$$

$$\rho_v = 0 \qquad \text{for } v > v_M$$

*b.* Plot $\rho_v$ as a function of $v$ of tungsten at the temperatures $T = 0°K$ and $T = 2500°K$. As ordinate, introduce the variable

$$y = \frac{h^3 \rho_v}{16\pi m^2 e}$$

and, as abscissa, introduce the variable $x = \sqrt{E}$ square-root volts.

**5-4.** Plot the distribution function $\rho_\tau$ as a function of energy $E$ for tungsten at the temperatures $T = 0°K$ and $T = 2500°K$. As ordinate, introduce the variable

$$y = \frac{h^3}{2m^3} \rho_\tau$$

**5-5.** Starting with the fact that $\rho_\tau$ $(= 2m^3/h^3)$ is a constant at absolute zero within a sphere of radius $v_M$ and that $\rho_\tau$ is zero outside of this sphere, find the expression for $v_M$ and $E_M$ in terms of the density of particles $N$.

**5-6.** Calculate the average age of the population of the United States according to the 1940 census (Table 5-1, page 119).

**5-7.** *a.* Calculate the maximum energy of the free electrons in metallic aluminum at absolute zero. Assume that there are three free electrons per atom. The specific gravity of aluminum is 2.7.

*b.* What is the average energy of the electrons under these conditions?

*c.* Repeat parts *a* and *b* for the electrons in metallic silver. The specific gravity of silver is 10.5. Assume that there is one free electron per atom.

**5-8.** *a.* Prove that the average speed of an electron in a metal at absolute zero is three-fourths the maximum speed.

*b.* Prove that the rms speed is 0.774 times the maximum speed.

*c.* Prove that the average value of the reciprocal speed is 1.5 times the reciprocal of the maximum speed. Note that the average value of the reciprocal speed is *not* equal to the reciprocal of the average speed.

*d.* Compute the numerical values of the above quantities for tungsten.

**5-9.** Prove that the specific heat of the electrons in a metal is

$$C_v = \frac{nR\pi^2}{2} \frac{E_T}{E_M} \qquad \text{cal/(°K)(mole)}$$

where $n$ is the number of free electrons per ion and $R$ is the gas constant in calories per degree Kelvin.

Calculate the numerical value of $C_v$ for tungsten at room temperature. Compare this with the classical value $3nR/2$ cal/(°K)(mole).

**5-10.** *a.* An electron within a metal whose volume is only $10^{-12}$ m³ has an energy of 1 ev. To within what percentage is its speed known, according to the uncertainty principle?

*b.* Explain why the uncertainty principle plays no important role in macroscopic measurements.

**5-11.** Write down the expression for the number of electrons per cubic meter of metal whose $X$-directed velocities are between 0 and $10^6$ m/sec, whose $Y$-directed velocities are between $10^5$ and $10^7$ m/sec, and whose $Z$-directed velocities are greater than $10^7$ m/sec. How is this interpreted geometrically in velocity space?

**5-12.** The $X$-directed FDS distribution function is given by Eq. (5-31), namely,

$$dN_x = \int_{v_y = -\infty}^{+\infty} \int_{v_z = -\infty}^{+\infty} \rho_\tau \, d\tau = \frac{2m^3}{h^3} \, dv_x \iint_{-\infty}^{+\infty} \frac{dv_y \, dv_z}{1 + e^{[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - eE_M]/eE_T}}$$

This function gives the number of electrons per cubic meter of metal that have $X$-directed velocities between $v_x$ and $v_x + dv_x$ and *any* $Y$- or $Z$-directed velocity whatsoever. Show that

$$dN_x = \frac{4\pi m^2 e E_T}{h^3} \ln\left(1 + \epsilon^{\frac{E_M - E_x}{E_T}}\right) dv_x$$

where $eE_x = \frac{1}{2}mv_x^2$.

HINT: The integration can be carried out by changing from Cartesian coordinates $v_y$ and $v_z$ to plane polar coordinates $R$ and $\theta$ as shown in the figure. Then

$$R^2 = v_y^2 + v_z^2$$

The element of area $dv_y\, dv_z$ must be replaced by $R\, dR\, d\theta$. The integration extends over the *entire* $v_y v_z$ plane, and hence the limits of integration are as follows: $\theta$ from zero to $2\pi$ and $R$ from zero to infinity. The integration over $R$ is carried out with the aid of the following identity:
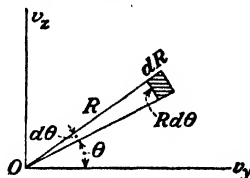
$$\int \frac{d\zeta}{1 + A\epsilon^\zeta} = -\ln(A + \epsilon^{-\zeta})$$



PROB. 5-12.

**5-13.** *a.* Using the result of the preceding problem, show that the completely degenerate $X$-directed distribution function of an FDS gas is

$$dN_x = \rho_x\, dv_x$$

where

$$\rho_x = \frac{4\pi m^2 e}{h^3}(E_M - E_x) \qquad \text{for } E_x < E_M$$

$$\rho_x = 0 \qquad \text{for } E_x > E_M$$

*b.* Plot the $X$-directed distribution function $\rho_x$ as a function of $E_x$ for tungsten at the temperatures $T = 0°\text{K}$ and $T = 2500°\text{K}$. As ordinate, introduce the variable

$$y = \frac{h^3\rho_x}{4\pi m^2 e}$$

**5-14.** Starting with the $X$-directed distribution function of Prob. 5-12, find the number of electrons per cubic meter whose energies are large enough so that they can escape from the metal $(E_x > E_B)$. Prove that the ratio of this number to the concentration of electrons is

$$\frac{3E_T^2}{8E_B^{\frac{1}{2}}E_M^{\frac{3}{2}}} \epsilon^{-E_W/E_T}$$

Prove that for tungsten at $3000°\text{K}$ but one electron out of every $2 \times 10^{11}$ electrons in the metal has enough energy to escape. This means that thermionic emission disturbs the FDS distribution inappreciably.

NOTE: For small values of $\zeta$

$$\ln(1 + \zeta) = \zeta$$

For large values of $\eta_0$

$$\int_{\eta_0}^\infty \epsilon^{-\eta^2}\, d\eta = \frac{\epsilon^{-\eta_0^2}}{2\eta_0}$$

**5-15.** *a.* Prove that the average $X$-directed velocity of a completely degenerate gas is three-eighths of the maximum speed $v_M$.

*b.* Prove that the average $X$-directed energy of an electron at absolute zero is one-fifth of the maximum, or Fermi, energy $E_M$.

**5-16.** Using the distribution function $dN_x'$ of Eq. (5-31), prove that the number of electrons crossing 1 m$^2$/sec at absolute zero in a metal is

$$\frac{\pi m^3 v_M{}^4}{2h^3}$$

**5-17.** Suppose that the work function $E_W$ is temperature dependent. If it is assumed that this dependency may be expressed by an equation of the form $E_W = E_{W0} + \alpha T$, what must be the magnitude of $\alpha$ in order that the Dushman-equation value of thermionic-emission current $J_{th}$, calculated with $A_0 = 120 \times 10^4$ amp/(m$^2$)($^\circ$K$^2$), be reduced by a factor of 2? (See S. Seely, *Phys. Rev.*, **59**, 75, 1941, for a discussion of the temperature dependence of work function.)

**5-18.** Carry through the integrals in Eq. (5-42) to prove that the average energy of the electrons leaving a metal is $2E_T$.

**5-19.** Starting with the distribution function $dN_r'$ of Eq. (5-47) for the electrons that have escaped from a metal, derive the thermionic-emission equation.

**5-20.** Starting with the distribution function $dN_x'$ of Eqs. (5-31) and (5-32), derive the thermionic-emission equation.

**5-21.** Show that the number of electrons having $X$-directed energies between $E_x$ and $E_x + dE_x$ is

$$(2em)^{\frac{3}{2}} \left( \frac{\pi E_T}{h^3} \right) \ln \left[1 + \epsilon^{(E_M - E_x)/E_T} \right] E_x{}^{-\frac{1}{2}} \, dE_x$$

HINT: Use the result of Prob. 5-12.

**5-22.** Prove that the number of electrons that strike 1 m$^2$ of the surface of a metal in 1 sec, having $X$-directed energies exceeding $2E_M$, is

$$\frac{4\pi m k^2 T^2}{h^3} \, \epsilon^{-E_M/E_T}$$

where the symbols have the meaning given in the text.

**5-23.** Prove that the energy distribution function of the electrons *that have escaped* from a metal is

$$dN_E = A \epsilon^{-E/E_T} E^{\frac{1}{2}} \, dE$$

where

$$A = \frac{8\pi \sqrt{2}(em)^{\frac{3}{2}}}{h^3} \, \epsilon^{-E_W/E_T}$$

This is one form of the Maxwell-Boltzmann distribution function which will be discussed in detail in Chap. 8.

**5-24.** Prove that the density of electrons in the immediate neighborhood of a thermionic cathode which is delivering saturation current is given by

$$\frac{2}{h^3} (2\pi m k T)^{\frac{3}{2}} \epsilon^{-E_W/E_T}$$

Find the numerical value for this density if the cathode is tungsten operating at 3000$^\circ$K.

**5-25.** Consider a cylindrical tube operating with a retarding potential. Carry through the integrations indicated in Sec. 5-18.

HINT: It will be found convenient to change to polar coordinates in a manner similar to that suggested in Prob. 5-12.

The result is

$$I = \frac{4I_{th}}{\sqrt{\pi}} \int_{\sqrt{E_r/ET}}^{\infty} \epsilon^{-x^2} x^2 \, dx$$

An integration by parts then yields Eq. (5-55) of the text.

**5-26.** A retarding-potential experiment on a tube having cylindrical symmetry (with the anode radius much larger than the cathode radius) yields the following data:

| $E_b$, volts | 0.60 | 0.70 | 0.80 | 0.85 | 0.90 | 1.0 | 1.3 | 1.8 | 2.4 |
|---|---|---|---|---|---|---|---|---|---|
| $I_b$, $\mu$a | 0.4 | 1.2 | 3.0 | 5.8 | 10.0 | 12.3 | 14.8 | 15.1 | 15.8 |

From the appropriate plot of these data, determine the contact difference of potential between the cathode and anode and the temperature of the cathode.

**5-27.** *a.* What must be the strength of the electric field at the surface of a tungsten filament maintained at 2300°K in order that the work function be decreased by 2 per cent?

*b.* What is the percentage increase in the thermionic emission under these circumstances?

*c.* What is the maximum normal-to-the-surface distance that an electron may wander outside of the metal and yet *not* be able to escape?

**5-28.** Show that the potential energy is a maximum at the escape distance in the Schottky theory of the effect of accelerating fields.

**5-29.** The following data were taken with a diode having a pure tungsten filament as an emitter:

| $E_b$, volts.... ........ | 4 | 8 | 20 | 36 | 64 | 100 |
|---|---|---|---|---|---|---|
| $I_b$ (1 $\mu$a = 250 units)....... | 122 | 190 | 231 | 246 | 260 | 274 |

| $E_b$, volts............... | 144 | 200 | 256 | 400 | 600 | 900 |
|---|---|---|---|---|---|---|
| $I_b$ (1 $\mu$a = 250 units)....... | 280 | 290 | 296 | 312 | 328 | 350 |

Plot the Schottky line log $I_b$ vs. $E_b^{\frac{1}{2}}$. (These data were taken from E. W. B. Gill, *Phil. Mag.*, **24**, 1093, 1937.)  Find the zero field thermionic current.

**5-30.** Consider two metals in contact, as in Fig. 5-9. The number of electrons per second per square meter leaving surface $A$ of metal 1 is  given by Eq. (5-47),

$$dN_r' = \frac{2m^3}{h^3} \epsilon^{-(E + E_{W1})/ET} v_x \, dv_x \, dv_y \, dv_z$$

If the contact potential between the metals is $E_{AB}$, what is the expression for the number of electrons per second per square meter that reach the surface $B$ of metal 2 after escaping from metal 1?  What is the expression for the number of electrons per second per square meter escaping from the surface $B$ of metal 2?  Equating these two

expressions, since they must be equal if a state of equilibrium is reached between metals 1 and 2, show that $E_{AB} = E_{W2} - E_{W1}$.

**5-31.** If two metals are in contact, show, from conservation of energy, that a double layer of charge must exist at the surface of contact.

### REFERENCES

1. FERMI, E., *Z. Physik*, **36**, 902, 1926.
2. DIRAC, P. A. M., *Proc. Roy. Soc.* (*London*), **112**, 661, 1926.
3. SOMMERFELD, A., and H. BETHE, "Handbuch der Physik," Vol. XXIV/2, p. 333, Verlag Julius Springer, Berlin, 1933.
   DARROW, K. K., *Bell System Tech. J.*, **8**, 672, 1929.
4. DOW, W. G., "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.
5. SOMMERFELD, A., *Z. Physik*, **47**, 1, 1928.
   MILLMAN, J., and S. SEELY, "Electronics," 1st ed., Appendix IV, McGraw-Hill Book Company, Inc., New York, 1941.
6. PAULI, W., *Z. Physik*, **31**, 765, 1925.
7. HEISENBERG, W., *ibid.*, **43**, 172, 1927.
8. RICHARDSON, O. W., *Phil. Mag.*, **28**, 633, 1914.
   DUSHMAN, S., *Phys. Rev.*, **21**, 623, 1923.
9. BECKER, J. A., *Revs. Modern Phys.*, **7**, 95, 1935.
   SEELY, S., *Phys. Rev.*, **59**, 75, 1941.
   SEITZ, F., "The Modern Theory of Solids," McGraw-Hill Book Company, Inc., New York, 1940.
10. HERRING, C., and M. H. NICHOLS, *Revs. Modern Phys.*, **21**, 185, 1949.
11. BECKER, J. A., *ibid.*, **7**, 95, 1935.
12. SOMMERFELD and BETHE, *op. cit.*
    GYSAE, B., and S. WAGENER, *Z. Physik*, **110**, 145, 1938.
13. GLASOE, G. N., *Phys. Rev.*, **38**, 1490, 1931.
    ANDERSON, P. A., *ibid.*, **47**, 958, 1935.
    HEINZE, W., *Z. Physik*, **109**, 459, 1938.
    KLEIN, O., and E. LANGE, *Z. Elektrochem.*, **44**, 542, 1938.
    GYSAE, B., and S. WAGENER, *Z. Physik*, **115**, 296, 1940.
14. ZISMAN, W. A., *Rev. Sci. Instruments*, **3**, 367, 1932.
15. POTTER, J. G., *Phys. Rev.*, **58**, 623, 1940.
16. SCHOTTKY, W., *Ann. Physik*, **44**, 1011, 1914.
17. GERMER, L. H., *Phys. Rev.*, **25**, 795, 1925.
18. NOTTINGHAM, W. B., *ibid.*, **41**, 793, 1932.
19. FAN, H. Y., *J. Applied Phys.*, **14**, 552, 1943.
20. SCHOTTKY, W., *Physik. Z.*, **15**, 872, 1914.
21. DEBRUYNE, N. A., *Proc. Roy. Soc.* (*London*), **120**, 423, 1928.
    PFORTE, W. S., *Z. Physik*, **49**, 46, 1928.
    NOTTINGHAM, W. B., *Phys. Rev.*, **47**, 806, 1935; **49**, 78, 1936.
    WHEATCROFT, E. L. E., *Phil. Mag.*, **29**, 16, 1940.
22. NOTTINGHAM, W. B., *Phys. Rev.*, **57**, 935, 1940.
    GUTH, E., and C. J. MULLIN, *ibid.*, **59**, 575, 1941.
23. GUTH, E., and C. J. MULLIN, *ibid.*, **59**, 867, 1941; **61**, 339, 1942.
24. TORREY, H. C., and C. A. WHITMER, "Crystal Rectifiers," Radiation Laboratory Series, Vol. 15, McGraw-Hill Book Company, Inc., New York, 1948.
    SHOCKLEY, W., "Electrons and Holes in Semiconductors," D. Van Nostrand Company, Inc., New York, 1951.

# CHAPTER 6

# CHARACTERISTICS OF THERMIONIC CATHODES

$T$HE modern concept of the electronic structure of a metal has been discussed in the foregoing two chapters. It is there shown that in order to release even the "free" electrons from the metal a certain amount of energy must be supplied to them. This energy may be supplied by any one of a number of different methods. One of the important methods is to heat the metal. In this way, some of the thermal energy supplied to the metal is transferred from the lattice of the heated-metal crystals into kinetic energy of the electrons.

**6-1. Determination of Emission Constants.** It has already been shown in Sec. 4-8 that the relation which exists between the thermionic-emission current and the temperature of the metal is expressible by the Dushman equations (4-9), which are rewritten for convenience,

$$I_{th} = A T^2 \epsilon^{-E_W/E_T} \qquad I_{th} = A T^2 \epsilon^{-b_0/T} \qquad \text{amp} \qquad (6\text{-}1)$$

where $A = A_0 S$, $S$ is the area of the emitting surface, $T$ is the temperature of the metal in degrees Kelvin, $E_W$ is the work function of the emitter in electron volts, $A_0$ is a constant, $b_0 = 11,600\, E_W$, and $E_T = T/11,600$.

The graphical representation between the thermionic-emission current and the temperature is generally obtained by taking the logarithm of Eq. (6-1), *viz.*,

$$\log_{10} I_{th} - 2 \log_{10} T = \log_{10} A - 0.434 b_0 \left( \frac{1}{T} \right) \qquad (6\text{-}2)$$

where the factor 0.434 represents $\log_{10} \epsilon$. In order to verify the thermionic-emission equation, one plots $\log_{10} I_{th} - 2 \log_{10} T$ vs. $1/T$, and the result should be a straight line having a slope equal to $-0.434 b_0$. The result obtained for the tungsten filament in a General Electric Co. high-vacuum diode, type FP 85, is shown in Fig. 6-1. From the slope of the straight line a value of $b_0 = 52,500$ deg is obtained. By using this value of $b_0$ in Eq. (6-2), the value of $A_0$ is found to be $50 \times 10^4$ amp/(m²)(°K²). These values compare favorably with the generally accepted values of $b_0 = 52,400$ and $A_0 = 60.2 \times 10^4$ given in Table 6-1, page 175.

The exact experimental verification of Eqs. (6-1) is extremely difficult. Many stray effects must be guarded against, and many precautions must

be observed.  In particular, an extremely high vacuum must be employed, and all adsorbed and occluded gases must be completely removed.  That is, quantitative verification of the theory necessitates a carefully planned schedule of pumping, baking of the glass envelope, outgassing of the supports, etc.  Details of these processes are contained in the excellent review of the subject by Dushman.[1]  The reader who is interested in specific details of experimental technique should consult the original literature.[2] The necessity for the elaborate precautions that must be taken in deter-
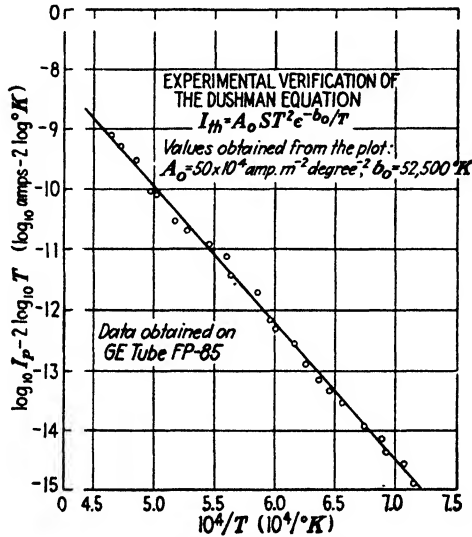


Fig. 6-1.  Experimental verification of the Dushman equation, using a tungsten filament.

minations of this kind arises from the fact that tiny traces of some foreign substances can change the emission from an element very materially.  This effect will be discussed in greater detail in Sec. 6-4.

In order to collect all the electrons that are emitted thermionically from a given surface, an accelerating field must be employed.  The application of an accelerating field at the surface of the metal results, however, in the variation of the height of the potential barrier at the surface of the metal, as discussed in Sec. 4-11.  This necessitates a careful determination of the current as a function of applied field, and then the results must be extrapolated to zero field to obtain the true thermionic currents.

The range of temperatures over which the emission current can be measured is limited at the low temperatures by the sensitivity of the instruments used to measure this current and at the high temperatures by the melting point of the metal.  In the case of tungsten, the extremes of

the range of investigation are 1000° and 3000°K. This results in a current ratio of approximately $10^{12}$. The currents of Fig. 6-1 extend from approximately $10^{-9}$ to $10^{-2}$ amp, a range of $10^7$. The temperature range covered extends from about 1400° to 2300°K.

**6-2. Temperature Determination.** In order to plot the curve of Fig. 6-1, it is necessary to know the cathode temperature accurately. In those cases where the cathode is sufficiently exposed, the temperature can most accurately be determined by means of an optical pyrometer. Often, however, it is difficult or entirely impossible to see the cathode. Under these conditions a method that is based upon the energy radiated by the cathode is usually employed.

If a certain amount of power is supplied to a cathode, it will become heated and the temperature will increase until temperature equilibrium occurs. Equilibrium exists when the rate of heat removal by all causes equals the rate of heat produced as a result of the electrical input. Since the cathode is generally a thin filament in a vacuum, no convection of heat can occur. A small amount of heat will be conducted away by the leads. Also, a slight cooling occurs because of the electron emission itself, the amount of power lost by this cause being the product of the emission current, in amperes, and $E_W + 2E_T$, in volts (see Prob. 6-15). Although this term may become appreciable in gas-filled tubes, where the emission current is high, it is generally negligible compared with the heat loss due to radiation, especially in high-vacuum thermionic tubes. Consequently, most of the heat loss is due to the radiant losses, which exist largely in the infrared region of the electromagnetic spectrum.

The resultant rate of the radiant-energy dissipation from a cathode depends, of course, upon the rate at which energy is radiated by the cathode to its surroundings and upon the rate at which energy is radiated to the cathode by the surroundings. In the case where no appreciable heat shielding exists, the contribution from the latter is usually very small. Under these conditions, the rate at which energy is radiated from the heated surface is expressed explicitly as a function of the temperature of the body by the Stefan-Boltzmann relation

$$P = 5.77 \times 10^{-8} e_T T^4 \qquad \text{watts/m}^2 \qquad (6\text{-}3)$$

where $P$ is the power radiated in watts per square meter, by the surface whose emissivity is $e_T$; the factor $5.77 \times 10^{-8}$ watts/$(\text{m}^2)(°\text{K}^4)$ is known as the Stefan-Boltzmann constant; and $T$ is the temperature in degrees Kelvin.

The value of $e_T$, the emissivity of the body, is always less than unity for all practical cases. It is a measure of the ratio of the energy radiated by the particular surface to the energy that would be radiated by a perfect radiator, or "black body," at the same temperature. This coefficient, as

indicated by the subscript, is a quantity that varies slightly with the temperature. The emissivity is a quantity that is characteristic of a surface, and it applies as readily to heat absorption as it does to heat radiation. Hence, surfaces that have high emissivities not only radiate energy readily but also absorb heat readily. The black body is, therefore, a perfect heat absorber as well as a perfect heat radiator. In short, the coefficient $e_T$ is the experimentally determined correction factor that must be applied to the theoretical law for a black body for the results to apply to any particular emitting surface.

If the emissivity were a constant, then a plot of the equation

$$\log_{10} P = \log_{10} (5.77 \times 10^{-8} e_T) + 4 \log_{10} T \qquad (6\text{-}4)$$

would be a straight line having a slope equal to 4. Actually, as already mentioned, the emissivity varies with the temperature and so must be determined experimentally for different temperatures for any substance. Forsythe and Worthing [3] and Jones and Langmuir [4] have determined the temperature of tungsten as a function of the input power per square centimeter, over wide ranges of temperature.

The results of the investigations of Jones and Langmuir have been plotted in such a way in Fig. 6-2 as to be most useful in connection with applications to the verification of the thermionic-emission equation. Here $\log_{10} P$ has been chosen as the independent variable and $2 \log_{10} T$ as the dependent variable. Included on the same curve sheet is a plot of $2 \log_{10} T$ vs. $10^4/T$. This curve materially simplifies the calculations involved in the verification of the Dushman equation. Thus $P$ is obtained by dividing the power input to the filament by its surface area. The value of $2 \log_{10} T$ corresponding to this value of power input per unit area is first obtained. The value of $10^4/T$ corresponding to this value of $2 \log_{10} T$ is then read. Thus, both functions of the temperature $T$ that enter into Eq. (6-2) are obtained directly from the curves.

For the cases where the cathode may be heat-shielded, resulting in a system that returns by reflection an appreciable portion of the heat energy that might otherwise be lost by radiation, the temperature of the surroundings may be appreciable. For such cases, Eq. (6-3) must be replaced by the expression

$$P = 5.77 \times 10^{-8} e_T (T^4 - T_0^4) \qquad \text{watts/m}^2 \qquad (6\text{-}5)$$

where $T_0$ represents the temperature of the surroundings into which the cathode radiates the energy. Such heat-shielded cathodes are suitable for use in gas-filled tubes but cannot be used in vacuum tubes, as will later be discussed.

It should be evident that the entire filament does not operate at a uniform temperature, since a temperature gradient exists near the ends of the
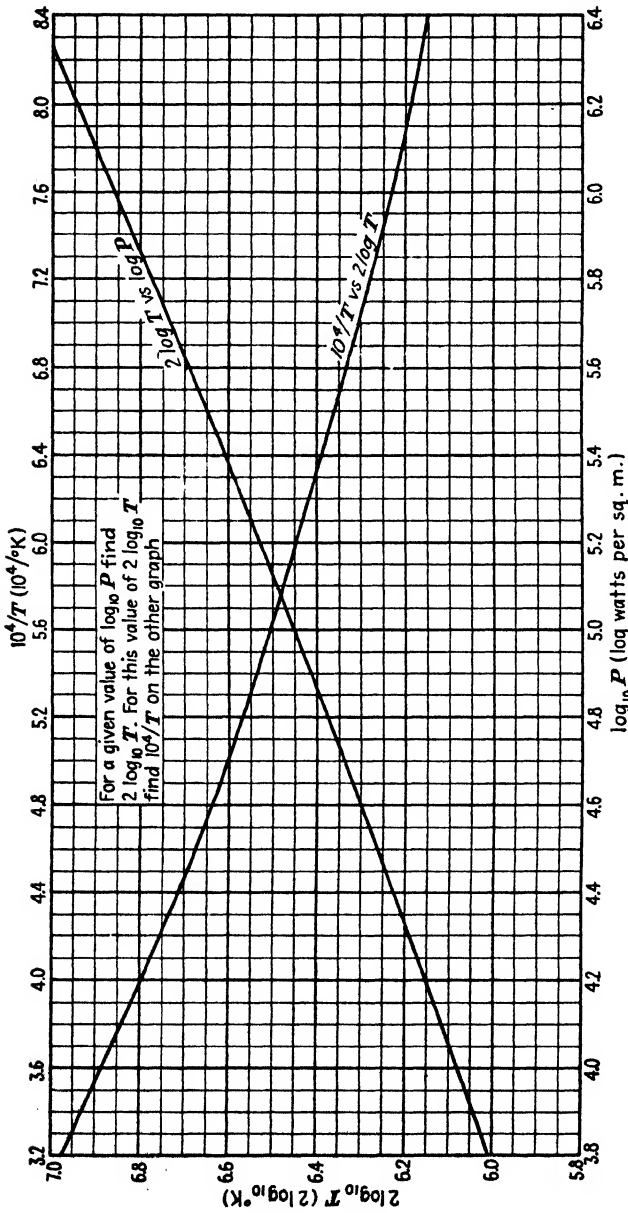
FIG. 6-2.  Graphs for determining the temperature functions $2 \log T$ and $10^4/T$ for given values of the power input per square meter of a tungsten filament.

filament. This temperature gradient arises from the fact that one end of each lead-in wire is at room temperature. Because of this temperature gradient, some of the heat is conducted away from the filament wires. Also, this decreased temperature at the ends of the filament results in a smaller thermionic-emission current from these portions. This, in turn, results in a different rate of cooling of the filament due to the "evaporation" of the electrons from the surface of the metal. Still another reason for a nonuniform temperature exists. The collecting electrode is connected to the cathode through a battery, the battery that supplies the field to attract the electrons from the cathode to the anode. Hence, the anode current is returned to the cathode. If the anode current is returned to one end of the filament, then the end of the filament to which the plate current is returned will be hotter than the other end, owing to the additional heating that results from the superposed anode current on the filament current. This contribution tends to compensate for the cooling due to the "evaporation" of the electrons, although the effects do not occur at the same portions of the filament.

**6-3. Other Forms of the Emission Equations.** An early form of an emission equation was suggested on the basis of the classical kinetic theory by a pioneer in the field of thermionic emission, O. W. Richardson.[5] The theoretical basis for this equation will be investigated in Chap. 8. The form of the equation is

$$I_{th} = A'T^{\frac{1}{2}}\epsilon^{-b'/T} \qquad (6\text{-}6)$$

where $A'$ is a quantity that depends upon the material, and $b'$ is a quantity related to, but not equal to, $b_0$. These parameters will be defined explicitly in Sec. 8-12.

Experimentally, it is impossible to distinguish between Eqs. (6-1) and (6-6). This arises from the fact that both equations predict the same exponential dependence upon the temperature; and since this factor is such a rapidly varying one, it overshadows the dependence upon the $T^{\frac{1}{2}}$ or the $T^2$ term. For example, it follows from Eq. (6-1), by taking the derivative of the natural logarithm of this equation, that

$$\frac{dI_{th}}{I_{th}} = \left(2 + \frac{b_0}{T}\right)\frac{dT}{T} \qquad (6\text{-}7)$$

For tungsten, $b_0 = 52,400$, so that at a normal operating temperature of 2400°K the fractional change in current $dI_{th}/I_{th}$ is $2 + 22$ times the fractional change in the temperature. It is to be noted that the term $22(= b_0/T)$ arises from the exponential term in the Dushman equation, and the term 2 arises from the $T^2$ term. Because of this slight dependence upon the power of the $T$ term, it is impossible to use the experimental results as a criterion to favor one or the other equation. In fact, the

experimental results are found to be in accord with an equation of the form given by Eq. (6-6) in which the exponent of the $T$ term has almost any value up to 4.

It is to be noted that if the thermionic current is known and it is desired to find the corresponding operating temperature of the filament, a transcendental equation must be used as the basis for calculating the temperature $T$. However, because of the foregoing observation that the exponential term is a rapidly varying one, the solution may be accomplished readily by a simple method of "successive approximations."

*Example.* A tungsten filament whose length is 1.0 in. and whose diameter is 0.01 in. gives a saturation current of 2 $\mu$a. At what temperature is the filament operating?

*Solution.* The cathode area is

$$S = \pi Dl = \pi \times 0.01 \times 1.0 \times 2.54^2 \times 10^{-4} = 0.203 \times 10^{-4} \text{ m}^2$$

and Eq. (6-2) becomes

$$\log_{10}(2 \times 10^{-6}) - 2\log_{10} T = \log_{10}(60.2 \times 0.203) - 0.434 \times 52,400 \times \frac{1}{T}$$

where the experimentally determined values $A_0 = 60.2 \times 10^{+4}$ amp/(m)$^2$(°K$^2$) and $b_0 = 52,400$°K are used (see Table 6-1, page 175). There results

$$T = \frac{22,750}{6.786 + 2\log_{10} T}$$

This equation cannot be solved in a straightforward manner. However, a little thought will show that the term $2\log_{10} T$ contained in the denominator arises from the $T^2$ term and is, in accordance with the foregoing discussion, a relatively unimportant factor. Consequently, any reasonable guess for $T$ can be used and substituted into the right-hand term of the expression, the resulting fraction being evaluated for $T$. This value of $T$ is then substituted into the logarithmic term that appears in the denominator, and the second approximation for $T$ is calculated. This process is continued until two successive approximations yield the same temperature within any specified accuracy.

Although the method may appear rather complicated, actually the calculations are straightforward, and the process converges very rapidly. In the present case, since the operating temperature of tungsten ranges between 1000° and 3000°K, choose, as a first approximation, $T = 2000$°K. Substituting this value of $T$ in the logarithmic term of the foregoing equation and solving for $T$ yields

$$T = \frac{22,750}{6.786 + 6.602} = 1700\text{°K}$$

By using $T = 1700$°K as a second approximation, there results

$$T = \frac{22,750}{6.786 + 6.462} = 1715\text{°K}$$

By inserting $T = 1715$°K as the third approximation, there results

$$T = \frac{22,750}{6.786 + 6.466} = 1715\text{°K}$$

Thus the temperature of the cathode is 1715°K within slide-rule accuracy.

Many other physical problems arise in which similar methods of successive approximation must be used. The general method of approach is very much like that illustrated in this example.

**6-4. Cathode Materials.** The filaments of the tubes used by the early investigators were chiefly of platinum, tungsten, or carbon, although many other materials were used. It was not until 1904 that Wehnelt [6] investigated the emission of electrons from metallic oxides, such as those of barium, strontium, or calcium, coated on a platinum wire, and found that these materials made excellent emitters. The work of Langmuir and Rogers [7] in 1914 showed that the application of 1 or 2 per cent of thorium oxide to a tungsten wire, when properly heat-treated, results in an emission current that greatly exceeds that of pure tungsten. The three most important practical emitters will now be discussed in detail.

*Tungsten.* Although tungsten becomes very brittle because of recrystallization, it is electrically rugged. That is, tungsten does not have an active surface layer that can be damaged by positive-ion bombardment, and so positive-ion bombardment does not influence the emitting properties appreciably. Hence, tungsten is used as the cathode in high-voltage high-vacuum tubes. These include X-ray tubes, diodes for use as rectifiers above about 5,000 volts, and large power-amplifier tubes for use in radio and other communication transmitters.

Certain gases attack tungsten appreciably when hot. This is particularly true of water vapor and oxygen. Hence a very high vacuum must be maintained. Furthermore, its emission is impaired by the presence of certain other gases, for example, nitrogen, although its emission is not appreciably affected by the presence of mercury vapor and the inert gases (helium, neon, argon, etc.).

Tungsten has the disadvantage that the *cathode emission efficiency*, defined as the ratio of the emission current, in milliamperes, to the heating power, in watts, is small. However, a copious supply of electrons can be provided by operating the cathode at a sufficiently high temperature. The higher the temperature, however, the greater will be the evaporation of the filament during its operation and the sooner it will burn out. Economic considerations dictate that the temperature of the filament be about 2500°K, which gives it a life of approximately 2,000 hr. The melting point of tungsten is 3650°K.

The important constants for a tungsten emitter are summarized in Table 6-1.

*Thoriated Tungsten.* In order to obtain copious emission of electrons at moderately low temperatures, it is necessary for the material to have a low work function. Such materials would then provide large emission with low heating power. Unfortunately, however, the low-work-function metals, such as cesium, rubidium, and barium, in some cases melt and in

other cases boil at temperatures necessary for appreciable thermionic emission.

It is possible to apply very thin layers of low-work-function material on filaments of tungsten, platinum, and other metals, the low-work-function layer being adsorbed by the base metal. In this way the base metal holds the adsorbed layer at high temperatures, even above the point at which the pure layer material would normally evaporate. This is the case with tungsten filaments that have their surfaces covered by a monatomic layer of thorium atoms. Such a filament has the desirable

TABLE 6-1

COMPARISON OF THERMIONIC EMITTERS

| Type of cathode | Dushman constants | | | Approximate operating temperature, °K | Efficiency,* amp/($m^2$) (watt) | Plate voltage, volts | Gas or vacuum tube |
|---|---|---|---|---|---|---|---|
| | $A_0 \times 10^{-4}$, amp/($m^2$) (°$K^2$) | $b_0$, °K | $E_W$, ev | | | | |
| Tungsten........... | 60.2 | 52,400 | 4.52 | 2,500 | 20-100 | Above 5,000 | Vacuum |
| Thoriated tungsten.. | 3.0 | 30,500 | 2.63 | 1,900 | 50-1,000 | 750-5,000 | Vacuum |
| Oxide-coated BaO + SrO | 0.01 | 11,600 | 1.0 | 1,000 | 100-10,000 | Below 750 | Vacuum or gas |

* K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.

mechanical properties of tungsten and possesses emission properties that are considerably better than those of the pure tungsten.

Thoriated-tungsten filaments are obtained by adding a small amount (1 or 2 per cent by weight) of thorium oxide to the tungsten. After the tungsten is drawn and mounted in a tube, the tube is exhausted and the glass and the metal structure are outgassed by baking the tube in an oven. The temperature to which the glass may be heated during the outgassing process depends upon the type of glass used. This temperature is about 400°C for soft glass and about 500°C for pyrex glass. The metallic parts are further heated with a high-frequency induction furnace. Just as the tube is sealed off from the pumping system, a getter is "flashed." The getter consists of an active chemical substance, such as magnesium, which absorbs the residual gas and tends to maintain a high vacuum. This getter is usually visible in commercial tubes, it being the silvery deposit on the glass near the base of the tube.

Following the evacuation process, the filament must be activated. This process requires essentially three steps. (1) The filament is heated to about 2800°K for several minutes. At this high temperature, the tungsten

surface is cleaned and some of the thoria inside the tungsten is reduced to metallic thorium. At this high temperature, any thorium that diffuses to the surface of the filament will immediately evaporate, so that the emission under these conditions is substantially that of the pure tungsten. (2) The filament is then activated by maintaining the temperature of the filament at about 2100°K. At these temperatures the rate of diffusion of the thorium to the surface is still rather high, although the rate of evaporation from the surface is rather low. In this way an adsorbed layer approximately one molecule thick of thorium atoms accumulates on the surface. Under these conditions, the emission from the activated surface is approximately 1,000 times that of the pure tungsten filament at the same temperature. (3) The temperature of the filament is reduced to the operating range, from 1800° to 2000°K, at which temperatures the evaporation and diffusion are so low that the metal retains its surface layer of thorium.

Under the conditions of normal operation the layer of thorium atoms slowly evaporates from the filament. This evaporation is compensated for by the continued diffusion of thorium atoms to the surface. However, even under proper use, the filament has a limited life. It can be rejuvenated by reactivation since only a limited amount of the thorium within the tungsten can be removed at the lower temperatures. In order to increase the life of thoriated-tungsten filaments, the process of *carbonization* has been developed. The filament is heated at a temperature about 1600°K in the presence of naphthalene or some other suitable hydrocarbon vapor. The hydrocarbon is reduced, and some of the carbon from this process diffuses into the tungsten and forms a surface layer of tungsten carbide. The effect of the process is to reduce the rate of evaporation of the thorium layer, the rate being about one-sixth that of the noncarbonized filament.[8]

The thoriated-tungsten emitter has a much higher cathode efficiency than a pure tungsten emitter, the uncarbonized filament being slightly more efficient than the carbonized one. The decrease of efficiency that results from the carbonization is more than compensated for by the decreased probability of impairment of the emission characteristics by oxidation. The limitation to the use of thoriated-tungsten emitters is the deactivation due to positive-ion bombardment. The effect of even a few ions is severe at high potentials, so that these filaments are confined to use in tubes that operate with potentials less than about 5,000 volts. Higher voltage tubes use pure tungsten filaments. Most of the "800" series of transmitting tubes use thoriated-tungsten filaments.

The effect of the thorium present on the tungsten surface is to lower the height of the potential barrier at the surface of the metal to a value below that of either the pure tungsten or the pure thorium. It is rather difficult to obtain representative values for the emission constants in the Dushman

equation, since these values depend markedly upon the fraction of the surface that is covered by thorium. Though several values of $A_0$ and $b_0$ are to be found in the literature, it will be here assumed that the values [9] given in Table 6-1 are representative. The presence of the monatomic layer of thorium on the surface of the tungsten does not alter the thermal properties of the tungsten; it is possible, therefore, to determine the temperature of the filament from the thermal curves of Fig. 6-2 for pure tungsten.

*Example.* The saturation current from a certain thoriated-tungsten filament operating at 2000°K is 100 ma. What would be the emission from a pure tungsten filament of the same area operating at the same temperature?

*Solution.* Equation (6-1) for tungsten is

$$I = (S)(60.2 \times 10^4)(2{,}000)^2 \epsilon^{-52{,}400/2{,}000}$$

Similarly, for thoriated tungsten, Eq. (6-1) becomes

$$100 \times 10^{-3} = (S)(3.0 \times 10^4)(2{,}000)^2 \epsilon^{-30{,}500/2{,}000}$$

Upon dividing these two equations, there results

$$\frac{I}{0.1} = \frac{60.2}{3.0} \epsilon^{-\frac{52{,}400}{2{,}000} + \frac{30{,}500}{2{,}000}}$$

or

$$I = 2.01 \epsilon^{-10.95}$$

This is evaluated with the aid of logarithms. Thus,

$$\log_{10} I = \log_{10}(2.01) - (0.434)(10.95)$$

$$= 0.30 - 4.75 = -4.45 = 0.55 - 5$$

$$I = 3.55 \times 10^{-5} \text{ or } 35.5 \ \mu a$$

*Oxide-coated Cathodes.*[10] The modern oxide-coated cathode is the most efficient type of emitter that has been developed commercially. It consists of a metallic base of platinum, nickel, nickel with a few per cent of cobalt or silicon, or Konal metal. Konal metal is an alloy consisting of nickel, cobalt, iron, and titanium. Konal-metal sleeves are used very extensively as the indirectly heated cathode of radio receiving tubes.

The wire filaments or the metallic sleeves are coated with oxides of the alkaline-earth group, especially barium and strontium oxides. This coating is applied to the surface of the cathode either by a spray or by dip, an organic binder being used to make the coating stick to the metal sleeve. Since the alkaline-earth oxides are unstable in the presence of the gases of the atmosphere, the coating usually consists of the alkaline-earth carbonate, nitrate, or hydroxide. Upon applying heat (about 1400°K) in a vacuum, the coating is reduced to the oxide, and the gas that is evolved is pumped away.

Before the oxide-coated cathode shows appreciable electron emission, it must be activated. The activation process consists essentially in operat-

ing the cathode for several minutes at a temperature from 1000° to 1500°K, which is above the normal operating temperature. The cathode is then maintained at a lower temperature for a longer period of time, an anode potential being applied so as to draw electron current. During this process, the emission increases rapidly to a high value and reaches an emission current that is as copious at 1000°K as an equivalent tungsten filament that is maintained at 2300°K. The corresponding cathode emission efficiency is about twenty times that of the tungsten emitter. It is for this reason that these cathodes are used so extensively. The activation process probably results in the reduction of some oxide to the pure metallic form, which will be distributed throughout the body of the coating.

There seems to be some difference of opinion among investigators as to whether the emission comes from the pure metal or the oxide, and also whether the material of the core affects the emission properties. The accumulated evidence seems to favor the view that the emission current arises principally from the free metal on the surface.[11] However, there seems to be some dependence of emission current on the core material.[12]

Three other characteristics of the coating exist that account for its extensive use. The first is its long life, several thousand hours under normal operating conditions being common. At reduced filament power several hundred thousand hours have been obtained. The second is the fact that it can easily be manufactured in the form of the indirectly heated cathodes. These indirectly heated cathodes are considered in greater detail in the next section. The third is its ability to give tremendous outputs under pulsed conditions. Thus, it has been found that for (microsecond) pulses current densities in excess of $10^6$ amp/m$^2$ may be obtained.[13] Such pulsed cathodes are used in magnetrons in radar applications. Cathode sparking or flashing limits the peak emission current density that can be obtained. Such a spark results in a physical loss of coating.

Oxide-coated cathodes are subject to deactivation by positive-ion bombardment and so are generally used in low-voltage tubes only. Furthermore, it is rather difficult to obtain high vacuum with them, although the vacuum requirements may not be very rigorous. Also, the evaporation of the oxide to other electrodes in the tube may lead to secondary-emission difficulties. This is especially true in high-power tubes in which the anodes operate at rather high temperatures.

The emission properties of an oxide-coated cathode are influenced by many factors, for example, the proportion of the contributing oxides, the thickness of the oxide coating, possibly the core material, and the details of the processing. The determination of the relationship between the power input and the emissivity or temperature is very difficult. The various methods found in the literature are summarized by Blewett,[10] who concludes that for an ordinary oxide-coated filament ($4 \times 10^{-4}$ m thick)

the emissivity can be taken as a constant and equal to 0.3 over the normal operating range. On the other hand, Dushman [1] quotes an emissivity that ranges between 0.6 and 0.7 for the temperature range 800° to 1200°K for the oxide-coated filaments used by the Western Electric Co.

Unique values of $E_W$ and $A_0$ for oxide-coated cathodes cannot be given. However, the mass of experimental data available indicates that a rea-
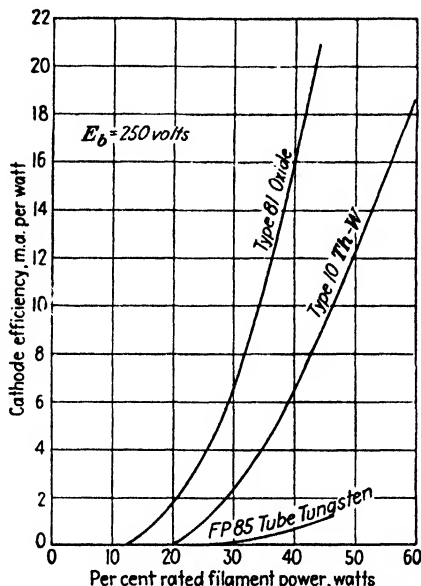


Fig. 6-3. Cathode efficiency curves for an oxide-coated filament, a thoriated-tungsten filament, and a pure tungsten filament.

sonable value for the work function is of the order of 1–1.5 volts and for $A_0$ of approximately $0.01 \times 10^4$ amp/$(m^2)(°K^2)$ as given in Table 6-1.

Curves showing the relative cathode efficiencies of tungsten, thoriated-tungsten, and oxide-coated cathodes are illustrated in Fig. 6-3. It will be noticed that tungsten has a considerably lower efficiency than either of the other two emitters. Since it is not possible to obtain saturation values of emission current for either the thoriated-tungsten or the oxide-coated cathodes, these curves were taken at a fixed plate voltage, $E_b = 250$ volts. Saturation currents from tungsten were possible with this value of $E_b$.

Since the emission characteristics vary so markedly from tube to tube, the reader may have wondered how tubes using these oxide-coated cathodes could serve satisfactorily in any circuit. The answer to this will become evident when the material of Chap. 7 is considered. It will there be shown that tubes usually operate under conditions of space-charge limita-

tion and not under conditions of temperature limitation. This means that *the current is determined by the plate voltage and not by the cathode temperature.* Thus, despite their rather unpredictable emission characteristics, oxide-coated cathodes make excellent tube elements provided only that their thermionic-emission current never falls below that required by the circuit.

Oxide-coated cathodes are used in the greatest percentage of commercial electron tubes. Almost all receiving tubes, many low-voltage transmitting tubes, and practically all gas tubes use such cathodes.

**6-5. Commercial Cathodes.** *Directly Heated Cathodes.* The cathodes used in thermionic tubes are frequently directly heated filaments in the
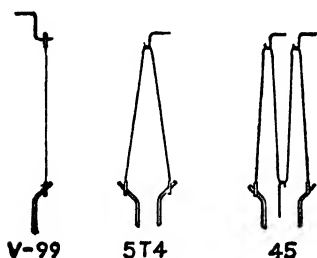
FIG. 6-4.  Typical  filamentary cathodes. (*Courtesy of Radio Corporation of America.*)
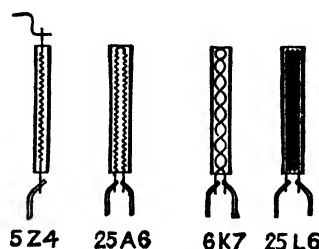
FIG.  6-5.  Typical  indirectly heated cathodes. (*Courtesy of Radio Corporation of America.*)

form of a V, a W, or a straight wire, although the modern tendency has been to use indirectly heated cathodes wherever possible. Typical directly heated cathodes are illustrated in Fig. 6-4.

Directly heated uncoated cathodes must be used for very high voltage systems, since at such voltages any coating will ultimately be destroyed by the vigorous bombardment of the few residual gas molecules that are always present in a tube.

*Indirectly Heated Cathodes.* This type of cathode was developed so as to minimize the hum arising from the various effects of a-c heater operation. The general forms of indirectly heated cathodes for use in vacuum tubes are illustrated in Fig. 6-5. The heater wire is contained in a ceramic insulator which is enclosed by a metal sleeve on which the oxide coating is placed. The cathode as a unit is so massive that its temperature does not vary appreciably with instantaneous variations in the magnitude of the heater currents. Further, since the sleeve is the cathode emitting surface, the cathode is essentially equipotential. The heater wire within the ceramic insulator is tungsten or an alloy of tungsten and molybdenum.

The ceramic insulator which acts to insulate the heater wire from the cathode must, of course, be a good heat conductor. Materials that are

extensively used for this purpose are the oxides of beryllium and aluminum. These materials must be free from any impurities [14] in order to avoid breakdown. The tungsten wire is usually sprayed with, or dipped into, a suspension of the oxide, to which an organic binder has been added. The coated heater is dried in air and is then heated to about 1400°K in a vacuum or in an atmosphere of hydrogen. A solid layer, approximately 0.5 mm thick, forms around the wire. The heater is then inserted into the nickel or Konal-metal sleeve. Under normal conditions of operation, the heater is maintained at about 1000°C, which results in the cathode temperature being at approximately 850°C.

If an excessive potential difference is maintained between the heater and the cathode, the thin layer of insulation may break down. This consideration is frequently of extreme importance in practice, where it may be desired to heat several cathodes that are at different potentials with respect to ground from the same heater circuit. The insulation between the heater and the cathode is considered satisfactory if it has a resistance of at least 10 megohms at normal operating temperature. Recommended practice is to limit the potential difference between the heater and the cathode to about 100 volts. If the difference in voltage between two cathodes is greater than this, then separate heater transformers must be used. .

**6-6. Heat-shielded Cathodes.** It was pointed out in Sec. 6-2 that most of the heat losses from the cathode of a tube result from the radiation of the heat energy. The amount of the heat radiated at a given temperature is proportional to the area of the emitting surface and the emissivity, according to Eq. (6-3). If either the emissivity of the surface or the area for radiation or both can be decreased, the efficiency of the cathode will be increased. Hull [15] accomplished more than simply decreasing these two terms. Not only did he decrease the effective area for losses by radiant energy, but he simultaneously increased the electron-emitting area so that currents in hundreds of amperes may be obtained from properly designed cathodes.

One form of heat-shielded cathode is so shaped as to resemble a furnace and is provided with oxide-coated vanes which extend radially outward from the center. Consequently, heat is radiated from vane to vane. That



FIG. 6-6. A heat-shielded cathode. (*Courtesy of General Electric Co.*)

such a cathode provides a large emitting surface is evident from an inspection of Fig. 6-6. The cathode is surrounded by a series of nickel cylinders which act as heat baffles, thereby providing for practically all the heat loss to occur from the top of the cylinder. This is a relatively small area com-

pared with the total surface area of the cylinder.  *This type of cathode has
a very large electron-emitting surface but a small effective radiating surface*,
which accounts for the increased efficiency of this "radiator" type of cath-
ode over that of the ordinary outward-radiating cathodes illustrated in
Fig. 6-5.

Another method of accomplishing a similar result is to wind a coated
cathode, in the form of a coated ribbon, either in the form of a spiral or in
a form that resembles the bellows of an accordion.  The adjacent turns
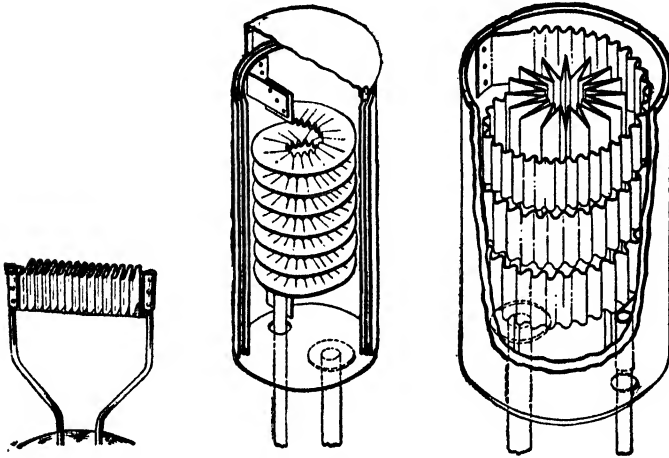


Fig. 6-7.   Three types of heat-shielded cathodes.   (*Lowry*.)

of the cathode will act as heat shields and so result in a decreased radia-
tion.  Cathodes of these types are illustrated in Fig. 6-7.[16]

Heat-shielded cathodes are used mainly in mercury-vapor rectifier
tubes of either the diode or the triode (thyratron) type.  They are not
used in vacuum tubes because in a vacuum tube the plate current is de-
termined by the plate voltage and not by the thermionic saturation cur-
rent (see Sec. 7-2).

## PROBLEMS

**6-1.** The tungsten filament of a General Electric Co. type FP 85 diode has a lighted
length of $3\frac{11}{16}$ in. and a diameter of 0.0085 in.

a. At what temperature does the filament operate if the power input is 40 watts?

b. What is the saturation current under these conditions?

c. Repeat parts a and b for power inputs of 20 and 5 watts.

**6-2.** Using the thermal curve for tungsten (Fig. 6-2), calculate the emissivity $e_T$.
Plot this parameter as a function of temperature.

**6-3.** A thermionic cathode is required to supply a temperature saturated current of
8 amp.

*a.* Compute the amount of filament power necessary if a tungsten cathode at 2500°K is to be used.

*b.* Repeat for a thoriated-tungsten cathode to be operated at 1850°K. Neglect the cooling effect discussed in Prob. 6-15.

**6-4.** If the emission from a certain cathode is 10,000 times as great at 2000° as at 1500°K, what is the work function of this surface?

**6-5.** *a.* If 30 watts is supplied to a thoriated-tungsten filament whose area is 2 cm², what is the emission current? Neglect the cooling effect discussed in Prob. 6-15.

*b.* What must the power input be in order to double the emission?

**6-6.** *a.* If the temperature of a tungsten filament is raised from 2300° to 2320°K, by what percentage will the emission change?

*b.* To what temperature must the filament be raised in order to double its emission at 2300°K?

**6-7.** By what percentage will the emission from a tungsten filament that is maintained at 2500°K change when the power input to the filament is changed by 5 per cent?

HINT: Use differentials to obtain fractional changes in current from the Dushman equation and fractional changes in power from the Stefan-Boltzmann equation.

**6-8.** At what temperature will a thoriated-tungsten filament give 5,000 times as much emission as a pure tungsten filament at the same temperature? The filament dimensions of the two emitters are the same.

**6-9.** *a.* The saturation current from a certain tungsten filament operated at 1840°K is 143 $\mu$a. What would be the emission from a thoriated-tungsten filament of the same area operating at the same temperature?

*b.* At what temperature would this thoriated-tungsten filament have to be operated in order to supply 14.3 ma?

**6-10.** At what temperature will an oxide-coated filament give five times as much emission as a pure tungsten filament of the same dimensions that is maintained at 2400°K? Take $E_W = 1.0$ ev; $A_0 = 100$ amp/(m²)(°K²) for the oxide-coated filament.

**6-11.** How much power must be supplied to an oxide-coated filament 1.8 cm² in area in order to maintain it at 1100°K? Assume that the heat loss due to conduction is 10 per cent of the radiation loss. Calculate the total emission current and the cathode efficiency of the cathode. Take $e_T = 0.7$; $E_W = 1.0$ ev; and $A_0 = 100$ amp/(m²)(°K²).

**6-12.** The type 45 triode is provided with a ribbon filament of the W type; this filament is 0.014 by 0.004 in., and it is 4.5 in. long. Calculate the filament temperature and the total cathode emission expected from this oxide-coated cathode when operated under normal filament power (1.5 amp at 2.5 volts). Take the emissivity to be 0.3, and assume the following values for the thermionic-emission constants: $E_W = 1.0$ ev and $A_0 = 100$ amp/(m²)(°K²).

**6-13.** A tungsten cathode is heated to a temperature of 2300°K. What must the retarding voltage be in order to limit the current density to $10^{-2}$ amp/m²? Neglect contact potential.

**6-14.** What field intensity must be applied to the surface of a cold tungsten metal in order to obtain a current density of $10^{-2}$ amp/m²?

HINT: Use the method of successive approximations outlined in connection with the thermionic-emission equation in Sec. 6-3.

**6-15.** Prove that the cooling of the filament of a thermionic cathode because of the emission current $I$ is due to the power loss $I$ $(E_W + 2E_T)$ watts. Refer to the energy level diagrams of Chap. 5 and consider (from the energy point of view) what happens to the electrons as they are emitted from the cathode; as they pass into space from cathode to anode; as they enter the anode; and finally as they return to the cathode through the plate battery. In particular, how much power must the plate dissipate?

How is the above picture modified if the cathode and anode are not the same material, so that a contact difference of potential exists between them?

**6-16.** *a.* Show that the Dushman equation $J = A_0 T^2 \epsilon^{-b_0/T}$ may be written in the form $J' = A'\epsilon^{-b'/T}$ and that $J = J'$ and $dJ/dT = dJ'/dT$ at $T = T_0$ provided that $b' = b_0 + 2T_0$ and $A' = A_0(\epsilon T_0)^2$.

*b.* Show that

$$T = \frac{b_0 + 2T_0}{2 + (2.303)[2 \log T_0 + \log (A_0/J)]}$$

The method of successive approximations may be applied to this equation and the convergence is very rapid. (H. F. Ivey and C. L. Shackelford, Westinghouse Electric Corp.)

### REFERENCES

1. DUSHMAN, S., *Revs. Modern Phys.*, **2**, 381, 1930.
2. For measurements on tungsten:
   DAVISSON, C., and L. H. GERMER, *Phys. Rev.*, **20**, 300, 1922.
   DUSHMAN, S., H. N. ROWE, J. W. EWALD, and C. A. KIDNER, *ibid.*, **25**, 338, 1925.
3. FORSYTHE, W. E., and A. G. WORTHING, *Astrophys. J.*, **61**, 146, 1925.
4. JONES, H. A., and I. LANGMUIR, *Gen. Elec. Rev.*, **30**, 354, 1927.
5. RICHARDSON, O. W., *Proc. Cambridge Phil. Soc.*, **11**, 286, 1901.
6. WEHNELT, A., *Ann. Physik*, **14**, 425, 1904.
7. LANGMUIR, I., and W. ROGERS, *Phys. Rev.*, **4**, 544, 1914.
8. MARSDEN, C. P., JR., *Electronics*, **11**, 22, 1938.
   KOLLER, L. R., "The Physics of Electron Tubes," 2d ed., p. 371, McGraw-Hill Book Company, Inc., New York, 1937.
9. DUSHMAN, S., and J. W. EWALD, *Phys. Rev.*, **29**, 857, 1927.
10. Excellent reviews of oxide-coated cathodes are given by:
    BLEWETT, J. P., Parts I and II, *J. Applied Phys.*, **10**, 668, 831, 1939; **17**, 643, 1946.
    EISENSTEIN, A. S., "Advances in Electronics," pp. 1–62, Academic Press, 1948.
    PRESCOTT, C. H., JR., and J. MORRISON, *Rev. Sci. Instruments*, **10**, 36, 1939.
11. SPANGENBERG, K. R., "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.
12. WRIGHT, D. A., *Proc. Phys. Soc. (London)*, **B62**, 188, 1949.
13. COOMES, E. H., *J. Applied Phys.*, **17**, 647, 1946.
    SPROULL, R. L., *Phys. Rev.*, **67**, 166, 1945.
14. KLEMPERER, H., *Elec. Eng.*, **55**, 981, 1936.
15. HULL, A. W., *Gen. Elec. Rev.*, **32**, 213, 1929.
16. LOWRY, E. F., *Electronics*, **6**, 280, October, 1933.

### General References

CHAFFEE, E. L.: "Theory of Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1933.
DOW, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.
EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," 3d ed., McGraw-Hill Book Company, Inc., New York, 1949.
HARNWELL, G. P., and J. J. LIVINGOOD: "Experimental Atomic Physics," McGraw-Hill Book Company, Inc., New York, 1933.

KOLLER, L. R.: "The Physics of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937.

McARTHUR, E. D.: "Electronics and Electron Tubes," John Wiley & Sons, Inc., New York, 1936.

MIT Staff: "Applied Electronics," John Wiley & Sons, Inc., New York, 1943.

REIMANN, A. L.: "Thermionic Emission," John Wiley & Sons, Inc., New York, 1934.

VAN DER BIJL, H. J.: "The Thermionic Vacuum Tube and Its Applications," McGraw-Hill Book Company, Inc., New York, 1920.

# CHAPTER 7

# DIODE CHARACTERISTICS

IN THE discussion of thermionic emission in the previous chapter, only the temperature saturated current is considered. This is the maximum current that can be obtained from a cathode at any particular temperature. This implies the existence of a collecting anode placed near to and in the same evacuated enclosure as the cathode. This chapter contains a discussion of the phenomena encountered when the anode potential is not high enough to collect the entire thermionic saturation current.

This chapter also discusses some of the properties of metallic diodes which possess rectifier characteristics which are roughly similar to those of the vacuum diode.

Gaseous diodes are not under consideration here, but are discussed in Chap. 11.

**7-1. The Potential Distribution between the Electrodes.** Consider a simple thermionic diode, the cathode of which can be heated to any desired temperature and in which the potential between the cathode and the anode may be maintained at any desired value. It will be assumed for the present that the cathode is a plane equipotential surface and that the collecting anode is also a plane which is parallel to it. The potential distribution between the electrodes is to be investigated for a given value of anode voltage.

It is assumed that the electrons are emitted from the cathode with zero initial velocities. Under these circumstances, the curves showing the variation of potential in the interelectrode space for various temperatures of the cathode are given in Fig. 7-1. The general shape of these curves may be explained as follows: At the temperature $T_1$ at which no electrons are emitted, the potential gradient is constant, so that the potential distribution is a linear function of the distance from the cathode to the anode. Use was made of this linear potential distribution in Chap. 2, where the motion of an electron in a constant electric field of force was under consideration.

At the higher temperature $T_2$, an appreciable density of electrons exists in the interelectrode space. The potential distribution will be somewhat as illustrated by the curve marked $T_2$ in Fig. 7-1. The increase in tem-

perature can change neither the potential of the cathode nor the potential of the anode. Hence, all the curves must pass through the fixed end points $K$ and $A$. Since negative charge (electrons) now exists in the space between $K$ and $A$, then, by Coulomb's law, the potential at any point will be lowered. The greater the space charge, the lower will be the potential. Hence, as the temperature is increased, the potential curves become more and more concave upward. At $T_3$, the curve has drooped so far that it is tangent to the $X$ axis at the origin. That is, the electric-field intensity
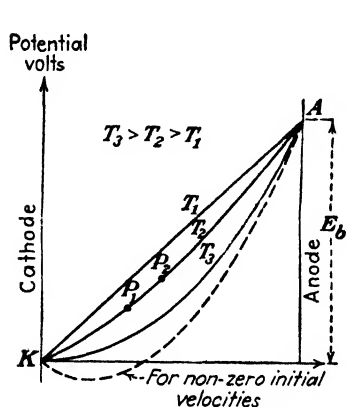


FIG. 7-1. The potential distribution between plane-parallel electrodes for several values of cathode temperature.
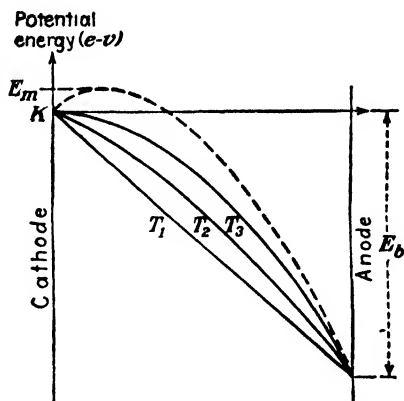


FIG. 7-2. The potential-energy distributions corresponding to the curves of Fig. 7-1.

at the cathode for this condition is zero. One may sketch the broken curve of Fig. 7-1 to represent the potential variation at a temperature higher than $T_3$. This curve, however, contains a potential minimum. Such a condition is physically impossible, if the initial velocities of the emitted electrons are assumed negligible. That this is so follows from the discussion given below.

Consider the potential-energy curves corresponding to Fig. 7-1. Since the potential energy is equal to the product of the potential $V$ and the charge $-e$, then the curves of Fig. 7-2 are simply those of Fig. 7-1 inverted, the unit of the ordinates being changed to electron volts. It is immediately evident that the broken curve represents a potential-energy barrier at the surface of the cathode. Several such potential-energy barriers have already been considered in Chap. 4. On the basis of our previous discussions, it is clear that only those electrons which possess an initial energy greater than $E_m$, the maximum height of the barrier, can escape from the cathode and reach the anode. Consequently, the assumed con-

dition of zero initial velocities of the emitted electrons precludes the possibility of any electrons being emitted. As a result, the barrier will be broken down, since the applied field will cause those electrons which produce the barrier to leave the interelectrode space and become part of the anode current. This automatic growth and collapse of the potential barrier outside of the cathode may be considered as a self-regulating valve that allows a certain definite number of electrons per second to escape from the cathode and reach the anode, for a given value of plate voltage.

It can be inferred from the foregoing argument that the maximum current that can be drawn from a diode for a fixed plate voltage and any temperature whatsoever is obtained under the condition of zero electric field at the surface of the cathode. Thus, for optimum conditions,

$$\mathcal{E} = -\frac{dV}{dx} = 0 \qquad \text{at } x = 0 \tag{7-1}$$

This condition is based on the assumption that the emitted electrons have zero initial energies. This is not strictly true, of course. Because of these initial velocities, the potential distribution within the tube may actually acquire the form illustrated by the broken curve of Fig. 7-1.

The effect of the initial velocities is equivalent to the establishment of a negative potential gradient at the cathode. The effect of this negative potential gradient will be that of a force tending to return the emitted electrons to the cathode. That is, this negative potential gradient which corresponds to a potential-energy barrier tends to repel those electrons which are emitted with small initial velocities. Evidently, the formation of such a barrier, which depends for its existence upon the presence of space charge near the cathode, tends to suppress the space charge that sustains it. Consequently, an equilibrium condition must exist between the emission of the electrons and the space charge. This means that a very slight potential barrier will exist near the surface of the cathode and that the anode actually draws the plate current from this negative space-charge region. Since the potential minimum in Fig. 7-1 is usually small in comparison with the applied potential, it will be neglected and the condition (7-1) will be assumed to represent the true status when space-charge current is being drawn.

In the next section the analytical expression for the potential as a function of distance will be derived. The starting point in this derivation is Poisson's equation, which relates the density of electrons with the potential at any point in the interelectrode space. The derivation of this equation is given in Appendix VI. It is simply a mathematical restatement of Coulomb's law relating the potential with the charge, except that the space-charge cloud is treated as a continuous volume density of charge rather than an ensemble of discrete point charges.

Poisson's equation is

$$\frac{d^2V}{dx^2} = \frac{\rho}{\epsilon_0} \qquad (7\text{-}2)$$

where $x$ is the distance from the cathode in meters, $V$ is the potential in volts, $\rho$ is the *magnitude* of the *electronic* volume charge density in coulombs per m$^3$, and $\epsilon_0$ is the permittivity of free space in the mks system.

The shapes of the curves in Fig. 7-1 conform to this equation. Thus, for $T = T_1$ where $\rho = 0$, Eq. (7-2) becomes

$$\frac{d^2V}{dx^2} = 0 \qquad \text{or} \qquad \frac{dV}{dx} = \text{const}$$

This is the equation of a straight line as shown in Fig. 7-1. Furthermore, the curve for any other temperature must be concave upward. This follows from Eq. (7-2), which states that $d^2V/dx^2$ is a positive number. A positive second derivative means that the change in slope $dV/dx$, between two adjacent points, must be positive.

It is readily verified that

$$\frac{dV}{dx}\bigg]_{P_2} - \frac{dV}{dx}\bigg]_{P_1}$$

is positive for any two neighboring points $P_1$ and $P_2$ of Fig. 7-1. Further, the change in slope is greater for larger values of $\rho$, corresponding to higher temperatures.

**7-2. Equations of Space Charge.** It was shown in Sec. 2-10 that if there are $N$ electrons per cubic meter, each of which carries a charge $e$ coulombs, then the motion of these electrons constitutes a current, the current density of which is given by

$$J = Nev = \rho v \qquad \text{amp/m}^2 \qquad (7\text{-}3)$$

where $v$ is the drift velocity of these electrons in meters per second and where $\rho \equiv Ne$ is the volume density of electric charge in coulombs per cubic meter. Both $\rho$ and $v$ are functions of the distance from the origin (the cathode). However, the product is constant, since the number of electrons passing through unit area per second must be the same for all points between a plane cathode and a parallel anode. This is actually an expression for the principle of *conservation of electric charge*. Therefore, at the cathode, where the velocity of the electrons is very small (the velocities being the initial velocities), the charge density must be very large. In the neighborhood of the anode, the velocity is a maximum, hence the charge density is a minimum. The initial velocities being neglected, the velocity of the electrons at any point in the interelectrode space may be determined

from the equation that relates the kinetic energy of the particle with the potential through which it has fallen, *viz.*,

$$\tfrac{1}{2}mv^2 = eV \tag{7-4}$$

which has been used many times before.

There results, from Eqs. (7-2), (7-3), and (7-4),

$$\frac{d^2V}{dx^2} = \frac{\rho}{\epsilon_0} = \frac{J}{v\epsilon_0} = \frac{J}{\left(2\dfrac{e}{m}\right)^{\frac{1}{2}}\epsilon_0} V^{-\frac{1}{2}} \tag{7-5}$$

In order to simplify the form of this equation, a transformation of the variable is made, by writing

$$V = ky \tag{7-6}$$

where $y$ is to be determined as a function of $x$ and where $k$ is a constant of proportionality that is independent of $x$. This equation defines $y$ for any given $k$. The most convenient value of $k$ will be chosen below. Under the specified transformation, Eq. (7-5) becomes

$$k\frac{d^2y}{dx^2} = \frac{Jk^{-\frac{1}{2}}}{\left(2\dfrac{e}{m}\right)^{\frac{1}{2}}\epsilon_0} y^{-\frac{1}{2}}$$

which reduces to the very simple form

$$\frac{d^2y}{dx^2} = y^{-\frac{1}{2}} \tag{7-7}$$

provided that $k$ is chosen so that

$$k^{\frac{3}{2}} = \frac{J}{\left(2\dfrac{e}{m}\right)^{\frac{1}{2}}\epsilon_0} \tag{7-8}$$

This change of variable has done considerably more than merely transform Eq. (7-5) into the simpler equation (7-7). It has transformed a differential equation in potential vs. distance with current density as a parameter to Eq. (7-7) which has been stripped of all parameters, the dependence upon current density having been transferred to Eq. (7-8). Thus it follows from (7-8) that

$$J = \left(2\frac{e}{m}\right)^{\frac{1}{2}}\epsilon_0 k^{\frac{3}{2}} = \left(2\frac{e}{m}\right)^{\frac{1}{2}}\epsilon_0 \frac{V^{\frac{3}{2}}}{y^{\frac{3}{2}}} \tag{7-9}$$

At the anode, $V = E_b$ and $y$ is a constant (independent of $E_b$). Therefore, *the plate current varies as the three-halves power of the plate potential.*

Note that it has been possible to obtain this result without solving the nonlinear differential equation (7-7).

In order to determine the dependence of the current on the interelectrode spacing, it is necessary to solve Eq. (7-7) for $y$ as a function of $x$. This is a second-order differential equation, so that two boundary conditions must exist. These conditions are already known, $viz.$: the potential $V = ky$ vanishes at the cathode ($x = 0$). Also, from Eq. (7-1) the electric-field intensity is zero at the cathode. That is,

$$y = 0 \quad \text{and} \quad \frac{dy}{dx} = 0 \quad \text{at } x = 0 \quad (7\text{-}10)$$

The solution of Eq. (7-7) is readily accomplished. Let $z = dy/dx$, whence the equation becomes

$$\frac{dz}{dx} = y^{-\frac{1}{2}}$$

or

$$dz = y^{-\frac{1}{2}} \, dx = y^{-\frac{1}{2}} \frac{dy}{z}$$

from which there results

$$z \, dz = y^{-\frac{1}{2}} \, dy$$

This integrates to

$$\frac{z^2}{2} = 2y^{\frac{1}{2}}$$

The constant of integration is zero since the boundary conditions (7-10) require that $y = z = 0$ at $x = 0$. By taking the root of the last equation, there results

$$z = \frac{dy}{dx} = 2y^{\frac{1}{4}}$$

and

$$y^{-\frac{1}{4}} \, dy = 2 \, dx$$

which integrates to

$$\tfrac{4}{3} y^{\frac{3}{4}} = 2x$$

The constant of integration is zero by virtue of Eq. (7-10). Finally,

$$y = (\tfrac{3}{2} x)^{\frac{4}{3}} \quad (7\text{-}11)$$

which is the desired functional dependence of $y$ upon $x$.

It is seen that the potential ($V = ky$) depends upon the four-thirds power of the interelectrode spacing. For example, the curve marked $T_3$ in Fig. 7-1 is expressed by the relation

$$V = \alpha x^{\frac{4}{3}} \quad (7\text{-}12)$$

where $\alpha$ is readily found in terms of constants and the current density $J$ from the foregoing equations. However, $\alpha$ may also be written as $E_b/d^{\frac{1}{3}}$, where $d$ is the separation of the electrodes and $E_b$ is the plate potential. This is so because Eq. (7-12) is valid for the entire interelectrode space, including the boundary $x = d$ where $V = E_b$.

The complete expression for the current density is obtained by combining Eqs. (7-9) and (7-11). The result is

$$ J = \frac{4}{9}\left(2\frac{e}{m}\right)^{\frac{1}{2}}\epsilon_0\frac{V^{\frac{3}{2}}}{x^2} \qquad \text{amp/m}^2 \qquad (7\text{-}13) $$

In terms of the boundary values, this becomes, upon inserting the value of $e/m$ for electrons and $\epsilon_0 = 10^{-9}/36\pi$,

$$ J = 2.33 \times 10^{-6}\frac{E_b^{\frac{3}{2}}}{d^2} \qquad \text{amp/m}^2 \qquad (7\text{-}14) $$

This equation was first considered in detail by Langmuir,[1] although it had been previously published in a different connection by Child.[2] It is known by several different names, for example, the *Langmuir-Child law*, the *three-halves power law*, or simply the *space-charge equation*.

It will be noticed that this equation relates the current density, and so the current, in terms only of the applied potential and the geometry of the tube. The space-charge current does not depend upon either the temperature or the work function of the cathode. Hence, no matter how many electrons a cathode may be able to supply, the geometry of the tube and the potential applied thereto will determine the maximum current that can be collected by the anode. Of course, it may be less than the value predicted by Eq. (7-14), if the electron supply from the cathode is restricted (because the temperature is too low). To summarize, *the plate current in a given diode depends only upon the applied potential*, provided that this current is less than the temperature-limited current.

The velocity of the electrons as a function of position between the cathode and anode can be found from Eq. (7-4) with the aid of Eq. (7-13). Then the charge density as a function of $x$ can be found from Eq. (7-3). It is easily found (see Prob. 7-1) that $v$ varies as the two-thirds power of $x$ and that $\rho$ varies inversely as the two-thirds power of $x$. The latter leads to the physically impossible result that at the cathode the charge density is infinite. This is a consequence of the assumption that the electrons emerging from the cathode all do so with zero initial velocity. Actually, of course, the initial velocities are small, but finite, and the charge density is large, though finite.

It will be found that conditions in certain portions of a gaseous discharge (see Sec. 10-19 for a discussion of sheaths) are precisely the same as those

discussed above, except that the current-carrying particles are positive ions of mass $m_i$ instead of electrons of mass $m_e$. Under these conditions, the space-charge equations must be modified in order to take into account this difference in mass and, if any, the difference in charge that is carried by the ion. Equation (7-13) is to be replaced in these cases by the expression

$$J = 2.33 \times 10^{-6} \sqrt{\frac{(e/m)_i}{(e/m)_e}} \frac{V^{\frac{3}{2}}}{x^2} \qquad \text{amp/m}^2 \qquad (7\text{-}15)$$

**7-3. Space Charge in Tubes Possessing Cylindrical Symmetry.** Systems that possess plane-parallel electrodes were considered above because the simplicity of this geometry made it easy to understand the physical principles involved. However, such tube geometry is almost never met in practice. More frequently, tubes are constructed with cylindrical symmetry, the anode being in the form of a cylinder that is coaxial with a cathode of either the directly or the indirectly heated type. The solution of the space-charge equation for such a system, subject to the conditions that field distortion due to the end effects is negligible, that both the anode and the cathode are equipotential surfaces, and that the electrons emerge from the cathode with zero initial velocities, is possible.

The solution is found to be [1]

$$I_b = \frac{8\pi \left(2 \dfrac{e}{m}\right)^{\frac{1}{2}} \epsilon_0}{9} \frac{l}{r_a} \frac{E_b^{\frac{3}{2}}}{\beta^2} \qquad \text{amp} \qquad (7\text{-}16)$$

In this expression $E_b$ is the plate voltage, $r_a$ is the plate radius, $l$ is the length of the plate, and $\beta^2$ is determined from the ratio $r_a/r_k$ from Fig. 7-3. For electrons, this expression reduces to

$$I_b = 14.6 \times 10^{-6} \frac{l}{r_a} \frac{E_b^{\frac{3}{2}}}{\beta^2} \qquad \text{amp} \qquad (7\text{-}17)$$

A comparison of Eqs. (7-14) and (7-17) reveals that the current varies inversely as the square of the distance between the electrodes in the case of plane-parallel electrodes and varies inversely as the first power of the anode radius in the case of cylindrical electrodes, provided that the ratio $r_a/r_k$ is sufficiently large so that $\beta^2$ may be considered approximately equal to unity.

Attention is called to the fact that the plate current depends upon the three-halves power of the plate potential for both the plane parallel and the system possessing cylindrical symmetry. This is a general relationship, since it is possible to demonstrate [3] that *an expression of the form $I_b = k_1 E_b^{\frac{3}{2}}$, where $I_b$ is the plate current, applies for any geometrical arrange-*
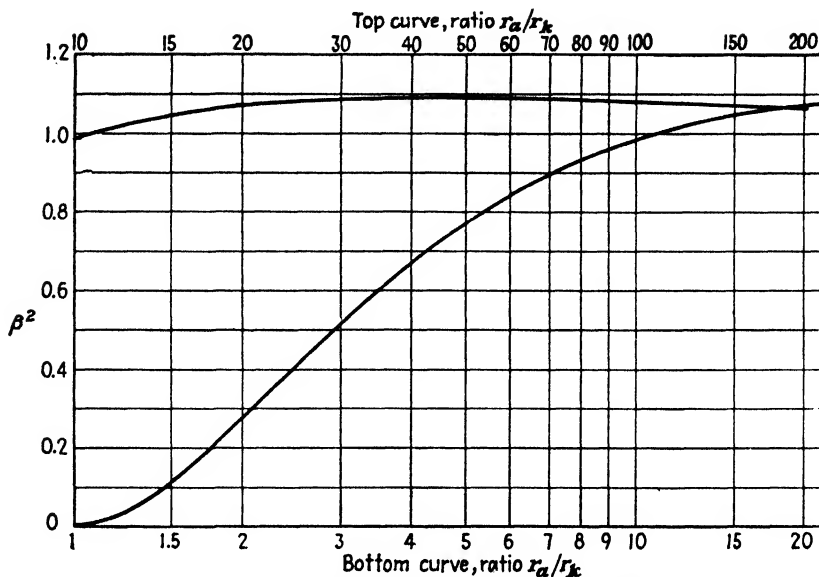
FIG. 7-3. Langmuir's space-charge parameter $\beta^2$.

*ment of cathode and anode*, provided that the same restrictions as imposed in the foregoing analyses are true. The specific value of the constant $k_1$ that exists in this expression cannot be analytically determined unless the geometry of the system is specified.

*Example.* Given a cylindrical cathode 1.8 cm long and 0.16 cm in diameter whose thermionic emission cannot exceed 22 ma/cm at the operating temperature. An anode 1.0 cm in diameter is coaxial with this cathode. What is the plate current when the plate potential is 50 volts? If the plate potential is increased to 100 volts, what is the new value of plate current?

*Solution.* The thermionic saturation current is $22 \times 1.8 = 39.6$ ma. The space-charge-limited current is calculated from Eq. (7-17). The value of $\beta^2$ corresponding to $r_a/r_k = 0.5$ cm/0.08 cm $= 6.24$ is found from Fig. 7-3 to be 0.85. Hence

$$I_b = 14.6 \times 10^{-6} \frac{l}{r_a} \frac{E_b^{\frac{3}{2}}}{\beta^2} = (14.6 \times 10^{-6}) \left(\frac{1.8}{0.5}\right) \left(\frac{50^{\frac{3}{2}}}{0.85}\right) = 21.8 \text{ ma}$$

Since this is less than the thermionic current, then the plate current will be the space-charge current, of 21.8 ma.

For $E_b = 100$ volts, which is double the voltage used above, the three-halves-power current will be $2^{\frac{3}{2}}$ times the value just calculated, or

$$I_b = 2^{\frac{3}{2}} \times 21.8 = 61.7 \text{ ma}$$

Since this exceeds the thermionic current, then the plate current will be the thermionic current, or 39.6 ma.

This example emphasizes the fact that, if the thermionic and space-charge currents are calculated, the plate current will always equal the smaller of these two values.

**7-4. Factors Influencing Space-charge Current.** Several factors modify to a great extent the equations for space charge given above. This is particularly true at low plate voltages. Among these factors are:

1. *Filament Voltage Drop.* The space-charge equations were derived on the assumption that the cathode is an equipotential surface. This is not a valid assumption for a directly heated emitter, and the voltage across the ends of the filament causes a deviation from the three-halves-power equation. In fact, the results depend on whether the plate current is returned to the positive or to the negative end of the filament.

If the plate current is returned to the negative end of the filament, the actual voltage of the various portions of the filament is less than the applied voltage, because of the repelling action of the filament voltage. Hence, the space-charge current is less than that which can be obtained from a geometrically similar equipotential cathode. If the cathode lead is returned to the positive end of the filament, an increased current is obtained. The effect of filament voltage drop is largely eliminated by returning the plate current to the filament mid-point, if this is available, or to a center-tapped resistor across the filament. If the filament is heated with a transformer, the plate is returned to the center tap of the secondary winding.

The voltage drop along the filament is nonuniform for several reasons. The resistance per unit length may not be constant, since the cross section may not be uniform. Also, the ends of the filament are naturally cooler than the center portion. These factors also cause deviations from the three-halves-power law.

2. *Contact Potential.* In every space-charge equation, the symbol $E_b$ must be understood to mean the sum of the applied voltage from plate to cathode plus the contact potential between the two. For plate voltages of only a few volts, this effect may be quite appreciable.

3. *Asymmetries in Tube Structure.* Commercial tubes seldom possess the ideal geometry assumed in deriving the space-charge equations.

4. *Gas.* The presence of even minute traces of gas in a tube can have marked effects on the tube characteristics. As soon as the voltage is sufficiently high to cause ionization of the residual gas molecules, the plate current will rise above that demanded by the space-charge equations. This arises because the positive ions that are formed neutralize the electronic-charge density. Modern vacuum tubes are exhausted to pressures of about $10^{-6}$ mm Hg, at which pressures the effects of residual gas are entirely negligible.

Residual gas in a vacuum tube may also have a deleterious effect on the thermionic emission itself. This arises from the fact that most gases react chemically with the hot cathode and (also in other ways) decrease the emission current.[1] Further, positive-ion bombardment of a cathode, par-

ticularly of the thoriated-tungsten or the oxide-coated types, may completely ruin the thermionic properties of these emitters.

5. *Initial Velocities of the Emitted Electrons.* It should now be evident that if the initial energies with which the electrons are emitted from a heated metal are taken into account the equations derived in Sec. 7-2 will have to be modified.

If the initial velocities of the electrons are not neglected, then the equilibrium condition will no longer be that of zero electric field at the cathode. Instead, the variation of potential with interelectrode spacing will be some-



Fig. 7-4. The potential distribution in a plane-parallel space-charge diode, taking into account the initial velocities of the electrons.

what as depicted by the broken curve of Fig. 7-1, which is reproduced in Fig. 7-4 for convenience. This represents a potential-energy barrier at the cathode surface, and so it is only those electrons whose energies are greater than the height $E_m$ of this barrier that can escape from the cathode. The height of this barrier is, from the results of Sec. 4-10, a fraction of 1 ev.

At a distance $x_m$ from the surface of the thermionic emitter, the point of the potential minimum, the electric-field intensity passes through zero. Hence the point $M$ may be considered as the position of a "virtual" cathode. Evidently, the distance that will enter into the resulting space-charge equation will be $d - x_m$, and not $d$. Likewise, the effective potential at any point in the interelectrode space will be $E + E_m$, and not $E$ alone. Both of these factors will tend to increase the current above that which exists when the initial velocities are neglected.

Another effect must be considered, in addition to the creation of a virtual cathode at the point $M$. The electrons that arise from the virtual emitter at $M$ do not possess zero velocity but are distributed according to the Maxwell-Boltzmann distribution function. This should be taken into account when solving Poisson's equation. The presence of the high-speed electrons directed toward the anode will permit higher currents to be obtained than are predicted by the three-halves-power law.

The exact mathematical formulation of the space-charge equation taking into account the Maxwellian distribution of the electrons is somewhat involved. It can be found in the literature.[4]

To summarize, the plate current in a diode is not strictly a function only of the plate potential but does depend, to a small extent, upon the temperature of the cathode.

If all the factors discussed above are neglected, theory predicts that a plot of log $I_b$ versus log $E_b$ should be a straight line having a slope equal to

1.5, since from the equation $I_b = k_1 E_b^{\frac{3}{2}}$ one has

$$\log_{10} I_b = \log_{10} k_1 + \tfrac{3}{2} \log_{10} E_b$$

Dishington [5] has determined the space-charge coefficient for 22 types of commercial diode. The values ranged from 1.24 to 1.49 with the average at 1.42.

**7-5. Diode Characteristics.** The discussion in this and in the preceding chapter has revealed that the two most important factors that determine the characteristics of diodes are thermionic emission and space charge.
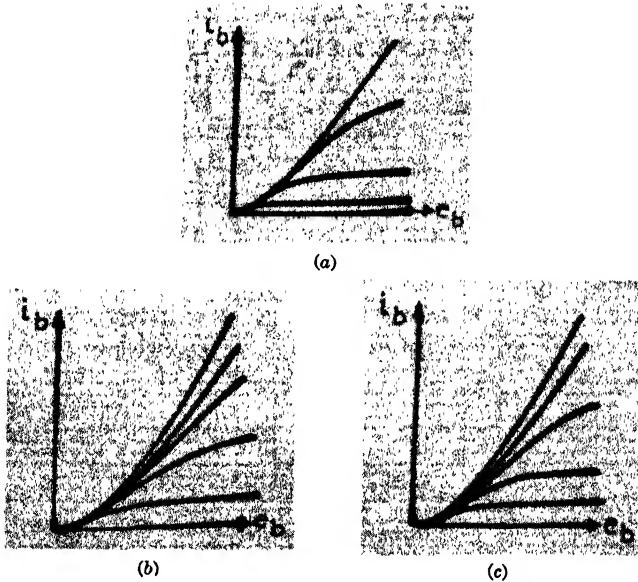


Fig. 7-5. Oscillograms showing volt-ampere diode characteristics for various filament temperatures. (*a*) Type FP 85. Tungsten filament. (*b*) Type 10. Thoriated-tungsten filament. Plate and grid connected together. (*c*) Type 81. Oxide-coated filament.

The first gives the temperature saturated value, *i.e.*, the maximum current that can be collected at a given cathode temperature, regardless of the magnitude of the applied accelerating potential. (This statement is strictly true only if the Schottky effect is neglected.) The second gives the space-charge-limited value, or the voltage saturated value, and specifies the maximum current that can be collected at a given voltage regardless of the temperature of the filament.

The volt-ampere characteristics for the types 10, 81, and FP 85 tubes are shown in the oscillograms of Fig. 7-5. It should be noted that the space-charge currents corresponding to the different temperatures do not, in gen-

eral, all coincide but that the currents decrease slightly as the temperature decreases. Further, there is no abrupt transition between the space-charge-limited and the temperature-limited portions of the curves, but rather a gradual transition occurs. Also, the current for the temperature-limited regions gradually rises with increased anode potentials. The latter is particularly true for the thoriated and oxide-coated cathodes.

The temperature-current curves that were obtained experimentally for a General Electric FP 85 tube are illustrated in Fig. 7-6. This particular
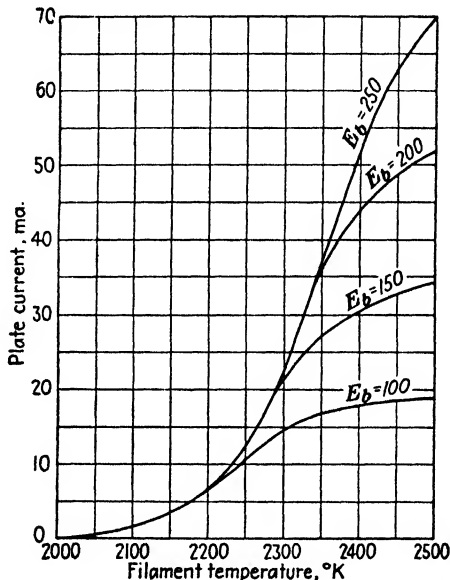


FIG. 7-6. Temperature-current curves of an FP 85 tungsten-filament diode with the plate voltage as a parameter.

tube has been chosen because it contains a directly heated tungsten filament, the temperature of which can be determined by the method outlined in Sec. 6-2. The curve for $E_b = 250$ volts is actually a plot of the Dushman equation for temperatures up to about 2400°K. Beyond this temperature, the curve shows a changing curvature that indicates that the current is becoming space-charge-limited. That is, the plate voltage is not high enough to collect saturation current. The bend in the curves is much more noticeable for the lower plate voltages. It should again be noted that there is no abrupt change from the temperature-limited to the space-charge-limited regions. As a matter of fact, the true space-charge current is not reached at all with the higher voltages and is only approached for $E_b = 100$ volts, as is evident from the curves.

The explanation of the shapes of the diode characteristics lies in a consideration of the numerous contributing factors. For example, in the case of a directly heated cathode, the drop in potential along the cathode will cause the different portions of the cathode to be at different values of voltage saturation. Another appreciable factor affecting the shape of the curves arises from the fact that the filament is not all at the same temperature. The consequence of the different temperatures existing at different parts of the cathode will result in temperature-limited currents that are limited at different values of temperature. Evidently the temperature effects are less pronounced when indirectly heated cathodes are used.

The application of the high anode potentials in order to collect the space current will give rise to an appreciable increase in current because of the Schottky effect. This effect may become particularly marked in the case of thoriated-tungsten and oxide-coated cathodes, since hot spots may result, the consequent thermionic emission being increased. In fact, it frequently happens that the diode characteristics of a tube containing a cathode having a composite surface will change appreciably after the application of a high plate voltage. In particular with oxide-coated filaments, the emission depends very markedly upon the past history of the tube.

In the case of filaments in the form of a V or W (hairpin filaments, in general), since adjacent portions of the filament are at different potentials with respect to each other, conduction may take place between portions of the same filament.

**7-6. Rating of Vacuum Diodes.** The rating of a vacuum diode, *i.e.*, the maximum current that it may normally carry and the maximum potential difference that may be applied between the cathode and the anode, is influenced by a number of factors.

1. A definite limitation is set by the *cathode efficiency* (see Sec. 6-4), which is the ratio of the emission current (in milliamperes) to the heating power of the cathode (in watts).

2. The temperature to which the glass envelope of the tube may be safely allowed to rise also furnishes a limitation to the normal rating. In order that the gas adsorbed by the glass walls should not be liberated, the temperature of the envelope must not be allowed to exceed the temperature to which the tube was raised in the outgassing process. As already mentioned, this is about 400°C for soft glass and about 600°C for pyrex glass. Because of this limitation, glass bulbs are seldom used for vacuum tubes of more than about 1 kw capacity, where the service is to be continuous.

3. Probably the most important factor limiting the rating of a tube is the allowable temperature rise of the anode. When a diode is in operation, the anode becomes heated to a rather high temperature. One of the fac-

tors that tends to increase its temperature is the power that must be dissipated by the anode. This may be considered either in terms of the kinetic energy carried by the electrons and transferred to the plate when they are collected, or in terms of the product of the plate current and the plate potential. The instantaneous power may be expressed in either of the forms

$$\tfrac{1}{2}Nmv^2 = Nee_b = i_b e_b \qquad \text{watts}$$

where $N$ is the number of electrons per second carried by the beam, $e$ is the charge of the electron in coulombs, $m$ is the electronic mass in kilograms, $v$ is the electronic velocity at the anode in meters per second, $i_b$ is the instantaneous plate current in amperes, and $e_b$ is the instantaneous plate potential in volts. In addition to the power carried by the anode current, the anode will also be heated to some extent by the heat that it intercepts from the cathode. This term depends upon the degree to which the anode surrounds the cathode and so intercepts heat radiation from it.

The temperature of the anode will rise until the rate at which the energy supplied to the anode from all sources just equals the rate at which the heat is dissipated from the anode in the form of radiation. The temperature to which the anode will rise will be determined by the power supplied to it, in accordance with Eq. (6-5). Consequently, for a given power supplied to the anode, the temperature to which it will rise will depend upon the area of the anode and the material of which it is constructed. The latter factor will determine the emissivity coefficient appearing in Eq. (6-5). The most common metals used for anodes are nickel and iron for receiving tubes and tantalum, molybdenum, and graphite for transmitting tubes. The surfaces are often roughened or blackened in order to increase the thermal emissivity. This permits higher power operation. These anodes may be operated at a cherry-red heat without excessive gas emission or other deleterious effects. For the larger tubes, it is necessary that the anodes be cooled. In the older type tubes, this is accomplished by circulating water through special cooling coils or, in special cases, by immersing the tube in oil, the temperature of the oil being maintained at a low value by means of water-cooling interchangers. Current practice, however, is to use forced air cooling on radiator fins.[6] These fins are attached to the anode and provide a large radiating surface in a relatively small space.

The design of the external anodes for high-power tubes was made possible by the development of metal-to-glass seals. It is necessary that the expansion and contraction of these seals be such that no excessive strains are introduced during operation, so that neither the glass nor the seal be ruptured. Although the glass and the metal may not have the same coefficients of expansion, the glass and metal are so proportioned, the metal being quite thin at the seal, that the seal can withstand the differential expansion that may arise. Another type of seal exists which has the same

thermal coefficient of expansion as certain glasses. These seals are made of alloys, generally consisting of iron, nickel, and cobalt.

4. The voltage limitation of a high-vacuum diode is not always determined by the permissible heating of the anode. For the case of a tube in which the filament and anode leads are brought out side by side through the same glass press, conduction may take place between the filament leads and the anode lead through the glass itself, especially if the voltage between these leads is high, as is true in a high-voltage rectifier. This effect is particularly marked if the glass is hot, and the resulting electrolysis will cause the glass to deteriorate and eventually leak. The highest voltage permissible between adjacent leads in glass depends upon the spacing of the leads and upon the type of glass but is generally of the order of 500 to 1,000 volts. For this reason, rectifiers are generally provided with filament leads and the anode lead at opposite ends of the glass envelope.

Also, the separation of the leads of high-voltage rectifiers must be large enough to preclude the possibility of flashover through the air. In fact, it is the highest voltage that may be safely impressed across the electrodes with no flow of charge that determines the safe voltage rating of a tube. Since, with an a-c potential applied between the cathode and anode, no current will exist during the portion of the cycle when the anode is negative with respect to the cathode, the maximum safe rating of a rectifying diode is known as the *peak inverse-voltage* rating.

Commercial vacuum diodes are made to rectify currents at very high voltages, up to about 200,000 volts. Such units are used with X-ray equipment, high-voltage cable testing equipment, and high-voltage equipment for nuclear-physics research. The dimensions and shapes of the glass envelopes of the diodes depend upon the current capacity of the tube, the shape of the anode, and whether the tube is to be air- or oil-cooled. Figure 7-7 is a photograph of a group of such tubes.

As was pointed out in Sec. 6-4, all high-voltage vacuum tubes have pure tungsten filaments. It is only for tubes that operate at about 5,000 volts or less that thoriated-tungsten filaments may be used. Oxide-coated cathodes are generally used in rectifier tubes that operate at about 750 volts or less and so find extensive application in home radio receiving sets.

**7-7. Metallic Rectifiers.**[7] A metallic rectifier cell comprises a combination of substances in the form of a sandwich. These sandwich combinations consist of a metal base plate, a semiconductor in intimate physical contact with this base plate, and a contact surface on the semiconductor. When the cell is formed, a thin *contact, barrier,* or *blocking layer* is produced between the base plate and the semiconductor. Such a cell possesses unilateral properties, there being a low resistance to current flow in one direction and a high resistance to current flow in the opposite direction. The rectifying action takes place in the blocking layer. The
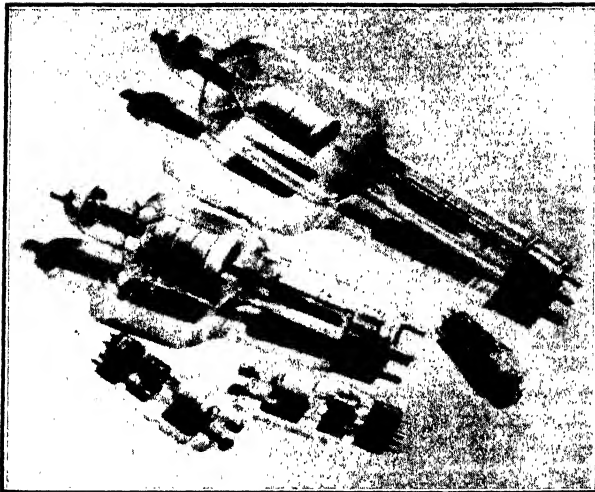
FIG. 7-7. A group of high-vacuum diodes. The smallest tube is the FP-400, which has a maximum anode dissipation of 15 watts, a typical operating anode voltage of 100 volts, and operating current of 0.025 amp. The largest tube is the GL-5625/KC-4, which has the following ratings: Peak inverse voltage, 150,000 volts; peak anode current, 1 amp; peak plate dissipation, 750 watts. (*Courtesy of General Electric Co.*)

general physical characteristics of such a cell are illustrated in Fig. 7-8. Such cells permit the ready flow of electrons from the base plate through the blocking layer to the semiconductor and then to the contact surface. Electron flow in the reverse direction is considerably impeded. In terms of the direction of conventional current flow, this means that the direction from semiconductor to base plate is the low-resistance direction, the reverse direction being a high-resistance path.
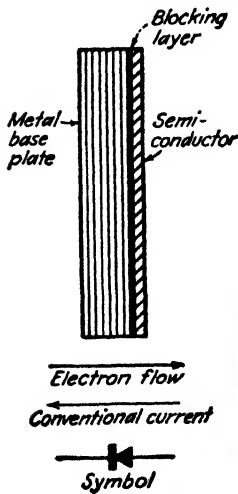


FIG. 7-8. The elements of a blocking-layer metallic rectifier cell.
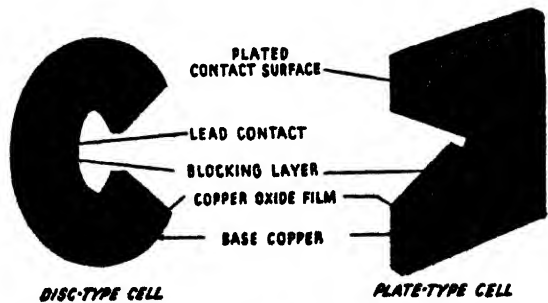


FIG. 7-9. Cross section of a copper-oxide rectifier disk and plate-type cells. (*Courtesy of General Electric Co.*)

A number of commercially important rectifier disks exist, these being known loosely as *copper-oxide* cells,[8] *selenium* cells,[9] and *copper sulfide* cells.[10] Each of these will receive some consideration below.

The copper-oxide cell is illustrated in two forms in Fig. 7-9. These cells are produced by heating a copper disk or plate in a furnace to approximately 1000°F and then quenching it in water. This treatment produces a thin layer of red cuprous oxide with an outer layer of black cupric oxide.
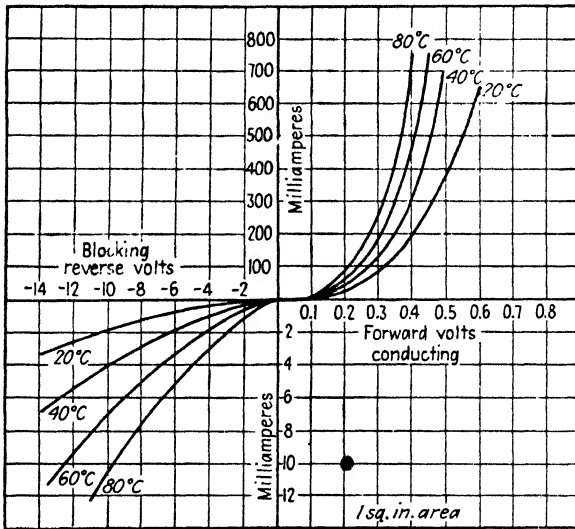


FIG. 7-10. Current-voltage characteristics of a copper-oxide rectifier cell. Note that the scales in the forward and backward directions are different. (*Courtesy of General Electric Co.*)

The cupric oxide is then removed, leaving the cell with a layer of cuprous oxide on the base copper. Contact with the oxide surface can be made in either of two ways. One method is to use a lead disk which is held against the oxide surface at a definite pressure. The other way is to plate a metallic conductor, such as nickel, on the oxide surface. This plated film is then used as the contact surface.

The unilateral characteristic of this cell is shown in Fig. 7-10. It should be noted that the forward resistance is quite low and the resistance in the backward direction is quite high. However, since the current in the reverse direction is not negligible for a high inverse voltage, it is desirable that this inverse voltage be kept as small as possible. Because of this, such cells are limited to low-voltage service, the inverse voltage per cell being restricted to about 10 volts. Higher voltage service is possible by connecting a number of such cells in series. A typical rectifier stack is shown in Fig. 7-11.

In order to prevent an excess inverse current, the maximum temperature of operation should be limited to about 35°C. The output of the rectifier may be increased by providing cooling fins for more effective heat transfer. It may be increased still more by providing forced air cooling.

It has been found that the resistance of a copper-oxide cell increases with time. This aging effect is accompanied by a reduced output voltage for a given input voltage, and also by a reduced efficiency. The cell resistance appears to reach a stable value at about twice the initial value in from 6 months to a year, depending on the amount of operation and the
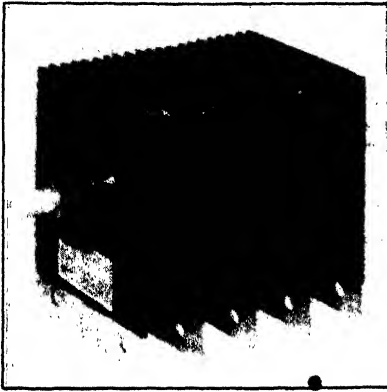


FIG. 7-11. Photograph of a typical metallic rectifier stack (a copper sulfide stack is shown). (*Courtesy of P. R. Mallory and Co.*)
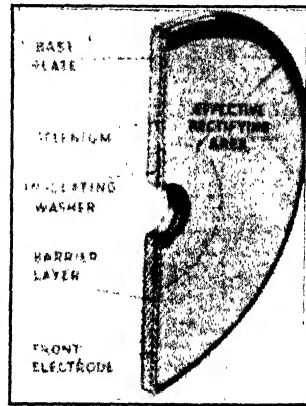


FIG. 7-12. Cross section of a selenium cell. (*Courtesy of Federal Telephone and Radio Corp.*)

temperature. For this reason, the rectifier is rated on the basis of the output which will be obtained after the unit is completely aged.

The selenium cell is physically not unlike the copper-oxide cell. It consists of a base plate of iron or aluminum which is covered by a thin film of selenium. A thin blocking layer forms between the base plate and the selenium upon the application of an a-c potential. Figure 7-12 shows the cross section of a selenium cell. A typical curve of the static characteristics of such a cell is given in Fig. 7-13.

Although the selenium cell possesses the same general properties as the copper-oxide cell, it possesses several distinct advantages. It will withstand a higher inverse voltage, which is approximately 25 volts for this cell. Also, owing to the different materials of construction, the weight of the selenium rectifier for a given rating will be considerably less than for the copper-oxide unit. One disadvantage of the selenium cell is that it shows a loss of some of its rectifying properties when not in use. However,

the original characteristics are restored in a few seconds with the application of an a-c voltage.

The copper sulfide cell consists of a base plate of magnesium, with a copper sulfide semiconductor layer. The active blocking layer, which is
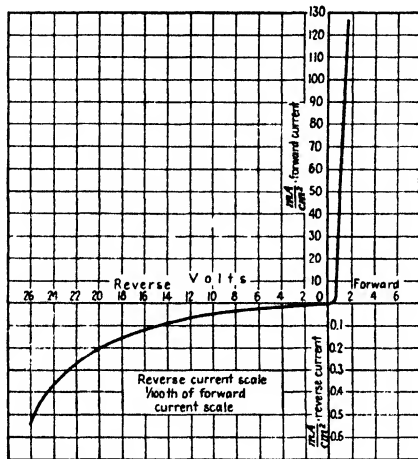


Fig. 7-13. Typical curve of static forward and reverse current characteristics of 26-volt selenium rectifier cell. (*Courtesy of Federal Telephone and Radio Corp.*)

formed between the magnesium and the copper sulfide by an electrochemical process, consists of a film of magnesium sulfide and an adjacent film of cuprous sulfide.

A comparison of the important characteristics of these metallic rectifier cells is contained in Table 7-1.

TABLE 7-1

COMPARISON OF METALLIC RECTIFIERS

| Type of cell | Peak inverse volts per cell, normal operation | Current density, amp/in.$^2$ (full-wave rectifier) | | Maximum operating temperature, °C, normal operation |
| --- | --- | --- | --- | --- |
| | | Normal cooling | Forced-draft cooling | |
| Copper oxide....... | 8.75–11.5 | 0.3 | 0.5 | 35 |
| Selenium........... | 25 | 0.3 | 0.75 | 35 |
| Copper sulfide...... | 5 | 25 | 50 | 40 |

## PROBLEMS

**7-1.** Prove that the following relationships are valid for a plane-parallel diode operating under space-charge-limited conditions:

$$V = E_b \left(\frac{x}{d}\right)^{\frac{4}{3}}$$

$$v = v_b \left(\frac{x}{d}\right)^{\frac{1}{3}}$$

and

$$\rho = \rho_b \left(\frac{x}{d}\right)^{-\frac{2}{3}}$$

where $v_b = (2eE_b/m)^{\frac{1}{2}}$ is the speed with which the electrons strike the plate and $\rho_b = E_b/81\pi d^2 10^9$ is the charge density at the plate. The other symbols have the meanings assigned in the text.

**7-2.** Show that the transit time of an electron from the cathode to the anode of a plane-parallel space-charge-limited diode is $T = 3d/v_b$, where $d$ is the cathode-anode spacing and $v_b = (2eE_b/m)^{\frac{1}{2}}$ is the speed with which the electron strikes the plate. The electron is assumed to leave the cathode with zero initial velocity.

Show that, if the space charge is negligibly small, the transit time is $T = 2d/v_b$. This is only two-thirds of the time taken under space-charge conditions.

**7-3.** Show that the tangent to the potential distribution curve at the anode of a plane-parallel space-charge-limited diode passes through the zero of potential at one-fourth the cathode-plate distance.

**7-4.** A diode having plane-parallel electrodes is operating under space-charge conditions. The plate current is 10 ma at 100 volts plate voltage.

  *a.* What must be the plate voltage in order that the plate current be doubled?

  *b.* What current will be obtained if the voltage is doubled ($E_b = 200$ volts)?

  *c.* If another diode is constructed having half the cathode-anode spacing and twice the electrode area, what current will be obtained if a potential of 100 volts is applied?

**7-5.** A plane-parallel diode having a 2-cm spacing is operated under space-charge conditions. The plate voltage is 100 volts. What is the space-charge density at a point halfway between the cathode and anode?

**7-6.** *a.* A plane-parallel diode with a cathode-anode spacing of 1 cm operates under space-charge-limited conditions at a plate voltage of 100 volts. How much power per square meter must the plate dissipate?

  *b.* If the voltage is increased to 400 volts, by what factor is the dissipation multiplied?

  *c.* If the emissivity of the anode material is 0.5, find the anode temperature in each case. Assume an ambient temperature of 20°C.

**7-7.** In a certain space-charge-limited diode a current of 5 ma results from the application of 100 volts. What is the maximum plate voltage that can be applied before the plate dissipation exceeds 16 watts?

**7-8.** The anode of a space-charge-limited diode is made of material whose emissivity is 0.7. When the plate voltage is 200 volts, the temperature of the anode is 50°C. The ambient temperature is 20°C. What is the plate current density?

**7-9.** Given a tungsten cathode and a cylindrical coaxial anode having the following dimensions:

Lighted length        = 1.0 in.
Cathode diameter = 0.005 in.
Anode diameter    = 0.5 in.

The filament voltage is 2.5 volts. The filament current is 1.0 amp. The plate voltage is 200 volts.

  *a.* Calculate the space-charge-limited current.

b. Calculate the temperature-limited current.

c. Calculate the plate current.

**7-10.** For a space-charge-limited diode the functional relationship between plate current and plate voltage is $I_b = KE_b^n$, where ideally $n = 1.5$. Determine whether or not the 5U4G satisfies this formula, and determine $n$. Use the data from Fig. A9-1.

**7-11.** Given a cylindrical tungsten filament whose lighted length is 1 in. and whose diameter is 4 mils. Coaxial with this emitter is a cylindrical plate whose diameter is 0.4 in.

a. What must be the power input to the filament if it is to operate at 2000°K?

b. What is the saturation current obtainable from this tube?

c. What plate current is obtained at a plate voltage of 100 volts?

**7-12.** a. Calculate the maximum current from a straight tungsten wire 1 mm in diameter and 3 cm long to a cylindrical anode 4 cm in diameter when the potential difference between cathode and anode is 120 volts.

b. What is the lowest temperature at which this emission will take place?

**7-13.** The tungsten filament of a high-vacuum rectifying diode has a diameter of 0.0085 in. and a lighted length of 1.00 in. The plate diameter is 0.75 in.

a. If the filament input is 5.0 amp at 2.0 volts, what is the saturation current?

b. At what potential will the current become temperature-limited?

**7-14.** Prove that the current density at the plate of a cylindrical diode is the same as that in a plane diode if the cathode-anode spacing is the same in both tubes and if the anode radius is very much greater than the cathode radius.

**7-15.** Prove that the space-charge density $\rho$ varies with distance $r$ according to the factor $r^{-\frac{4}{3}} (\beta^2)^{-\frac{1}{3}}$, where $\beta^2$ is defined in Sec. 7-3.

**7-16.** In either a plane or a cylindrical diode all the dimensions are enlarged by the same factor. If the voltage remains constant, show that the space-charge current is unchanged.

**7-17.** Consider a space-charge diode having an arbitrary geometry. Show that if the potential at any point in space changes by a factor $K$ the magnitude of the current density at that point changes by the three-halves power of this factor.[3]

HINT: Use Poisson's equation in three dimensions [Eq. (A6-3)].

### REFERENCES

1. LANGMUIR, I., *Phys. Rev.*, **2**, 450, 1913.
   LANGMUIR, I., and K. B. BLODGETT, *ibid.*, **24**, 49, 1924.
2. CHILD, C. D., *ibid.*, **32**, 492, 1911.
3. LANGMUIR, I., and K. T. COMPTON, *Revs. Modern Phys.*, **3**, 191, 1931.
4. FRY, T. C., *Phys. Rev.*, **17**, 441, 1921; **22**, 445, 1923.
   LANGMUIR, I., *ibid.*, **21**, 419, 1923.
   WHEATCROFT, E. L. E., *IEEJ*, **86**, 473, 1940.
   PAGE, L., and N. I. ADAMS, JR., *Phys. Rev.*, **76**, 381, 1949.
5. DISHINGTON, R. H., *Elec. Eng.*, **67**, 1043, 1948.
6. OSTLUND, E. M., *Electronics*, **13**, 36, June, 1940.
7. HEMISCH, H. K., "Metal Rectifiers," Oxford University Press, New York, 1949.
8. HAMANN, C. E., and E. A. HARTY, *Gen. Elec. Rev.*, **36**, 342, 1933.
9. RICHARDS, E. A., *IEEJ*, **88**, 238, 1941.
   CLARKE, C. A., *Elec. Commun.*, **20**, 47, 1941.
10. RUBEN, S., *Trans. Electrochem. Soc.*, **87**, 275, 1945.

### General References

For general references, see end of Chap. 6.

# CHAPTER 8

# THE KINETIC THEORY OF GASES

IN THE case of the structure of metals, it was mentioned that the ions that constitute the metal are arranged in a regular pattern, this pattern being maintained against the action of externally applied forces by means of the very strong interionic forces. In the case of gases, no such intermolecular forces exist, so that the molecules are free to move more or less independently of each other, except for the effect of the collisions that may occur between them. The kinetic theory of gases is concerned with the study of these colliding molecules. This chapter will emphasize those features which have a bearing on the understanding of the operation of electronic-discharge devices.

**8-1. Classical Concept of a Gas.** According to the classical kinetic theory of gases, any type of molecule is considered as a simple spherical particle that possesses a size and a mass that are characteristic of the gas under survey. This is at variance with the known fact that certain gases are monatomic, *i.e.*, one for which each atom exists as a separate entity, whereas other gases are diatomic or polyatomic. Despite the limitations imposed by this basic assumption, the kinetic theory of gases yields considerable significant information about gas behavior.

This theory supposes that the molecules are in continual motion, the direction of flight of any particle constantly undergoing changes owing to the collisions with other molecules or with the walls of the container. A calculation of the number of such collisions per second, the average distance between collisions, the average speed of the molecules, and similar properties is given in a later section.

The kinetic theory of gases is concerned with the explanation of macroscopic gaseous phenomena in terms of the dynamic behavior of the particles constituting the gas. In this way the features of pressure, viscosity, thermal conductivity, diffusion, and others, all measurable quantities, are reduced to problems in mechanics, specifically, to the kinetics of elastically colliding spheres.

In Chaps. 4 and 5 the distribution in energy of the electrons in a metal is discussed. It is found from kinetic theory that the molecules of a gas, likewise, are distributed in energy but in accordance with the *Maxwell-Boltz-*

*mann* (MB) *law* instead of the Fermi-Dirac-Sommerfeld (FDS) law. The original derivation of the MB distribution function was based upon the mechanical model of the gaseous system, as described above.

The kinetic theory of gases supposes that in an ensemble of colliding particles the velocity of any one particle changes at each collision. Nevertheless, since there are such an enormous number of particles present, an equilibrium stage will be reached within any given velocity range. Thus the fundamental problem is to obtain a distribution function that has as its basis the fact that the number of particles entering a given velocity range per second because of collisions is equal to the number of particles leaving this same range in the same time. In addition, the solution is determined solely with the aid of the elementary laws of mechanics, *viz.*, the laws of conservation of energy and momentum. Temperature enters into the distribution functions because the propounders of this theory made the fundamental assumption that *the average kinetic energy of the molecules is proportional to the temperature.* This one decisive step allows agreement to be had between many theoretically derived equations and experimentally obtained results.

Although the basic laws on which the solution of this problem is based are simple, nevertheless the mathematical formulation of the solution is complicated.[1] The most important concepts of the theory and some of the end results are summarized in the next few sections. Further details, including some derivations, are given in the last part of this chapter.

**8-2. Perfect-gas Laws.** What is the kinetic-gas-theory explanation for the pressure exerted by a gas upon the walls of the enclosing container? The answer is given in terms of the physical picture of the gas discussed above. By definition, the term *pressure* is the *force per unit area.* In a gaseous system, the force is that resulting from the bombardment of the walls of the container by the molecules. Since Newton's second law of motion dictates that force is the rate of change of momentum, the pressure of a gas is the change in momentum of all the particles that strike unit area of the container in unit time.

If the pressure is evaluated as outlined above (see Sec. 8-13), it is found that

$$p = \tfrac{2}{3} N e \bar{E} \tag{8-1}$$

where $p$ is the pressure in newtons per square meter, $N$ is the concentration in molecules per cubic meter, $e$ is the electronic charge in coulombs, and $\bar{E}$ is the average total energy per molecule in electron volts. The fundamental assumption of kinetic theory is that the average energy is proportional to the absolute temperature, or

$$e\bar{E} = \tfrac{3}{2} k T \tag{8-2}$$

where $T$ is in degrees Kelvin and the proportionality factor $k$, called the *Boltzmann constant*, is in joules per degree Kelvin. Combining these two equations gives

$$p = NkT \qquad (8\text{-}3)$$

An extremely important conclusion is contained in this expression: *At a given temperature and pressure all gases must contain the same number of molecules per cubic meter.* This is called *Avogadro's principle.*

Consider two containers occupying the same volume, one filled with gas $A$ and the other with gas $B$. If these are at the same temperature and pressure, then, by the above principle, the two containers enclose equal number of molecules. Hence, the weight of the gas in the first container is to that in the second as the weight of gas molecule $A$ is to the weight of molecule $B$. This leads to the concept of *molecular weight*.

The molecular weight of any substance is the ratio of the mass of one molecule of that substance to that of another which has been chosen as a standard. The molecular weight of diatomic oxygen is arbitrarily taken as 32.0000 (actually, the atomic weight of monatomic oxygen is taken as 16.0000). Hence the molecular weight of any substance is simply a numeric. Also, the term *gram-molecular weight* denotes a quantity of substance equal to this numeric, in grams. It thus follows that a gram-molecular weight, or *mole*, of any substance must contain the same number of molecules as the gram-molecular weight of any other substance. This number, known as *Avogadro's number*, is $6.02 \times 10^{23}$ molecules per mole. Furthermore, from the foregoing, a mole of any gas must occupy the same volume. This volume (the gram-molecular volume) has been found by experiment to be 0.0224 $\text{m}^3$, or 22.4 liters, under standard conditions of 0°C and 760 mm Hg pressure. Hence, there are $6.02 \times 10^{23}/0.0224 = 2.69 \times 10^{25}$ molecules per cubic meter of any gas at these standard conditions. This quantity is known as *Loschmidt's number*.

This is a large concentration, and yet it is only about one one-thousandth the density of electrons in a metal, which in Chap. 4 is found to be approximately $10^{28}$ electrons per cubic meter. Furthermore, the volume occupied by these molecules is (under standard conditions) only about one one-hundredth of 1 per cent of volume of the container. This follows from the fact that the radius of a molecule is approximately 1A ($10^{-10}$ m) (see Table 8-1), and hence the volume occupied by all the molecules in 1 $\text{m}^3$ is $(\frac{4}{3}\pi)(10^{-10})^3 \times 2.69 \times 10^{25} = 10^{-4}$ $\text{m}^3$. Most of the gas container is empty!

According to Eq. (8-3), the concentration varies directly as the pressure and inversely as the temperature. Thus, at any temperature $T$°K and pressure $p$ mm Hg the concentration is

$$N = 2.69 \times 10^{25} \left(\frac{p}{760}\right)\left(\frac{273}{T}\right)$$

$$= 9.68 \times 10^{24} \frac{p}{T} \qquad \text{molecules/m}^3 \qquad (8\text{-}4)$$

Equation (8-3) is one form of the equation of state of an ideal gas. The form that is more familiar in elementary physics and chemistry may be obtained as follows. Each member of Eq. (8-3) is multiplied by $V$, the volume of the gas in cubic meters. Then

$$pV = NVkT$$

Since $N$ is the number of molecules per cubic meter, then $NV$ gives the total number of molecules in the container of volume $V$. However, Avogadro's number $N_0$ gives the number of molecules in 1 mole of gas. Hence if the volume $V$ contains $n$ moles of gas, then $NV = nN_0$, and

$$pV = nN_0kT$$

Finally, the introduction of the familiar *gas constant per mole*, $R$, which is defined by the relation

$$R \equiv N_0k \qquad (8\text{-}5)$$

leads to the well-known equation of state, *viz.*,

$$pV = nRT \qquad (8\text{-}6)$$

where $p$ is in newtons per square meter, $V$ is in cubic meters, $n$ is a numeric, $R$ has the dimensions joules per degree Kelvin per mole, and $T$ is in degrees Kelvin. Obviously, if $p$ and $V$ are expressed in any other units, all that is necessary is to convert $R$ to this same system of units and Eq. (8-6) will remain valid. A still more elementary form of Eq. (8-6) is the equation

$$\frac{pV}{T} = \text{const} \qquad (8\text{-}7)$$

The foregoing derivation throws an interesting light on the significance of the Boltzmann constant $k$. If $R$ is the gas constant per mole, then, according to Eq. (8-5), $k$ is the *gas constant per molecule*. In fact, the numerical value of $k$, namely, $1.380 \times 10^{-23}$ joule/°K, is determined from Eq. (8-5). This is done indirectly in the following way: Starting with 1 mole of a gas, the temperature, volume, and pressure are measured experimentally. Then with the aid of Eq. (8-6) the magnitude of $R$ is calculated. In order to determine $N_0$, an electrolytic process is employed. The number of coulombs $F_0$ necessary to deposit out of solution 1 gram-atomic

weight of a monovalent substance is measured. Then, since the charge per atom is $e$, the total measured charge will be $F_0 = N_0 e$. From other experiments, already referred to, the electronic charge is a known quantity, whence $N_0$ may be calculated. The quantity $F_0$ is known as *Faraday's constant*. The numerical values of $F_0$ and $R$ are given in Appendix I.

**8-3. Distribution Functions of Gas Molecules.** If the energy is expressed in units of $E_T$ electron volts, the MB distribution function has a universal form valid for any gas at any temperature. $E_T$ is defined by Eq. (5-6), namely,

$$eE_T \equiv kT \tag{8-8}$$

If $\eta$ is defined by $\eta \equiv E/E_T$, the energy distribution is shown in Sec. 8-10 to be given by

$$\rho_\eta = \frac{2N}{\sqrt{\pi}} \eta^{\frac{1}{2}} \epsilon^{-\eta} \tag{8-9}$$

This is plotted in Fig. 8-1. $\rho_\eta$ represents the energy density, and so the number of molecules having (dimensionless) energies between $\eta$ and $\eta + d\eta$
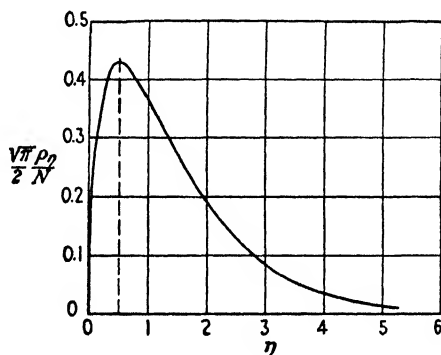
Fig. 8-1. The Maxwell-Boltzmann energy distribution function, plotted in terms of the variable $\eta = E/E_T$.

is $\rho_\eta \, d\eta$. It is seen that the molecules possess all energies from zero to infinity. Since the curve approaches the axis exponentially, there are very few molecules having very large energies. The energy at which the maximum of the distribution occurs is the *most probable energy*, since more particles have energies in this neighborhood than any other. This is given by $\eta = \frac{1}{2} = E/E_T$, or at an energy of $E_T/2$. At room temperature (say, 300°K) this most probable energy is only 0.013 ev. At absolute zero, $\rho_\eta = 0$, which indicates that all of the gas molecules have zero energy.

This distribution should be compared with Fig. 4-9 for electrons in a metal. Even at absolute zero, the electrons are much more energetic than

the gas molecules at room temperature. The most probable electronic energy is $E_M$, which is of the order of a few electron volts, or about 500 times the most probable molecular energy at room temperature.

The average energy of the gas molecules may be calculated from the distribution (8-9) (see Sec. 8-10) and is found to be $\frac{3}{2}E_T$ electron volts, or $\frac{3}{2}kT$ joules.

The molecules are distributed in speed (see Fig. 8-5), and the peak of this curve corresponding to the *most probable speed* is found to be

$$v_c \equiv \sqrt{\frac{2kT}{m}} \qquad (8\text{-}10)$$

The numerical value of this speed for the case of monatomic hydrogen at room temperature is easily calculated to be $2.2 \times 10^3$ m/sec, or approximately 5,000 mph. Even though the energy of a molecule is small, its speed is terrific (compared with that of macroscopic objects) because its mass is so small.

The average speed is given by (see Sec. 8-9)

$$\bar{v} = \frac{2}{\sqrt{\pi}} v_c = 1.128 v_c \qquad (8\text{-}11)$$

The random current density $J_r$ is the charge per second crossing 1 m$^2$ in the gas. A calculation of this (see Sec. 8-11) leads to the expression

$$J_r = \frac{1}{4} Ne\bar{v} = Ne \sqrt{\frac{kT}{2\pi m}} \qquad (8\text{-}12)$$

where $J_r$ is expressed in amperes per square meter, $N$ is the number of ions per cubic meter, $e$ is the charge in coulombs on each ion, $\bar{v}$ is the average speed in meters per second, and $m$ is the mass in kilograms.

**8-4. Mean Free Paths of Electrons and Gas Molecules.** If molecules travel at the tremendous speeds indicated in the preceding section, then it might be expected that gases should have extremely fast rates of diffusion and that all the gas contained in a given volume should escape in a small fraction of a second after removing the cover of the container. Actually this is not so, not only because the gas molecules strike the walls of the container, but because they also strike each other an enormous number of times each second. In fact, it is the collisions with the walls of the gas container that constitute the gas pressure, as was shown above.

The average distance that a molecule travels between successive collisions with other molecules is called the *molecular mean free path*. In electronic devices in which both molecules and electrons exist, the *electronic mean free path* is defined as the average distance that an electron travels

between collisions with gas molecules. The electronic and molecular mean free paths will be denoted, respectively, by the symbols $l_e$ and $l_m$.

The problem involved in the calculation of the mean free path of a molecule among similar molecules is essentially the following: Given a container of gas at a known pressure, the molecules of which obey the Maxwellian distribution of velocities. Each molecule will collide with another molecule, be deflected, move a short distance, collide with another molecule, again be deflected, and so on. What is the average distance traveled by a gas molecule between successive collisions? The solution of this problem may be shown to be [2]

$$l_m = \frac{1}{4\sqrt{2}\sigma N} \qquad \text{meters} \qquad (8\text{-}13)$$

where $\sigma = \pi r^2$ is the molecular cross section ($r$ is the molecular radius in meters) and $N$ is the concentration per cubic meter. This expression shows that the mean free path varies inversely with the number of molecules present per cubic meter. This is entirely reasonable, since it shows that the mean free path is longer when fewer molecules are present.

Numerical values for the radii of molecules have been obtained indirectly through the use of this expression for mean free path. Specifically, it can be shown by the kinetic-theory methods of this chapter that the viscosity of a gas depends upon the mean free path. From the experimental measurements of viscosity, values of $l_m$ are calculated. The radii $r$ calculated from these values of $l_m$ are found to vary only slightly from element to element. Typical values of molecular radii obtained from such measurements are tabulated in Table 8-1. They are all seen to be between 1 and 2 A.

TABLE 8-1 *

MOLECULAR RADII
(Determined from viscosity measurements)

| Molecule | Radius, A |
|----------|-----------|
| Ne | 1.17 |
| A | 1.43 |
| Kr | 1.59 |
| Xe | 1.75 |
| Hg | 1.82 |
| $H_2$ | 1.09 |
| $N_2$ | 1.58 |
| $O_2$ | 1.48 |

* From L. B. Loeb, "Kinetic Theory of Gases," p. 643, McGraw-Hill Book Company, Inc., New York, 1934.

*Example.* Calculate the number of collisions per second made by a nitrogen molecule at room temperature and atmospheric pressure.

*Solution.* If the temperature is taken as 300°K, then the concentration is, according to Eq. (8-4),

$$N = (9.68 \times 10^{24})(\tfrac{760}{300}) = 2.45 \times 10^{25} \text{ molecules/m}^3$$

Upon using the molecular radius of nitrogen listed in Table 8-1, then

$$l_m = \frac{1}{(4\sqrt{2})(\pi)(1.58)^2(10^{-20})(2.45)(10^{25})} = 9.18 \times 10^{-8} \text{ m}$$

Since $l_m$ represents the distance (in meters) traveled per collision, then the reciprocal of $l_m$ gives the number of collisions per meter. Hence

$$\frac{1}{l_m} = \frac{1}{9.18 \times 10^{-8}} = 1.09 \times 10^7 \text{ collisions/m}$$

The average molecular speed is, according to Eqs. (8-11) and (8-10),

$$\bullet \qquad \bar{v} = 1.128 v_c = 1.128 \left(\frac{2kT}{m}\right)^{\frac{1}{2}}$$

$$= 1.128 \left(\frac{2 \times 1.38 \times 10^{-23} \times 300}{28 \times 1.66 \times 10^{-27}}\right)^{\frac{1}{2}} = 4.75 \times 10^2 \text{ m/sec}$$

Hence there are

$$\left(1.09 \times 10^7 \frac{\text{collisions}}{\text{m}}\right)\left(4.75 \times 10^2 \frac{\text{m}}{\text{sec}}\right) = 5.18 \times 10^9 \text{ collisions/sec}$$

This example is not intended as a rigorous quantitative calculation. It does give some idea, however, of the incessant activity that is taking place in a gas. Thus, even though the molecules travel with extremely high average speeds, they travel only an extremely short distance before they make a collision with another molecule and so suffer a change of direction. This readily accounts for the slow diffusion rates of gases.

It will be shown in Sec. 8-6 that the electronic mean free path is given by the following expression:

$$l_e = \frac{1}{\sigma N} \tag{8-14}$$

Hence, by comparing this expression with Eq. (8-13), it is seen that

$$l_e = 4\sqrt{2}l_m = 5.66 l_m$$

Thus the electronic mean free path among nitrogen molecules at the physical conditions specified in the foregoing illustrative example is $l_e = 5.66(9.18 \times 10^{-8}) = 5.20 \times 10^{-7}$ m.

Since the mean free path varies inversely as the concentration, then if the pressure is reduced sufficiently the mean free path can be made very large. For example, in a "vacuum" tube in which the pressure is about $10^{-6}$ mm Hg, the mean free path will be $5.20 \times 10^{-7} \times 760 \times 10^6 = 3.95 \times 10^2$ m. This distance is so very much larger than the distance between the electrodes in a tube that very few of the electrons leaving the

cathode will collide with gas molecules in the interelectrode space. Because of the lack of collisions, little ionization by collision can take place, and the tube will act as a vacuum tube and not as a gaseous-discharge device.

**8-5. The Distribution of Free Paths.** It is not implied by nor is it to be inferred from the foregoing that each molecule or electron will travel a distance exactly equal to the *mean* free path, but simply that the average distance traveled between collisions will be equal to $l$. What fraction of the particles will travel some other distance? The answer to this query
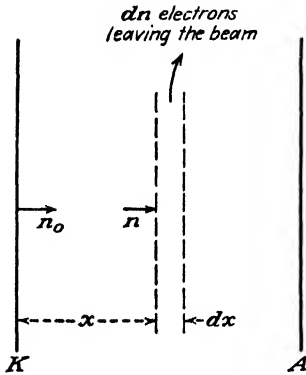


FIG. 8-2. Electrons leaving the cathode $K$ are scattered out of the beam by the gas molecules in the tube.

will also furnish an answer to the following practical question: What fraction of the electrons leaving the cathode in a given electronic device in which a gas at low pressure is contained will make collisions with the gas molecules before they reach the collector?

Suppose that the electrons are moving along the $X$ direction in a gas. Let $n_0$ denote the number of electrons per second leaving the emitter at $x = 0$, and let $n$ be the corresponding number that have arrived at any distance $x$. Refer to Fig. 8-2. There are $n$ electrons that have *not* yet made a collision, since it is assumed that each collision removes the particles from the beam. Since $l$ is the distance between collisions, then $1/l$ is the number of collisions per meter per electron. The number of collisions made by the $n$ electrons in traversing a distance $dx$ is $(n/l)\ dx$. This also gives the number of electrons removed from the electron beam in this distance. Thus

$$dn = -\frac{n}{l} dx$$

(the negative sign appears in this equation because $dn$ represents a negative number; that is, $n$ is decreasing).

Upon integration, this expression leads to

$$n = n_0 \epsilon^{-x/l} \tag{8-15}$$

Combining this value of $n$ with the foregoing equation yields

$$dn = -\frac{n_0}{l} \epsilon^{-x/l} dx \tag{8-16}$$

Since $dn$ expresses the number of collisions that occur in the interval between $x$ and $x + dx$, then Eq. (8-16) gives the number of paths that

terminate in this interval. It is for this reason that this equation is often referred to as the *distribution function in free paths*.

Suppose that the method of averages of Sec. 5-10 is employed, using the present distribution function. Thus, the mean distance is

$$\bar{x} = \frac{\displaystyle\int_0^\infty x\, dn}{\displaystyle\int_0^\infty dn} = \frac{\displaystyle\int_0^\infty -\frac{xn_0}{l}\,\epsilon^{-x/l}\,dx}{n_0} = l$$

This shows that the average distance between collisions is the mean free path, as it must be.

Equation (8-15) is sometimes given the designation of the "survival equation," since it gives the number of particles which have *not* yet suffered a collision in traveling a distance $x$. The fraction $n/n_0$ is plotted in
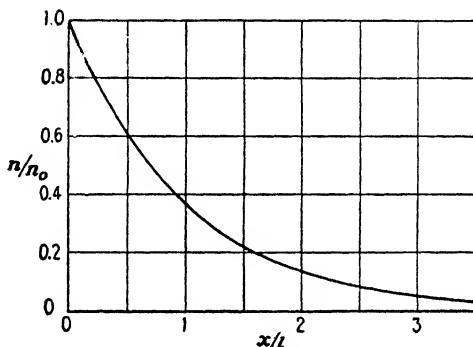


Fig. 8-3. Graph of the "survival equation" giving the fraction of particles that travel a distance $x$ in a gas *without* making a collision. The distance is measured in terms of the mean free path $l$.

Fig. 8-3. It is seen from this curve that the longer paths are much less probable than the shorter ones. When $x = l$, then $n/n_0 = \epsilon^{-1} = 0.368$. Hence, only 36.8 per cent of the particles travel a distance equal to one mean free path *without* colliding with another particle. A little thought will convince the reader that this result is not incompatible with the definition of mean free path. Thus, from Fig. 8-3, though most particles have paths that are shorter than $l$, nevertheless there are enough particles having longer paths to yield a mean value equal to $l$.

If it is desired to have very few collisions in a tube, the pressure must be reduced to a value for which the mean free path will be very large compared with the interelectrode spacing $x$. As a first guess, one might consider $l$ equal to 100 times $x$ as a reasonable value. Actually, however, such

a corresponding pressure would be much too high for a "vacuum" tube. Thus, if $x/l$ is small, then $\epsilon^{-x/l}$ can be approximated by the first two terms of its series expansion, *viz.*,

$$\epsilon^{-x/l} = 1 - \frac{x}{l} \tag{8-17}$$

an expression that is accurate to better than 0.5 per cent if $l > 10x$. Then Eq. (8-15) becomes

$$n = n_0\left(1 - \frac{x}{l}\right)$$

or

$$\frac{n_0 - n}{n_0} = \frac{x}{l} \tag{8-18}$$

Since $n_0$ equals the total number of particles and $n$ gives those particles which have not made collisions, then $(n_0 - n)/n_0$ gives the fraction of all the electrons that do collide with molecules in the interelectrode space.

If, as chosen, $x/l = \frac{1}{100}$, then 1 per cent of the electrons leaving the cathode will collide with the residual air molecules. If the applied potential were high enough so that ionization resulted, this 1 per cent of collisions would cause the formation of many more than enough positive ions to neutralize completely the space charge, as discussed later. Consequently, the pressure in a "high-vacuum" tube must be reduced to approximately $10^{-6}$ mm Hg. Under these conditions, $l \doteq 100$ m so that, for an interelectrode spacing of about 1 cm, Eq. (8-18) states that only one out of $10^4$ electrons leaving the cathode will collide with a molecule. This will not provide enough positive ions to destroy the space-charge characteristics of the tube.

**8-6. Collision Cross Sections.** The expression for the mean free path of an electron among molecules given in Eq. (8-14) will now be derived. In order to derive this quantity, it is noted that the diameter of a gas molecule is of the order of $10^{-10}$ m, whereas that of an electron is approximately $10^{-15}$ m. Relatively, therefore, the electron may be considered to be a point. Further, since the thermal motion of the electron is so much greater than that of the molecules, it may be assumed, as a first approximation, that the molecules are stationary and that the electron collides with these rigid spheres.

Consider an electron to be traveling in the gas in the direction $DD'$ of Fig. 8-4. An imaginary parallelepiped is erected, with the surface of area $A$ normal to this direction of motion, and of length $d$. The circles shown on this figure denote the molecules that are contained in this volume element. What is the probability that the electron will collide with a molecule in this parallelepiped?

The following analogous problem in ballistics will clarify the analysis: Suppose that a gun of small caliber were fired at random at a wall of area $A$ in which a number of holes had been drilled. What is the probability that the bullet will pass through one of these holes? Evidently, if the total area of the holes is, say, $\frac{1}{4}A$, then there is 1 chance in 4 that the bullet will not strike the wall material. That is, the probability is given as the ratio of the total area of the holes to the total area of the wall.

Similarly in the electronic problem, the probability that an electron will collide with a molecule within the distance $d$ is the ratio of the total projected area of all the molecules in the parallelepiped (as seen by the colliding electron) to the area $A$. If the radius of a molecule is $r$, then the projected area per molecule is $\sigma = \pi r^2$. The total projected area of all the molecules in the parallelepiped is $\sigma N A d$, where $N$ is the concentration (molecules per cubic meter). Consequently, the probability that an electron will collide with a molecule in a distance of 1 m of path ($d = 1$) is $\sigma N$. This quantity



Fig. 8-4. To determine the mean free path of electrons among gas molecules.

is by definition the *probability of collision* $P_c$.   Therefore

$$P_c \equiv \sigma N \qquad (8\text{-}19)$$

Suppose, for example, that $P_c$ equals $\frac{1}{4}$. Then there is 1 chance in 4 that an electron will collide with a gas molecule in a distance of 1 m. Therefore, on the average, an electron will travel a distance of 4 m before making a collision, so that $l = 4$ m. In other words, the mean free path is the reciprocal of the probability of collision.   Thus

$$l = \frac{1}{P_c} = \frac{1}{\sigma N} \qquad (8\text{-}20)$$

It has been implicitly assumed in this derivation that the projected areas of no molecules overlap each other, so that the total projected area in the direction of the oncoming electron is the sum of the projected areas of all the molecules in the parallelepiped. If the gas were so dense that this assumption could no longer be made, then the expression (8-19) would no longer be valid. Actually, however, the representation of a gas molecule as a solid sphere of radius $r$ is known not to be the true physical picture. Furthermore, according to this derivation, $P_c$ should be independent of the speed of the impinging electron. Experiments on the collisions of electrons among gas molecules, to be described in the next chapter, show that this is not true and that $P_c$ depends very markedly upon the energy of the electron.

What then is the meaning to be associated with the cross section $\sigma$? In view of the foregoing discussion, $\sigma$ will be called the *collision cross section of a molecule* and will be *defined* by the relationship $\sigma \equiv P_c/N$. The reasons for this name and for this choice of definition are evident from the manner of the development. Thus, the total collision cross section per cubic meter of the gas is defined as $\sigma N$ and is equal to $P_c$. Since $P_c$ gives the number of collisions made by an electron traveling a distance of 1 m through a gas, then $\sigma N$ is merely a measure of the obstruction offered to this electron by the molecules in 1 $m^3$ of gas. An experimental method of measuring $P_c$ is discussed in Sec. 9-14.

Since the probability of collision will depend, in general, upon the temperature and pressure, it is customary to specify the values of $P_c$ for a pressure of 1 mm Hg and a temperature of 0°C. For other than these specified conditions, corrections must be made to the mean free path. The mean free path at a pressure of $p$ mm Hg and a temperature of $T$°K is given by

$$l = l_0 \frac{1}{p} \frac{T}{273} \tag{8-21}$$

where $l_0$ is the mean free path at 1 mm Hg and 0°C.

**8-7. The Maxwell-Boltzmann Distribution Function.** The kinetic theory method of obtaining this distribution is outlined in Sec. 8-1. This method will not be followed here, but instead it will now be shown that the MB function may be obtained as a limiting case of the FDS distribution. This latter function, given by Eq. (5-18), is

$$\rho_T = \frac{2m^3}{h^3} \frac{1}{1 + \epsilon^{(E-E_M)/E_T}} \tag{8-22}$$

If the unity in the denominator were neglected, this expression would be much simplified. In order to investigate the physical circumstances under which such an approximation is justified, Eq. (8-22) is rewritten in the form

$$\rho_T = \frac{2m^3}{h^3} \frac{1}{1 + (1/a)\epsilon^{E/E_T}} \tag{8-23}$$

where, for convenience, the constant $\epsilon^{-E_M/E_T}$ has been replaced by the factor $1/a$.

The smallest value that the exponential term in the denominator may attain is unity, since $E/E_T$ is never negative. This condition prevails when the energy $E$ of the particle is zero. Hence, if $a$ is very small, then the second term in the denominator becomes large compared with the first under all conditions of energy, it being thus possible for the first term to

be neglected. Under these conditions, Eq. (8-23) may be written in the form

$$\rho_r = \frac{2m^3}{h^3} a \epsilon^{-E/E_T} \qquad (8\text{-}24)$$

In accordance with the principles of Chap. 5, this expression is equivalent to

$$dN_r = \frac{2m^3 a}{h^3} \epsilon^{-E/E_T} dv_x \, dv_y \, dv_z \qquad (8\text{-}25)$$

This equation expresses the number of particles per unit volume having velocity components in the range between $v_x$ and $v_x + dv_x$, $v_y$ and $v_y + dv_y$, and $v_z$ and $v_z + dv_z$. This is the MB distribution function, the Maxwellian distribution function, or the classical or nondegenerate distribution function.

The conditions under which the quantity $a$ is small, which is the criterion for the validity of Eq. (8-25), are best investigated by finding an expression for the constant $a$. This constant is found by integrating Eq. (8-25) over the whole of velocity space and equating the result to the density or concentration $N$ of particles. If use is made of Eq. (8-8), which defines $E_T$, then

$$N = \frac{2am^3}{h^3} \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} \epsilon^{\frac{-m(v_x{}^2 + v_y{}^2 + v_z{}^2)}{2kT}} dv_z \qquad (8\text{-}26)$$

This is equivalent to the product of three integrals, each of the form $\int_{-\infty}^{\infty} \epsilon^{-\frac{1}{2}\frac{mv^2}{kT}} dv$. Upon performing the indicated integrations (see Appendix V), there results,

$$a = \frac{Nh^3}{2(2\pi mkT)^{\frac{3}{2}}} \qquad (8\text{-}27)$$

In order that $a \ll 1$, the criterion for the validity of Eq. (8-25), it is observed that $N$ must be small, $m$ must be large, or $T$ must be large, or any combination of these must be true. For the case of electrons *in metals*, for which $m = 9.107 \times 10^{-31}$ kg and $N$ is of the order of magnitude of $5 \times 10^{28}$ electrons per cubic meter, calculations show that $T$ must exceed $50,000°K$ in order that $a$ may become less than unity. Consequently, electrons in metals *never* obey the MB distribution function.

A gas under standard conditions of temperature and pressure is a system of particles whose density is much smaller and whose mass is much larger than those of electrons in metals. For example, if numerical values are substituted into Eq. (8-27), a value of $a$ equal to $1.6 \times 10^{-5}$ is obtained for monatomic hydrogen. It is evident from an inspection of Eq. (8-27)

that the value of $a$ for any gas heavier than hydrogen will be smaller than the value that has just been calculated. Consequently, Eq. (8-25) is valid for *any* gaseous system under normal conditions of pressure and temperature. If, however, the temperature of the gas is decreased to a very low value and the pressure is increased (so that $N$ increases), then $a$ may become comparable with unity. Under these circumstances the FDS function, and not the MB function, must be used. This was, in fact, the application suggested for the FDS statistics before it was applied to the description of phenomena associated with the electrons in metals. It explains the term "degenerate gas distribution function." Actually, however, no gases can be cooled to a temperature low enough so that the FDS function applies, since the gases liquefy before these extreme conditions are reached.

In view of the foregoing, molecular gases in general satisfy the distribution function obtained by combining Eq. (8-27) with (8-25), *viz.*,

$$dN_\tau = N \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \epsilon^{-E/E_T} d\tau \qquad (8\text{-}28)$$

which forms the basis of the kinetic theory of gases. In this equation $E$ is expressed in electron volts. If it is desired to use $E$ in joules, then it is necessary only to replace $E_T$ electron volts by its equivalent $kT$ joules.

Although this MB or classical distribution function has here been derived as a special case of the degenerate electron-gas distribution function, nevertheless, historically, Eq. (8-28) was known long before the introduction of the FDS function. It is not at all surprising, therefore, that Planck's constant $h$, which was introduced with the birth of the quantum theory, does not appear in this resulting equation.

It must not be inferred, however, that systems of gas molecules are the only physical systems that obey the MB statistics. In fact, it was shown in Sec. 5-16 that the electrons which have escaped from a metal also follow the classical law. The physical reason for this should be apparent. Whereas the electron density inside the metal is extremely high, only those relatively few electrons which have energies higher than the surface barrier can escape. For these, $N$ is small enough to reduce $a$ to the order of magnitude that justifies the conversion of the FDS into the MB function.

**8-8. The Boltzmann Relation.** The distribution function (8-28) can easily be extended to include cases where the particles are subjected to external forces, for example, the case of the gas molecules in the atmosphere which are subject to the earth's gravitational field. Under such circumstances the quantity $E$ in Eq. (8-28) is to be interpreted not as the kinetic energy of the particle but rather as the total energy of the particle, potential plus kinetic energies.

If Eq. (8-28) is integrated over all possible values of velocity, the result is

$$N = N_0 \epsilon^{-U/kT} \tag{8-29}$$

where $N$ is the density of particles in a region of space where the potential energy is $U$ joules and where $N_0$ is a constant arising from the integration. $N_0$ is easily interpreted as the density at that point in the region at which $U = 0$. This expression is known as the *Boltzmann relation*.

In the earth's gravitational field, $U = mgy$ joules, where $y$ is the height in meters above the surface of the earth and $g$ is the acceleration of gravity in meters per second squared. The Boltzmann relation becomes

$$N = N_0 \epsilon^{-mgy/kT}$$

which expresses the variation of the atmospheric density, or pressure, with the height $y$. This expression is frequently referred to as the *barometric equation*.

As a second example of the applicability of the Boltzmann relation, consider two neighboring portions of a gaseous discharge in which the electrons are known to obey the classical distribution function. If the potential in one region of the discharge is $E_1$ and that in another region is $E_2$, then

$$N_1 = N_0 \epsilon^{-E_1/E_T} \qquad N_2 = N_0 \epsilon^{-E_2/E_T}$$

so that the ratio

$$\frac{N_2}{N_1} = \epsilon^{-(E_2 - E_1)/E_T} \tag{8-30}$$

gives a measure of the concentrations of electrons in the two regions of the discharge.

**8-9. The Classical Speed Distribution Functions.** It will be assumed in what follows that the potential energy of the particles is zero. Then the distribution in speed is obtained by transforming Eq. (8-28) from a relationship in $E$ to one in $v$. This necessitates the use of the spherical element of velocity space, $4\pi v^2 \, dv$ instead of the element $d\tau$. The result so obtained is exactly the counterpart of the expression obtained in Sec. 5-4. It is

$$dN_v = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} N \epsilon^{-mv^2/2kT} 4\pi v^2 \, dv \tag{8-31}$$

This distribution function expresses the number of particles whose speeds lie in the range between $v$ and $v + dv$.

Since the power of an exponential must be a dimensionless quantity, then the term $2kT/m$ must have the dimensions of the square of a velocity.

In view of this, the "characteristic" velocity, designated by the symbol $v_c$, is defined by the relation

$$v_c \equiv \sqrt{\frac{2kT}{m}}$$

or, equivalently,

$$\tfrac{1}{2}mv_c{}^2 = kT = eE_T \qquad \text{joules} \tag{8-32}$$

The distribution function (8-31) becomes, by virtue of Eq. (8-32), the simplified expression

$$dN_v = \frac{4N}{\sqrt{\pi}} \frac{v^2}{v_c{}^3} \epsilon^{-v^2/v_c{}^2} dv \tag{8-33}$$

By introducing the dimensionless variable

$$u = \frac{v}{v_c} \qquad \text{or} \qquad v = uv_c \tag{8-34}$$

this equation reduces to the form

$$dN_u = \frac{4N}{\sqrt{\pi}} u^2 \epsilon^{-u^2} du \tag{8-35}$$

It is observed that the general method of designating the dependent variable in the distribution function, introduced in Chap. 5 in connection with the FDS function, has been adhered to. For this reason the subscript of $dN$ has been changed from $v$ to $u$ in Eq. (8-35). Similarly, the corresponding density function has the form $\rho_u = dN_u/du$. The density $\rho_u$ gives the number of molecules per cubic meter whose (dimensionless) speeds are between $u$ and $u + du$. This practice will be followed throughout this chapter.

Since this·expression contains neither $m$ nor $T$, then a single plot of

$$\frac{\sqrt{\pi}}{4} \frac{1}{N} \frac{dN_u}{du} = \frac{\sqrt{\pi}}{4N} \rho_u = u^2 \epsilon^{-u^2}$$

will give the speed distribution function of any nondegenerate gas at any temperature $T$. This distribution is illustrated in Fig. 8-5.

It is observed from this diagram that the function $dN_u/du$ vs. $u$ starts at the origin, rises to a maximum at the point $u = 1$ (or $v = v_c$), and then decreases asymptotically to zero. This shows that the particles may possess all speeds from zero to infinity, although the number of particles that move with large speeds is very small. Of course, the speed at which the maximum of the function occurs is the "most probable" speed, since

more particles have speeds in the neighborhood of this value than any other speed.

In order to verify mathematically that the most probable velocity is equal to the characteristic velocity $v_c$, it is necessary to differentiate Eq.
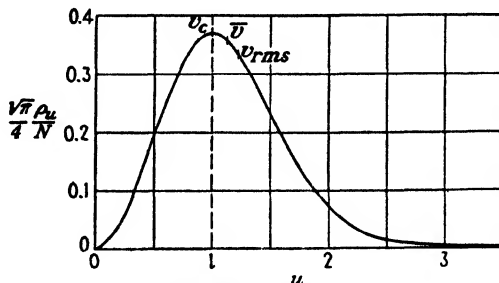


FIG. 8-5. The Maxwell-Boltzmann speed distribution function, plotted in terms of the variable $u = v/v_c$.

(8-35) and then set the derivative equal to zero, the usual method for determining the maximum of a given function. Thus

$$\frac{d\rho_u}{du} = \frac{d}{du}\left(\frac{4N}{\sqrt{\pi}} u^2 \epsilon^{-u^2}\right) = \frac{4N}{\sqrt{\pi}} (2u\epsilon^{-u^2} - 2u^3\epsilon^{-u^2}) = 0$$

from which it follows that $u = 1$, or $v = v_c$, at the position of the maximum.

In order to determine the average value of any function of the speed, use is made of the results of Sec. 5-10, where, of course, the MB distribution function is used instead of the FDS function.

*Example.* Calculate the average speed $\bar{v}$ of a gas molecule.

*Solution.* According to the general expression for the average value of a function,

$$\bar{v} = \frac{\displaystyle\int_0^\infty v\,dN_v}{\displaystyle\int_0^\infty dN_v} = \frac{\displaystyle\int_0^\infty uv_c\,dN_u}{N} = \frac{4v_c}{\sqrt{\pi}} \int_0^\infty u^3\epsilon^{-u^2}\,du$$

which reduces to (see Appendix V),

$$\bar{v} = \frac{2}{\sqrt{\pi}} v_c = 1.128v_c \tag{8-36}$$

Another speed function that is frequently found in the literature is the *root-mean-square* (rms) *velocity.* As the name implies, this is the value which is obtained by taking the square root of the average (or mean) value of the square velocity, exactly analogous to rms quantities in electric-circuit theory. By proceeding as above,

$$v_{rms} = \sqrt{\overline{v^2}} = \sqrt{\frac{3}{2}}\,v_c = 1.224v_c \tag{8-37}$$

The positions of the different speed functions $v_c$, $\bar{v}$, $v_{\text{rms}}$ are indicated on the graph of Fig. 8-5.

It should be carefully noted that the mean of the squared velocity is not the same thing as the square of the mean velocity. That is $\overline{v^2} = 1.5v_c^2$, whereas $(\bar{v})^2 = 4v_c^2/\pi$. Consequently, the utmost care must be exercised when calculating an average value so that the desired mean is actually evaluated.

**8-10. The Energy Distribution Function.** With the aid of the familiar relation $eE = \frac{1}{2}mv^2$, the distribution in speed can be transformed into the distribution in energy. For the nondegenerate gas, this transformation leads to the equation

$$dN_E = \frac{2N}{\sqrt{\pi}} \frac{E^{\frac{1}{2}}}{E_T^{\frac{3}{2}}} \epsilon^{-E/E_T} dE \qquad (8\text{-}38)$$

where, as usual, $E$ is expressed in electron volts. If it is desired to use $E$ in joules, then it is necessary only to replace $E_T$ electron volts by its equivalent $kT$ joules.

The mass has completely disappeared from this equation, and it is seen that the energy distribution depends solely upon the temperature of the gas (if the obvious proportionality factor $N$, the concentration, is overlooked). Thus, two different gases at the same temperature will be distributed according to identical distribution functions. Equation (8-38) implies that an intimate connection exists between the temperature of a gas and the energy of its molecules.

If, for convenience, the dimensionless variable

$$\eta = \frac{E}{E_T} \qquad \text{or} \qquad E = \eta E_T \qquad (8\text{-}39)$$

and

$$\rho_\eta \equiv \frac{dN_\eta}{d\eta} \qquad (8\text{-}40)$$

are introduced into Eq. (8-38), this expression becomes

$$\rho_\eta = \frac{2N}{\sqrt{\pi}} \eta^{\frac{1}{2}} \epsilon^{-\eta} \qquad (8\text{-}41)$$

A graph of this function is shown in Fig. 8-1 and is repeated, for convenience, in Fig. 8-6. It is possible to deduce the energy distribution of *all* nondegenerate gases at *all* temperatures from this one expression.

If a procedure similar to that used in the previous section is employed in order to determine the position of the maximum of the energy function,

it is easy to show that the most probable energy is $E_T/2$ electron volts, or $kT/2$ joules.

*Example.* Calculate the average energy of a gas molecule.

*Solution.* Proceeding according to the method of Sec. 5-10, the average energy is given by the expression

$$ E = \frac{\int_0^\infty E \, dN_E}{\int_0^\infty dN_E} = \frac{\int_0^\infty \eta E_T \, dN_\eta}{\int_0^\infty dN_\eta} = \frac{2E_T}{\sqrt{\pi}} \int_0^\infty \eta^{\frac{3}{2}} \epsilon^{-\eta} \, d\eta $$

By making a change of variable, namely, $\eta = x^2$, this expression reduces to one of the standard forms given in Appendix V. The evaluation gives

$$ E = \tfrac{3}{2} E_T \text{ ev} \qquad \text{or} \qquad \tfrac{3}{2} kT \text{ joules} \tag{8-42} $$

The same result may be obtained somewhat more simply as follows: The energy is

$$ eE = \tfrac{1}{2} m v^2 \qquad \text{joules} $$

Therefore

$$ e\bar{E} = \tfrac{1}{2}\overline{m v^2} = \tfrac{1}{2}m(\tfrac{3}{2}v_c^2) = \tfrac{3}{2}kT $$

according to the results of Eqs. (8-37) and (8-32) for $\overline{v^2}$ and $v_c^2$, respectively.
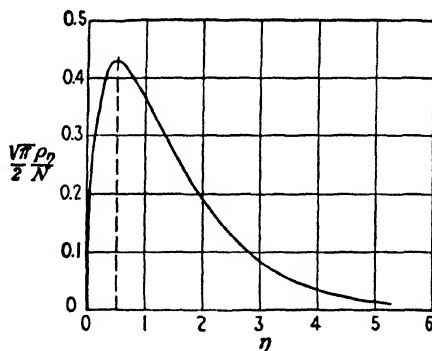


FIG. 8-6. The Maxwell-Boltzmann energy distribution function, plotted in terms of the variable $\eta = E/E_T$.

This example illustrates clearly the importance of the rms velocity. It is this value which expresses the average energy of the molecules, just as the rms current is involved in the determination of the average power dissipated in an electrical circuit.

Consider the expression

$$ \tfrac{1}{2}m v^2 = \tfrac{1}{2}m(v_x{}^2 + v_y{}^2 + v_s{}^2) $$

The average total energy equals the sum of the average $X$-directed, the average $Y$-directed, and the average $Z$-directed energies. A little thought should convince the reader that the average value of a sum of functions is equal to the sum of the average values of these functions. Further, since the distribution function is symmetrical in $v_x$, $v_y$, and $v_z$, then,

$$\tfrac{1}{2}\overline{mv_x^2} = \tfrac{1}{2}\overline{mv_y^2} = \tfrac{1}{2}\overline{mv_z^2}$$

Therefore

$$\tfrac{1}{2}\overline{mv^2} = 3(\tfrac{1}{2}\overline{mv_x^2}) = \tfrac{3}{2}kT$$

or

$$\tfrac{1}{2}\overline{mv_x^2} = \tfrac{1}{2}kT \qquad \text{joules} \tag{8-43}$$

This relation shows that the average energy to be associated with each degree of freedom of a molecule is the same and is equal to $kT/2$ joules. It expresses a very important law of the kinetic theory of gases, which is known as the *theorem of the equipartition of energy*. The theorem has already been used in connection with Sec. 5-10. This result is general and applies to systems which may possess more than three degrees of freedom.

**8-11. Random Current Density and Diffusion.** The concept of random current density was discussed in Chap. 5 and served as an introduction to the problem of thermionic emission from metals. It is shown in Eqs. (5-31) and (5-33) that the random current density is expressed by the following definite triple integral:

$$J_r = e \int_{v_x=0}^{\infty} \int_{v_y=-\infty}^{\infty} \int_{v_z=-\infty}^{\infty} v_x \, dN_r \tag{8-44}$$

This definite integral becomes, by substituting the MB distribution (8-28) for $dN_r$,

$$J_r = eN \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \int_0^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} v_x \epsilon^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2kT}} \, dv_z$$

This expression is simply the product of three definite integrals, each of which may be easily evaluated. The result is

$$J_r = Ne \left(\frac{kT}{2\pi m}\right)^{\frac{1}{2}} = \frac{Nev_c}{2\sqrt{\pi}} = \frac{1}{4}Ne\bar{v} \tag{8-45}$$

where $J_r$ is expressed in amperes per square meter, $N$ is the number of ions per cubic meter, $e$ is the charge in coulombs on each ion, $\bar{v}$ or $v_c$ is expressed in meters per second, and $m$ is the mass in kilograms.

The number of molecules crossing 1 $m^2$/sec at any point in a gas is $n = \tfrac{1}{4}N\bar{v}$. If the gas is all at one temperature and pressure, then $N$ and $\bar{v}$ are independent of position. Hence the same number of particles pass a

given cross section in either direction, and there is no net movement or diffusion of the gas molecules. However, if the concentration varies from point to point, then $n$ is different at two adjacent planes $x$ and $x + dx$ (see Fig. 8-2). This results in a movement of molecules in the $X$ direction. Hence a concentration gradient in a gas causes diffusion of the molecules in the direction from the high to the low values of density.

**8-12. Richardson's Equation.** Originally it was assumed that the electrons in a metal obeyed the nondegenerate distribution function, since the existence of energy levels for the electrons in the metal had not yet been postulated. Under these conditions, the analysis of the phenomenon of thermionic emission would lead to Eq. (8-44) with the lower limit of the integral in $v_x$ replaced by $v_{xB}$. This is the velocity corresponding to the height of the potential barrier $E_B$ at the surface of the metal. The existence of the potential barrier at the surface of the metal had been explained in terms of the image-force concept of Sec. 4-7.

The change in the limit of the integral does not complicate its evaluation. The final expression for the thermionic current density, first deduced by Richardson,[3] becomes

$$J_{th} = A'T^{\frac{1}{2}}\epsilon^{-E_B/E_T}$$

$$= A'T^{\frac{1}{2}}\epsilon^{-b'/T} \tag{8-46}$$

where $A' \equiv eN\sqrt{k/2\pi m}$ and where $b' \equiv eE_B/k$. This equation differs from the corresponding equation determined by using the FDS function in the appearance of the factor $T^{\frac{1}{2}}$ instead of $T^2$, in the factor $A'$ instead of $A_0$, and in the appearance of $E_B$ instead of $E_W$ in the exponent. This means that $E_B$ was formerly considered to be the net work function.

As discussed in Sec. 6-3, it is experimentally impossible to favor one or the other of the two thermionic-emission equations because the greatest variation with temperature is that of the exponential factor. Theoretically, for the reasons outlined in Chap. 5, the Dushman $T^2$ equation is to be preferred, and it is the $T^2$ equation that is always employed for the determination of the thermionic-emission current density.

**8-13. The Pressure-Energy Equation.** Equation (8-1) will now be derived. It must be remembered that the pressure of a gas is the change in momentum of all the particles that strike unit area of the container in unit time.

According to Eq. (5-31), the number of particles that strike unit area of surface in unit time with $X$-directed velocities in the range between $v_x$ and $v_x + dv_x$ and with all possible $Y$- and $Z$-directed velocities is

$$dN_x' = \int_{v_y = -\infty}^{\infty} \int_{v_z = -\infty}^{\infty} v_x \, dN_r \tag{8-47}$$

where the $X$ direction is that normal to the surface under consideration. The momentum of a particle perpendicular to the wall is $mv_x$ before it strikes the surface, and if an elastic collision is assumed, it will be $-mv_x$ after the collision. Hence, the change in momentum per molecule is $2mv_x$ for all particles having $X$-directed velocities in the range between $v_x$ and $v_x + dv_x$. Therefore the total change in momentum per second per square meter of all molecules having any velocity $v_x$ (directed, of course, only *toward* the surface) is obtained by multiplying Eq. (8-47) by $2mv_x$ and integrating over $v_x$ from zero to infinity. This yields, for the pressure $p$,

$$p = \int_0^\infty dv_x \int_{-\infty}^\infty dv_y \int_{-\infty}^\infty dv_z (2mv_x^2 \rho_r)$$

$$= \int_{-\infty}^\infty dv_x \int_{-\infty}^\infty dv_y \int_{-\infty}^\infty dv_z (mv_x^2 \rho_r) \qquad (8\text{-}48)$$

The integral over $v_x$ between the limits 0 and $\infty$ is just one-half of that between the limits $-\infty$ and $\infty$, since the integrand is an even function of $v_x$.

Although these integrals may readily be evaluated, this is not necessary, since a simple correlation between this expression and that for the average energy per molecule is possible. Thus,

$$e\overline{E} = \tfrac{1}{2}m(\overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}) = 3(\overline{\tfrac{1}{2}mv_x^2}) \qquad \text{joules}$$

which, in integral form, may be written as

$$e\overline{E} = \frac{\dfrac{3}{2}\int mv_x^2\, dN_r}{\int dN_r} = \frac{3}{2N} \int_{-\infty}^\infty dv_x \int_{-\infty}^\infty dv_y \int_{-\infty}^\infty dv_z (mv_x^2 \rho_r) \qquad (8\text{-}49)$$

Comparing Eq. (8-48) with Eq. (8-49), it is seen that the simple relationship (8-1), *viz.*,

$$p = \tfrac{2}{3}Ne\overline{E} \qquad (8\text{-}50)$$

exists, where $p$ is the pressure in newtons per square meter, $e\overline{E}$ is the average total energy per molecule, in joules, and $N$ is the number of molecules per cubic meter. In this derivation the explicit form of the classical distribution function $\rho_r$ is not needed. However, the following two properties are employed: $\rho_r$ is an even function of $v_x$, and it is symmetrical in $v_x$, $v_y$, and $v_z$. Thus, the relationship (8-50) is valid for particles obeying any distribution function that satisfies these two conditions. In particular, it applies for an FDS distribution.

**8-14. The Specific Heat of Gases.** The average energy of all the molecules contained in 1 mole of a gas *per degree of freedom* is

$$e\overline{E}_{\text{mole}} = (\tfrac{1}{2}kT)N_0 \qquad \text{joules}$$

since there are $N_0$ molecules per mole, and, by Eq. (8-43), the average energy per degree of freedom is $\frac{1}{2}kT$. This becomes, by Eq. (8-5),

$$e\bar{E}_{\text{mole}} = \tfrac{1}{2}RT \qquad \text{joules} \qquad (8\text{-}51)$$

The specific heat at constant volume $C_v$ is the rate of change of this expression with respect to the temperature, or

$$C_v = \tfrac{1}{2}R = \tfrac{1}{2} \times 1.987 \doteq 1 \text{ cal/(deg)(mole)(deg of freedom)} \quad (8\text{-}52)$$

The number of degrees of freedom of a gas molecule depends upon whether it is monatomic or polyatomic. This result has been experimentally verified.

### PROBLEMS

**8-1.** Calculate the concentration of gas molecules in a "vacuum" tube at $10^{-6}$ mm Hg pressure and room temperature,

a. If the gas is nitrogen.

b. If the gas is neon.

**8-2.** If 1.0 g of argon gas is confined in a volume of 1 liter at 40°C, what is the pressure of the gas?

**8-3.** a. Calculate the mean free path of a mercury molecule in a rectifier operating at a mercury condensation temperature of 40°C (see Fig. 11-1).

b. Calculate the number of collisions per second.

c. Calculate the random current density. Assume that $\frac{1}{10}$ of 1 per cent of the molecules are ionized.

**8-4.** A gas photocell contains argon at a pressure of 0.5 mm Hg and room temperature of 20°C. On the average, how many collisions are made by an electron leaving the cathode on its way to the anode, which is 0.015 m away?

**8-5.** Calculate the number of collisions per second made by a neon molecule, if the pressure is 1 mm Hg and the temperature is 100°C.

**8-6.** To what pressure must a diode be evacuated in order that only 1 out of each 1,000 electrons leaving the cathode makes a collision before reaching the anode, a distance of 1 cm away? Assume that the tube is at 30°C and that the gas in the tube is nitrogen.

**8-7.** Show that the effective radius of an atom for collision processes is

$$r = 0.03P_c^{\frac{1}{2}} \qquad \text{A}$$

where $P_c$ is the probability of collision (in reciprocal meters) under standard conditions of 1 mm Hg pressure and 0°C.

**8-8.** What fraction of the molecules per cubic meter in a gaseous tube have speeds within 1 per cent of the most probable speed, i.e., within the range of $0.99v_c$ to $1.01v_c$?

**8-9.** A gas tube contains argon at a pressure of 15 mm Hg and 25°C. How many molecules per cubic meter have energies within the range of 0.100 to 0.101 ev?

**8-10.** At what height above the earth will the density of the nitrogen in the atmosphere be 90 per cent of its value at the earth's surface? Assume that the atmosphere is at a constant temperature of 50°C.

**8-11.** At what height above the earth is the concentration of oxygen in the atmosphere equal to one-fifth that of the nitrogen? Assume that at the earth's surface the atmosphere consists of 20 per cent oxygen and 80 per cent nitrogen and that the temperature is constant at 50°C.

**8-12.** Calculate the average value of the reciprocal speed of the oxygen molecules in the atmosphere at room temperature.

**8-13.** Calculate the average value of the square of the energy of a neon molecule at 1 mm Hg pressure and 50 °C.

**8-14.** Obtain the MB energy function directly from the FDS energy function without recourse to the distribution in $\tau$ space.

**8-15.** Neon gas at 25 °C and at atmospheric pressure is sealed in a very large container. A small hole $10^{-6}$ m$^2$ in area is cut in the side of the container. How many molecules per second escape from the container through the opening?

**8-16.** Prove, using the MB distribution, that the average energy of particles crossing a surface is $2E_T$ electron volts.

**8-17.** Show that Eq. (8-4) is equivalent to Eq. (8-3), without using Loschmidt's number.

**8-18.** Find the numerical value of $R$ (the gas constant) in atmosphere-liters per degree Kelvin per mole.

**8-19.** Calculate the pressure of the electron gas in a tungsten metal at absolute zero. Does this enormous pressure seem illogical physically?

**8-20.** *a.* Show that the distribution function for gas molecules having $X$-directed velocity components in the range between $v_x$ and $v_x + dv_x$ and with any velocity in the other two directions is

$$dN_x = N \left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}} \epsilon^{-mv_x^2/2kT} \, dv_x$$

$$= \frac{N}{\sqrt{\pi}} \epsilon^{-w^2} \, dw$$

where $w = v_x/v_c$ and the other symbols have the meanings given in the text.

*b.* Derive this distribution from the corresponding distribution function for an FDS gas (see Prob. 5-12).

**8-21.** *a.* Show that the average $X$-directed velocity of a gas molecule is zero if the averaging is done over all molecules and is $v_c/\sqrt{\pi}$ if the averaging is done over those molecules traveling only in one direction (say, the positive $X$ direction).

*b.* Show that the rms $X$-directed velocity of a gas molecule is $v_c/\sqrt{2}$.

**8-22.** Prove that the average normal-to-the-surface energy of the gas molecules striking the walls of the enclosing container is $kT$ joules.

**8-23.** Show that the number of gas molecules that strike unit area of the container in unit time with normal-to-the-surface velocities greater than $v$ is

$$\frac{Nv_c}{2\sqrt{\pi}} \epsilon^{-(v/v_c)^2}$$

where $N$ is the concentration and $v_c$ is the characteristic speed.

**8-24.** *a.* What is the most probable $X$-directed velocity of the molecules in a container of gas?

*b.* If there is a tiny hole in the side of the container, what is the most probable $X$-directed velocity of the molecules that escape through the hole?

Express the answers in terms of the characteristic speed $v_c$. Why are the answers to parts $a$ and $b$ not the same?

**8-25.** *a.* Show that the fraction of the molecules in a gas that have energies in excess of the energy $E$ is

$$\Psi(\zeta) = \frac{4}{\sqrt{\pi}} \int_\zeta^\infty u^2 \epsilon^{-u^2} \, du = \frac{2}{\sqrt{\pi}} \zeta \epsilon^{-\zeta^2} + 1 - \operatorname{erf} \zeta$$

where $\zeta = \sqrt{E/E_T}$ and where erf $\zeta$ is known as the *error function* and is defined by

$$\text{erf } \zeta = \frac{2}{\sqrt{\pi}} \int_0^\zeta \epsilon^{-u^2} du$$

(Numerical values of the error function can be found in B. O. Pierce, "A Short Table of Integrals," 3d ed., Ginn & Company, Boston, 1929.)

*b.* Show that the fraction of the molecules in a gas that have speeds in excess of the speed $v$ is $\Psi(v/v_c)$, where $\Psi$ is the function defined in part *a*.

## REFERENCES

1. See, for example, LOEB, L. B., "Kinetic Theory of Gases," 2d ed., Chap. IV, McGraw-Hill Book Company, Inc., New York, 1934.
2. LOEB, *op. cit.*, p. 95.
3. RICHARDSON, O. W., *Proc. Cambridge Phil. Soc.*, **11**, 286, 1901.

### General References

DOW, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

JEANS, J. H.: "An Introduction to the Kinetic Theory of Gases," The Macmillan Company, New York, 1940.

JEANS, J. H.: "Dynamical Theory of Gases," Cambridge University Press, London, 1916.

KENNARD, E. H.: "Kinetic Theory of Gases," McGraw-Hill Book Company, Inc., New York, 1938.

LOEB, L. B.: "Kinetic Theory of Gases," 2d ed., McGraw-Hill Book Company, Inc., New York, 1934.

SLEPIAN, J.: "Conduction of Electricity in Gases," Westinghouse Electric & Manufacturing Company, Educational Department, Course 38, 1933.

TOLMAN, R. C.: "Statistical Mechanics with Applications to Physics and Chemistry," Reinhold Publishing Corporation, New York, 1927.

# CHAPTER 9

## FUNDAMENTAL PROCESSES IN GASES

A GAS, according to the classical kinetic theory, was envisaged as a system of solid spherical molecules which collide elastically with each other and with the walls of the confining container. However, whenever electronic gaseous devices are discussed, it becomes necessary to abandon this simple model of solid atoms and introduce the concept of ionization. This immediately carries with it the hypothesis that an atom has certain loosely bound electrons which can be torn away from it, thereby leaving the atom in an ionized state. It is necessary to elaborate on this modern theory of matter, since without it, the phenomena that occur in gaseous discharges would be inexplicable.

**9-1. The Bohr-Rutherford Theory of the Atom.** According to this theory, the atom consists of a nucleus of positive charge that contains nearly all of the mass of the atom. Surrounding this central positive core are negatively charged electrons. This concept of the atom was developed by Rutherford [1] in order to explain the results of his experiments on the scattering of $\alpha$-particles (ionized helium particles which are emitted spontaneously in radioactive transformations) in passing through matter. Based on the results of theory and experiment, the diameter of the nucleus is found to be of the order of $10^{-15}$ m. This is comparable to the diameter of the electron.

As a specific illustration of the foregoing atomic model, consider the hydrogen atom. This atom consists of a positively charged nucleus (a proton) and a single electron. The charge on the proton is positive and is equal in magnitude to the charge on the electron. Therefore the atom as a whole is electrically neutral. Because the proton carries practically all of the mass of the atom, it will remain substantially immobile, whereas the electron will move about it in a closed orbit. The force of attraction between the electron and the proton follows Coulomb's law. It can be shown from classical mechanics that the resultant closed path will be a circle or an ellipse under the action of such a force. This motion is exactly analogous to that of the planets about the sun. Although the force of attraction in this latter case is gravitational rather than electrical in nature, nevertheless the resultant motions are similar since the forces in

234

both systems vary in the same way, *viz.*, the force varies inversely as the square of the distance between the particles.

Assume, therefore, that the orbit of the electron in this planetary model of the atom is a circle, the nucleus being supposed fixed in space. It is a simple matter to calculate its radius in terms of the total energy, $W$ joules, of the electron. The force of attraction between the nucleus and the electron is $e^2/4\pi\epsilon_0 r^2$ newtons, when the electronic charge $e$ is in coulombs, the separation $r$ between the two particles is in meters, and $\epsilon_0$ is the permittivity of free space. By Newton's second law of motion, this must be set equal to the product of the electronic mass $m$ in kilograms and the acceleration $v^2/r$ toward the nucleus, where $v$ is the speed of the electron in its circular path, in meters per second. Then

$$\frac{e^2}{4\pi\epsilon_0 r^2} = \frac{mv^2}{r}$$

or

$$\frac{e^2}{4\pi\epsilon_0 r} = mv^2 \tag{9-1}$$

Furthermore, the potential energy of the electron at a distance $r$ from the nucleus is $-e^2/4\pi\epsilon_0 r$ joules, and its kinetic energy is $\frac{1}{2}mv^2$ joules. Then, according to the conservation of energy,

$$W = \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r} \quad \text{joules} \tag{9-2}$$

Combining this expression with (9-1) produces

$$W = -\frac{e^2}{8\pi\epsilon_0 r} \quad \text{joules} \tag{9-3}$$

which gives the desired relationship between the radius and the energy of the electron. This equation shows that the total energy of the electron is always negative and that the energy of the electron at an infinite distance from the nucleus is zero. The negative sign arises because the potential energy has been so chosen as to be zero when $r$ is infinite. This expression also shows that the energy of the electron becomes smaller (*i.e.*, more negative) as it approaches closer to the nucleus.

The foregoing discussion of the planetary atom has been considered only from the point of view of classical mechanics. If it is recalled that the system is electrical in nature, then a radiating system is actually under survey. This is so because an accelerated charge must radiate energy, in accordance with the classical laws of electromagnetism. Also, if the charge is performing oscillations of a frequency $f$, then the radiated energy will

also be of this frequency. Hence, classically, it must be concluded that the frequency of the emitted radiation equals the frequency with which the electron is rotating in its circular orbit.

There is one feature of this picture that cannot be reconciled with experiment. If the electron is radiating energy, then its total energy must decrease by the amount of this emitted energy. As a result the radius $r$ of the orbit must decrease, in accordance with Eq. (9-3). Consequently, as the atom radiates energy, the electron must move in smaller and smaller orbits, eventually falling into the nucleus. Since the frequency of oscillation depends upon the size of the circular orbit, the energy radiated would be of a gradually changing frequency. Such a conclusion, however, is incompatible with the sharply defined frequencies of spectral lines.

This difficulty was resolved by Bohr in 1913.[2] He enunciated two postulates which are fundamental to the interpretation of spectra and which have since become basic in atomic and nuclear studies. Bohr made the following assumptions:

1. Not all energies as given by classical mechanics are possible, but the atom can possess only certain discrete energies. While in orbits corresponding to these discrete energies, the electron does *not* emit radiation. Under such conditions the electron is said to be in a "stationary," or nonradiating, state.

2. In a transition from one stationary state corresponding to a definite energy $W_2$ to another stationary state, with an associated energy $W_1$, radiation will be emitted. The frequency of this radiant energy is given by

$$f = \frac{W_2 - W_1}{h} \qquad \text{cps} \qquad (9\text{-}4)$$

where $h$ is Planck's constant in joule-seconds and where the $W$'s are expressed in joules.

Bohr gave the quantitative rule whereby the energies of the stationary states could be calculated. Then, upon making use of Eq. (9-4), the exact frequencies found in the hydrogen spectrum were obtained (see Prob. 9-2). This remarkable achievement of the Bohr theory was shadowed by the rather mystical nature of the laws involved. Thus it was necessary to introduce a separate atomic mechanics, the principles of which differed from those of classical mechanics, in order to explain the hydrogen atom. It should be pointed out that the semiclassical model of Bohr has since been abandoned for the modern wave-mechanical and matrix-mechanical (known as *quantum-mechanical*) methods, which are less pictorial and more mathematical than the Bohr theory.

Justification of Bohr's radical assumptions can be given in terms of these more advanced theories. Furthermore, the newer theories go much further than the old. Not only do they explain the spectra of complex atoms, but they also furnish an explanation of other related atomic phe-

nomena that had remained inexplicable. However, despite the many innovations that have been introduced into atomic theory, the original ideas of Bohr as expressed in the two postulates have remained essentially unchanged.

**9-2. The Electronic Structure of the Elements.** Since mercury vapor is the most common gas used in gaseous rectifier tubes, it will be used to serve as an illustration of the application of the Bohr postulates. However, the mercury atom contains 80 electrons and so is a much more complicated system than the hydrogen atom with its 1 electron. Hence, as a necessary digression, the electronic structure of the elements will be discussed briefly in this section. However, to present the mass of direct and indirect chemical and physical evidence that supports the picture to be described below would take us too far astray. Hence, only the conclusions that are drawn from this evidence will be presented.



FIG. 9-1. The charge distribution in the sodium atom. The nucleus consists of a positive core of 11 electronic units. The heavy shading represents the 10 tightly bound electrons, the light shading the loosely bound (valence) electron. (W. Shockley; see Fig. 4-1.)

Normally the atom is neutral in charge, the number of electrons that surround the nucleus in the form of a "cloud" being just sufficient to neutralize the positive charge of the nucleus. Such a system is illustrated in Fig. 9-1. The exact motions of the electrons in this cloud are of no importance in this discussion. For purposes of visualization and insight, the individual electrons may be considered to describe, very approximately, the Bohr circles or ellipses.

If a deficiency of electrons within an atom exists for any reason whatsoever, the atom as a whole exhibits properties that are characteristic of positively charged bodies. If, correspondingly, the atom contains an excess of electrons, it then exhibits properties of negatively charged bodies. Atoms that have a dearth of electrons, being positively charged bodies, are classified as positive ions; those that have an excess of electrons are negatively charged and are classified as negative ions.

Although the constitution of the nucleus is not under survey, it may be well to mention that the nucleus consists of protons and neutrons. Each proton has associated with it a charge equal to that of the electron, though of positive sign. The neutron is an uncharged particle that has a mass practically equal to that of the proton (1,837 times the mass of the electron). The atomic number $Z$ of an atom gives the number of protons contained within the nucleus and consequently the number of electrons surrounding the nucleus under normal conditions. The atomic weight of an atom gives the weight of the nucleus, since the contribution of the weight of the electrons to the total weight is negligible. The difference between

the total weight of the atom and the weight of the protons contained therein is accounted for largely by the weight of the neutrons.

The electrons surrounding the nucleus orient themselves with respect to it in certain groups, called *shells*. Each shell may contain only a definite maximum number of electrons. Once any particular group is filled with its maximum possible number of electrons, another shell must be formed. The series of shells are located at progressively increasing distances from the nucleus. The inner-shell electrons are very strongly bound to an atom and cannot be easily removed.

The electrons that surround the nucleus in the respective shells may exist in any one of a number of energy levels. These energy levels are similar to those which exist in metals. In the case of the atom, however, the separation of the atomic energy levels, which depends upon the particular atom, is measured in electron volts, whereas the separation of the electronic energy levels in metals may be shown to be a very small fraction of an electron volt (of the order of $10^{-22}$ ev). When the electrons in an atom occupy the lowest possible energy levels, the atom is said to be in its "normal" state.

**9-3. Atomic Energy Levels.** The amount of energy involved in discharge devices is seldom large enough to perturb the inner-shell electrons. This greatly simplifies the discussion of the behavior of complex atoms. For example, most of the activity of the sodium atoms in a sodium-vapor lamp results from the single outermost electron. The other 10 electrons remain undisturbed by the discharge. Similarly, 78 of the 80 electrons that surround the nucleus of a mercury atom are tightly bound. Hence any energy changes that the atom undergoes affect only the two outermost electrons.

Though it is theoretically possible to calculate the various energy states of the atoms of the simpler elements, these energies must be obtained indirectly from spectroscopic data in the more complicated cases. Thus, starting with the frequencies $f$ of the observed spectral lines, Bohr's relationship (9-4) is used to deduce the energy values $W_1$ and $W_2$. This method of deduction is obviously not a unique one, but the problem that the spectroscopist sets for himself is to account for all of the observed lines in the spectrum of an element with the minimum number of assumed energy states. There are other clues that aid him in this search, although they need not concern us here. Several such clues will be mentioned in subsequent sections. The result of an analysis of this type applied to mercury is shown in Fig. 9-2. This is called an *energy-level diagram*.

The numbers to the left of the horizontal lines give the energy of these levels in electron volts. The arrows represent some of the transitions that have been found to exist in actual spectra, the attached numbers giving the wave length of the emitted radiation, expressed in angstrom units

$(10^{-10}$ m).   The light emitted in these transitions gives rise to the lumi-
nous character of the gaseous discharge.   However, all of the emitted
radiation need not appear in the form of visible light but may exist in the
ultraviolet or infrared regions.   The meaning of the broken lines will be
explained in Sec. 9-11.

It is usual practice to express the energy values of the energy levels in
electron volts, as has been done, rather than in joules, as expressed in Eq.
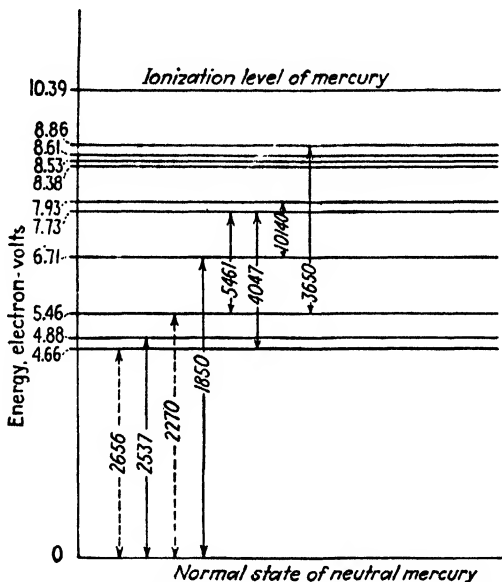


FIG. 9-2.   The lower energy levels of atomic mercury.

(9-4).   Also, it is more common to specify the characteristics of the emitted
radiation in terms of wave length, rather than in terms of the frequency of
the light.   The wave length and the frequency of the radiation are related
by the expression $\lambda f = c$, where $\lambda$ is the wave length in meters, $f$ is the
frequency in cycles per second, and $c$ is the velocity of light in meters per
second.   If the $E$'s are expressed in electron volts and $\lambda$ is in angstrom
units, Eq. (9-4) may be rewritten in the form

$$\lambda = \frac{12,400}{E_2 - E_1} \quad A \qquad (9\text{-}5)$$

Since only differences of energy enter into this expression, the zero state
may be chosen at will.   It is convenient and customary to choose the
lowest energy state as the zero level.   This was done in Fig. 9-2.   The

lowest energy state is called the "normal" level, and the other stationary states of the atom are called "excited," "radiating," "critical," or "resonance" levels.

The most intense line in the mercury spectrum is that resulting from the transition from the 4.88-ev level to the zero state. The emitted radiation, as calculated from Eq. (9-5), is 12,400/4.88 = 2,537 A, as indicated in the diagram. It is primarily this line that is responsible for the ultraviolet burns which arise from mercury discharges.

**9-4. The Photon Nature of Light.** The mean life of an excited state ranges from $10^{-7}$ to $10^{-10}$ sec, the excited electron returning to its previous state after the lapse of this time.[3] In this transition, the atom must lose an amount of energy equal to the difference in energy between the two states that it has successively occupied, this energy appearing in the form of radiation. According to the principle of Bohr, this energy is emitted in the form of a photon of light, the frequency of this radiation being given by Eq. (9-4). The term *photon* denotes an amount of radiant energy equal to Planck's constant $h$ times the frequency.

The photon concept of radiation may be difficult to comprehend at first. Classically, it was believed that the atoms were systems that emitted radiation *continuously* in all directions. According to the foregoing theory, however, this is not true, the emission of light by an atom being a discontinuous process. That is, the atom radiates only when it makes a transition from one energy level to a lower energy state. In this transition, it emits a definite amount of energy of one particular frequency. That is, it emits one photon $hf$ of light. Of course, when a luminous discharge is observed, this discontinuous nature of radiation is not suspected because of the enormous number of atoms that are radiating energy and, correspondingly, because of the immense number of photons that are emitted in unit time.

*Example.* Given a 50-watt mercury-vapor lamp. Assume that 0.1 per cent of the electrical energy supplied to the lamp appears in the ultraviolet line, 2,537 A. Calculate the number of photons per second of this wave length emitted by the lamp.

*Solution.* The energy per photon is, according to Eq. (9-5),

$$E = \frac{12,400}{2,537} = 4.88 \text{ ev/photon}$$

The total power being transformed to the 2,537-A line is 0.05 watt, or 0.05 volt × coulomb/sec. Since the charge per electron is $1.60 \times 10^{-19}$ coulomb, then the power radiated is

$$\frac{0.05 \text{ volt} \times \text{coulomb/sec}}{1.60 \times 10^{-19} \text{ coulomb/electron}} = 3.12 \times 10^{17} \text{ ev/sec}$$

Hence, the number of photons per second is

$$\frac{3.12 \times 10^{17} \text{ ev/sec}}{4.88 \text{ ev/photon}} = 6.40 \times 10^{16} \text{ photons/sec}$$

This is an extremely large number.

This calculation shows why the discontinuous nature of light is never perceptible when macroscopic (large-scale) phenomena are being considered. There are far too many photons for their individuality to be observed. However, any attempts to understand the generation or absorption of light or the interaction between radiation and matter (say, electrons) are futile if the photon nature of light is not presupposed.

One must not gain the impression that the photon theory of light originated with Bohr. It was invented by Planck in 1900. In fact, the birth of the quantum theory occurred when Planck made the fundamental assumption that radiant energy can exist only in discrete quantities, instead of being continuous. With the aid of this concept of quanta of magnitude $hf$, Planck was able to correlate satisfactorily the theory and measurements of infrared radiation. The same fundamental assumption was employed by Bohr in order to correlate theory and experiment in the field of spectroscopy. Einstein applied the same quantum hypothesis to give a description of the phenomena involved in the interaction of light and matter and thereby put the mechanism of photoelectricity on a firm theoretical basis, as will be described in Chap. 15. So much experimental evidence exists in favor of this quantum hypothesis that little doubt exists as to its validity.

TABLE 9-1

| Gas or vapor | Ionizing potential, ev | First radiative excitation potential, ev | First resonance wave length, A | First metastable excitation potential, ev | Discharge color |
|---|---|---|---|---|---|
| A | 15.7 | 11.6 | 1,065 | . . . . . | Blue |
| Cd | 8.96 | 3.78 | 3,260 | . . . . . | Red |
| He | 24.5 | 20.6 | 600 | 19.8 | Yellow |
| Hg | 10.4 | 4.88 | 2,537 | 4.66 | Purple |
| Na | 5.12 | 2.09 | 5,896 | . . . . . | Yellow |
| Ne | 21.5 | 16.6 | 743 | 16.6 | Orange |

**9-5. Ionization.** It is now possible to obtain a better insight into the process of ionization, mention of which has been made on several occasions. As the most loosely bound electron of an atom is given more and more energy, it moves into stationary states which are farther and farther away from the nucleus. When its energy is large enough to move completely out of the field of influence of the ion, it becomes "detached" from it. The energy required for this process to occur is called the *ionization potential* and is represented as the highest state in the energy-level diagram.

From an inspection of Fig. 9-2, this is seen to be 10.39 volts for mercury. The alkali metals have the lowest ionization potentials, whereas the inert gases have the highest values, the ionizing potentials ranging from approximately 4 to 25 ev.

Table 9-1 gives the values of the ionization potential, the first excitation potential, and the corresponding first resonance wave length of some of the elements that play important roles in discharge tubes.

**9-6. Multiple Ionization.** Once an atom has been ionized, it is possible for it to receive additional energy so that a second electron will be ejected from the ion. This gives rise to a doubly ionized atom. If a third electron is removed, then a trebly ionized particle remains. Bleakney [4] has detected as many as five stages in the ionization of mercury vapor.

The amount of energy that is necessary to free an electron from an ion depends upon the number of electrons that have already been removed and also upon the electronic configuration of the resulting ion. Since the inner-shell electrons are more tightly bound than those which are at a greater distance from the nucleus, each successive ionization requires more energy than the previous one. For example, an energy of 10.4 ev is required to singly ionize mercury vapor so that $Hg^+$ is formed, whereas an energy of approximately 230 ev is required to produce $Hg^{+++++}$.

It is not possible for multiple ionization to occur in a discharge device unless the physical conditions are properly adjusted. If the excitation arises from electron impact with the atom, the pressure of the gas must be sufficiently low so that the electron can gain the required high energy within a mean free path.

**9-7. Collisions of Electrons with Atoms.** The foregoing discussion has shown that in order to excite or ionize an atom, energy must be supplied to it. This energy may be supplied to the atom in various ways. The most important method by which an energy interchange occurs in a gaseous discharge is through the medium of electron impact. Other methods of ionization or excitation of atoms will be considered below.

Suppose that an electron is accelerated by the potential field applied to a discharge tube. When this electron collides with an atom, one of several effects may occur. A slowly moving electron suffers an "elastic" collision, *i.e.*, one that entails an energy loss only as required by the laws of conservation of energy and momentum. The direction of travel of the electron will be altered by the collision although its energy remains substantially unchanged. This follows from the fact that the mass of the gas molecule is large compared with that of the electron.

If the electron possesses sufficient energy, the amount depending upon the particular gas or vapor present, it may transfer enough of its energy to the atom to elevate the outer-shell electron to one of the higher quantum states. Collisions, in which all or a portion of the kinetic energy of an

incident electron before impact with a gas molecule or atom is used in raising an electron in the atom from a state that it normally occupies to a higher unoccupied state, are known as collisions of the *first kind*. The amount of energy necessary for this process, when expressed in electron volts, is the excitation, or radiation, potential of the atom. If the impinging electron possesses a higher energy, say an amount at least equal to the ionization potential of the gas, it may deliver this energy to an electron in the outermost shell of the atom and completely remove it from the parent atom. Three charged particles result from such a collision, two electrons and a positive ion.

It must not be presumed that the incident electron must possess an energy corresponding exactly to the energy of a stationary state in an atom in order to raise the atom into this level. If the bombarding electron has gained more than the requisite energy from the electric field to raise an atom into a particular energy state, then the amount of energy in excess of that required for excitation will be retained by the incident electron as kinetic energy after the collision. Or if the process of ionization has taken place, the excess energy divides between the two electrons. This is so because the energy gained by the ion as a result of the impact is negligible because of its large mass compared with the electronic mass.



FIG. 9-3. An experimental method for determining stationary states within an atom.

Shortly after the existence of stationary states within an atom was postulated by Bohr, direct experimental confirmation of this hypothesis was obtained by Franck and Hertz [5] using a method of electronic collisions based on the foregoing discussion. The circuit, in simplified form, is shown in Fig. 9-3. Electrons from a hot cathode $K$ are accelerated by means of the potential $E_a$ toward the grid $G$. A small retarding potential $E_r$ (say, 0.5 volt) is applied between the grid and the plate. The experiment [6] consists in observing the current indicated by the meter $M$ in the plate circuit as the accelerating potential $E_a$ is increased from zero.

According to Bohr's theory, the electrons must make elastic collisions for all values of $E_a$ below the first excitation potential. Hence, as $E_a$ is varied from 0.5 volt upward to the first critical energy, the electrons are able to surmount the potential retardation between $G$ and $A$ and are recorded by the plate meter. The current thus rises continuously as $E_a$ increases. However, when the potential $E_a$ reaches the first excitation potential, some of the electrons transfer their energy to atoms, thereby exciting them. As a result of the collision process, these bombarding electrons no longer possess sufficient energy to overcome the retarding field $E_r$.

Hence the current indicated by the meter $M$ decreases, since these electrons do not reach the plate.

As $E_a$ is further increased, the residual energy of the electrons increases and these electrons again reach the plate despite the inelastic collision that they have made. Thus the curve of current vs. accelerating potential rises until the lowest excitation potential is reached, when it takes a sudden dip and then commences to rise again. When $E_a$ reaches the second excitation potential, a second dip occurs in the curve. At this point, some of the electrons lose most of their energy and so cannot surmount the retarding-potential barrier between the grid and the plate. It is possible to detect several of the critical potentials in this way.

A slight modification of the experimental arrangement shown in Fig. 9-3 must be made in order to observe the ionization potential of the gas directly.[6] The anode is now made 0.5 volt negative with respect to the cathode, instead of being negative with respect to the grid. Hence no electrons emitted from the cathode can reach the plate. However, any positive ions that are formed in the gas will be recorded by the meter $M$. Of course, very few ions exist in the tube until $E_a$ is made equal to the ionization potential, when ionization occurs suddenly and the meter reading increases rapidly.

As soon as the first excitation voltage has been reached in the above experiment, photons will be emitted by the gas. Some of this radiation will be intercepted by the plate, resulting in photoemission of electrons, which will then proceed to the cathode (since this electrode is positive with respect to the plate). This is equivalent to a flow of positive charges toward the plate. Hence the photocurrent will add to the positive-ion current due to ionization. In order to separate the true ionization current from the photocurrent, Davis and Goucher [7] added a screen grid to the tube. If this screen is maintained slightly negative with respect to the plate, any photoelectrons emitted from the plate are repelled by the screen and return to the plate and are not recorded as a current. Any photoelectrons emitted by the screen are collected by the plate and are recorded as a *negative* current. However, when ionization sets in, the positive ions traveling to the plate are recorded as a *positive* current. The onset of ionization is very marked in this experiment, the slope of the volt-ampere characteristic changing abruptly from negative to positive at this potential.

**9-8. Collisions of Photons with Atoms.** Another important method by which an atom may be elevated into an excited energy state is to have radiation fall on the gas. This possibility has probably been predicted by the reader on the basis of the Bohr transition rule. An atom may absorb a photon of frequency $f$ and thereby be elevated from the level of energy $W_1$ to the higher energy level $W_2$, where $W_2 = W_1 + hf$ joules.

That such a process does occur is made evident from the results of the following experiment: [8]  Nonluminous ("cold") sodium vapor is illuminated by a collimated beam of radiation of wave lengths 5,890 and 5,896 A, the strong yellow lines in the sodium spectrum. It is then found that the vapor becomes luminous, yellow light being emitted in all directions. This "scattering" of the original beam is readily explained on the hypothesis that the photons of the incident light are absorbed by the vapor, some of the atoms of which are raised to excited levels. After the lapse of an extremely short time (the mean life of the excited state), the excited atoms return to the normal state, the liberated energy appearing as radiation having the same wave length as that of the absorbed radiation. The emitted beam is no longer collimated, and the vapor appears luminous.

A continuous spectrum may be obtained from an incandescent solid, for example, a heated tungsten filament. The energy levels of the electrons in a solid are so close together that upon excitation an enormous number of lines in all parts of the spectrum are emitted. These lines are so close together that the resulting spectrum is said to be continuous.

Another experiment suggests itself. Instead of using a source of monochromatic light, suppose that the vapor is illuminated with radiation of all frequencies, $i.e.$, a continuous spectrum. The light transmitted through the vapor is then passed through a spectroscope. It is found that the transmitted radiation has certain lines missing. For example, if the gas under consideration is sodium vapor, then the yellow lines 5,890 and 5,896 A, among others, will not be present in the emerging beam. In fact, all the lines that result from transitions from excited states *to the normal level* of sodium will be absent. Such a spectrum is called an "absorption spectrum." Such absorption spectra are extremely useful in locating and assigning the various energy levels within an atom, since *the absorption lines are those that result from transitions from the normal level (of energy zero) to a higher state*. Under these circumstances, the energy of an excited state $E_x$ is specified by the expression $12,400/\lambda_x$, where $\lambda_x$ is the wave length of the observed missing line.

Actually, however, the excited atoms will emit the energy that has been absorbed from the continuous spectrum. Since a fraction of this energy will be in the direction of the spectroscope, then the beam of light that has been transmitted through the vapor will emerge with certain of the lines greatly reduced in intensity. It is these absorption lines that are said to be "missing." In fact, the "dark" lines in the spectrum of the sun, which were first observed by Fraunhofer, find their explanation in this theory. The absorbing atoms are those in the cooler vapor that surrounds the hot core of the sun.

When a photon is absorbed by an atom, the excited atom may return to its normal state in one jump, as has already been discussed, or it may

do so in more than one step. If the atom falls into one or more excitation levels before finally reaching the normal state, then it will emit several photons. These will correspond to energy differences between the successive excited levels into which the atom falls. None of the emitted photons will have the frequency of the absorbed radiation! This *fluorescence* cannot be explained by classical theory but is readily understood once Bohr's postulates are accepted.

If the absorbing vapor is itself a region of a discharge, then many excited atoms will be present. Under this condition many more lines of the absorption spectrum will be missing than if the absorbing gas were "cold." The additional missing lines result from the transitions that take place from one excited level to another.

It is possible for the photons that induce the excitation of certain atoms to originate within the discharge itself. That is, not all of the photons must originate external to the discharge. For example, the following chain process has been observed to occur in mercury vapor: An atom near the center of a discharge tube emits the ultraviolet 2,537-A line. This photon is captured by a neighboring atom, thereby exciting it to the first resonance potential. About $10^{-7}$ sec later, this excited atom returns to its normal state with the emission of a 2,537-A photon. This new photon in turn excites another atom, with the subsequent reemission of a 2,537-A photon. This absorption and emission process may occur several thousand times before an atom near the boundary of the discharge emits a photon that escapes from the discharge. This process has been called the "imprisonment of radiation" in a discharge tube.[9]

An extremely important feature of excitation by photon capture is that *the photon will not be absorbed unless its energy corresponds exactly to the energy difference between two stationary levels of the atom with which it collides.* Consider, for example, the following experiment: The 2,537-A mercury radiation falls on sodium vapor in the normal state. What is the result of this irradiation? The impinging photons have an energy of 12,400/2,537 = 4.88 ev, whereas the first excitation potential of sodium is only 2.09 ev. It is conceivable that the sodium atom may be excited and that the excess energy $4.88 - 2.09 = 2.79$ ev would appear as another photon of wave length 12,400/2.79 = 4,440 A. Actually, however, the 2,537-A line is transmitted without absorption through the sodium vapor, neither of the two lines appearing. It must be concluded, therefore, that the probability of excitation of a gas by photon absorption is negligible unless the energy of the photon corresponds exactly to the energy difference between two stationary states of the atoms of the gas.

If the frequency of the impinging photon is sufficiently high, it may have enough energy to ionize the gas. The photon vanishes, with the appearance of an electron and a positive ion. Unlike the case of photo-excitation, the photon need not possess an energy corresponding exactly

to the ionizing energy of the atom. It need merely possess *at least* this much energy. If it possesses more than ionizing energy, the excess energy will appear as the kinetic energy of the emitted electron and positive ion. It is found by experiment that the maximum probability of photoionization occurs when the energy of the photon is equal to the ionization potential, the probability decreasing rapidly for higher photon energies.

The phenomena of photoexcitation and photoionization discussed above are frequently referred to as the "photoelectric" effect in gases.

**9-9. Collisions of Positive Ions with Atoms.** One might expect that a certain amount of excitation and ionization will result from collisions between positive ions and neutral gas atoms with which they may collide. Actually, however, the probability of ionization by such a collision is very much smaller than that for an electron collision,[10] so that this process plays a minor role in the formation of additional ions. It has been found that it is only when ions have speeds of the order of the speeds of electrons that possess ionizing energy that the probability of ionization by ion impact becomes appreciable. This requires that the energy of the ions be 1,000 ev or more to be able to produce ionization, although ionization of Hg by $Na^+$ ions having 35 ev of energy has been observed.[11]

Unless, therefore, an ion can acquire enough energy between collisions with atoms, it cannot produce ionization. Since the fields in most discharges are much too small to supply a high energy to the ions, the contribution of this process to the production of secondary ionization in gases may be neglected.

**9-10. Other Ionizing Agents.** Minute traces of radioactive contamination are always present in all materials. These emit particles and radiations that are capable of producing ionization. Ultraviolet rays, X rays, and cosmic rays can also produce ionization. Because one or more of these ionizing agents is always present, then a (very small) percentage of any gas is ionized. These few free electrons play a most important role in some types of discharge, as will be seen in the next chapter.

Since the average thermal energy of a gas molecule is $3E_T/2 = T/7,730$ ev, then at the temperatures that are prevalent in most discharges this quantity is smaller than the lowest excitation potential of any atom. However, owing to the wide distribution in energies of the particles in a gaseous discharge, some few may possess enough energy to excite or even to ionize the atoms with which they collide. This condition plays an insignificant role, however, except in very high-temperature high-pressure arcs.

**9-11. Metastable States.** The features of the experiment by which Franck and Hertz showed the existence of stationary states within an atom were discussed in Sec. 9-7. When this experiment is performed on mercury vapor, it is found that the first dip in the curve, indicating the point at which the electrons make inelastic collisions with the mercury

atoms, occurs at 4.66 volts. A transition between this stationary state and the ground level should give rise to radiation having a wave length $12,400/4.66 = 2,656$ A. However, no such line exists in the spectrum of mercury. Moreover, no absorption line of this wave length is present in the absorption spectrum of mercury.

It is evident, therefore, that stationary states may exist which can be excited by electron bombardment, but not by photoexcitation. Levels from which transitions to the normal state, with the resultant emission of radiation, are forbidden are called *metastable* states. The two metastable states of extreme importance in the mercury spectrum are indicated on Fig. 9-2. The forbidden transitions are indicated by dotted arrows on the energy-level diagram. Transitions from higher levels to a metastable state are permitted. Some of these are shown in Fig. 9-2.

The mean life of a metastable state is found to be very much longer than the mean life of a radiating level. Representative times are $10^{-2}$ to $10^{-4}$ sec for metastable states and $10^{-7}$ to $10^{-10}$ sec for radiating levels. The long lifetime of the metastable states arises from the fact that a transition to the normal state with the emission of a photon is forbidden. How then can the energy of a metastable state be expended so that the atom may return to its normal state? One method is for the metastable atom to collide with another molecule and give up its energy to the other molecule as kinetic energy of translation, or potential energy of excitation. It may even ionize the molecule with which it collides (see the next section). Another method is that by which the electron in the metastable state receives additional energy by any of the processes enumerated in the preceding sections. The metastable atom may thereby be elevated to a higher energy state from which a transition to the normal level can take place, or else it may be ionized. If the metastable atom diffuses to the walls of the discharge tube or to any of the electrodes therein, either it may expend its energy in the form of heat or the metastable atoms might induce secondary emission. Because of various methods whereby the metastable atom can lose its energy, the mean life of the metastable state depends upon the surroundings in which the atom finds itself.

Owing to the relatively long lifetime of a metastable state, the probability of cumulative ionization is greatly increased. It is very common in an arc discharge to have ionization take place at a voltage that is appreciably lower than the ionization potential of the gas. For example, the drop across a mercury-arc tube may be considerably less than 10.4 volts. Thus any electron that has fallen through a potential greater than only 5.73 volts which collides with an atom in the 4.66-ev metastable state can ionize the atom (since $5.73 + 4.66 = 10.39$).

**9-12. Collisions of the Second Kind.** What happens when an excited atom collides with another atom, this atom being either in the normal state or else in an excited state? Several resulting processes are possible.

The excited atom may receive sufficient additional energy in this collision to detach an electron, thereby becoming ionized. It may be raised to a higher excited state from which the transition to the normal state may follow. Or the excited atom may fall to a lower state as a result of the impact. The energy released as a result of this transition may cause the unexcited atom to be raised to a higher level or may simply give extra kinetic energy to the colliding atoms. Reactions also occur in which an excited atom collides with a molecule and yields its excitation energy to dissociate the molecule into the constituent atoms.

A collision in which one of the atoms is in an excited quantum state and as a result of an impact falls to a lower state without the emission of radiation is known as a *collision of the second kind*. Collisions of the second kind sometimes play an important part in certain types of excitation phenomena. Direct experimental evidence of this has been found.[12] These experiments show that when the 2,537-A radiation from a mercury discharge falls on the vapors of thallium or silver none of this radiation is absorbed. This is in accord with the discussion of Sec. 9-8. However, when a small amount of mercury vapor is mixed with the thallium or the silver vapors and the mixed vapors are irradiated with the 2,537-A photons, not only is radiation of 2,537 A emitted, but many lines in the thallium or silver spectra are also excited. This showed with certainty that the thallium or silver lines were excited only when excited Hg atoms were present in the mixed vapors. This excitation could occur only by collisions of the second kind. Of course, the excess energy of a collision (the difference between the 4.88 volts energy of the excited Hg atom and the energy of the emitted radiation of the thallium or silver atom) appears as kinetic energy after the collision.

Another illustration of collisions of the second kind, of some practical importance, is the *Penning effect*.[13] Penning found that neon, to which a little argon had been added, has a lower sparking potential than pure neon. The explanation is that neon has a metastable level at 16.6 ev, which is higher than the ionization level of 15.7 ev of argon. Hence a metastable neon atom may ionize an argon atom with which it collides.

As a check on the above explanation Penning illuminated the discharge with strong light from a glow discharge in neon. The emitted photons can be absorbed by the metastable neon atoms, which are thus removed from their metastable states. Hence ionizing collisions of the second kind with argon become less frequent, and it is found that the breakdown voltage is increased.

**9-13. Recombination.** It is desired to examine whether an electron and an ion may combine in the body of a discharge. In such a process an amount of energy equal to the ionization potential is released. Hence, by the principle of conservation of energy, the resulting neutral molecule must have an energy equal to the sum of the energies of the electron, the

ion, and the ionization potential. Moreover, since momentum must also be conserved, the momentum of the molecule after neutralization must equal the sum of the momentum of the electron and the ion. These two principles result in two equations for the one unknown (the velocity of the molecule), and a solution is not possible. In other words, in the body of the discharge simple electron and ion recombination does not take place. Clearly, recombination may occur if a third body is involved in the process.

Suppose that the third particle in the recombination process is a photon. The laws of conservation of energy and momentum can be satisfied, and such a process is theoretically possible. Actually recombination with the emission of radiation (apart from the regular spectral lines of the molecule in question) has been observed,[14] but the intensity of these lines is weak, indicating that this process is an improbable one.

The third body might be another electron, an ion, or a molecule that is present in the gas. However, the probability of such a triple encounter is not very great, except at very high pressures.

Recombination is possible when a positive ion and a negative ion (a neutral molecule to which an electron has become attached) interact with each other. The result of such an encounter is the formation of two neutral molecules.

Each of the foregoing processes has a low probability of occurrence, and it must be concluded that *very little deionization takes place in the body of the gas.*

Recombination takes place principally at the surface of the container. Thus, when the excitation is removed from a gaseous discharge, the ions diffuse to the walls or to any surfaces in the gas. They also drift to the electrodes if any field is present. The presence of the ion, the electron, and the surface permits satisfying both conservation of energy and momentum, with the consequent neutralization of the ion.

**9-14. Collision Probabilities.** Suppose that a beam of electrons is directed into a gas; the probability of collision $P_c$ * of an electron with a molecule is determined from Eqs. (8-15), (8-20), and (8-21), *viz.*,

$$n = n_0 \epsilon^{-p P_c x} \qquad (9\text{-}6)$$

where the pressure $p$ is in millimeters Hg and where $P_c$ is specified at 1 mm Hg and 0°C. Although this expression was deduced on the basis of the kinetic theory of gases, which views the molecule as a rigid sphere, nevertheless the derivation of this expression did not explicitly involve the molecular model. The equation was based merely on the definition of

---

* In the literature $P_c$ is given as the number of collisions *per centimeter* at 1 mm Hg and 0°C. Hence in this and the next section the unit of length is the centimeter instead of the meter.

$1/l_0 = P_c$ as the number of collisions made by an electron in traveling 1 cm through a gas that is at a pressure of 1 mm Hg and at 0°C. The value of $P_c$ may therefore be determined from this expression.

An experimental apparatus for this determination consists of an electron gun that projects a sharply defined beam of electrons of known energy into a gas at a pressure $p_1$ and known temperature. A collector at a distance $x$ away from the electron source is used to measure the current $i_1$ arising
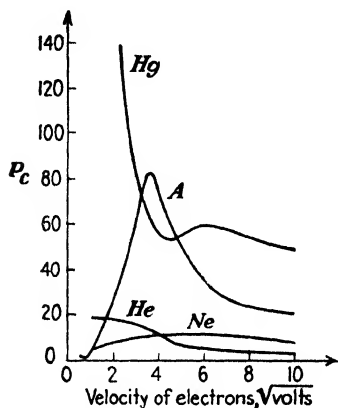


FIG. 9-4. The probability of collision of Hg, A, He, and Ne as a function of the velocity of the impinging electrons. (*R. B. Brode.*)
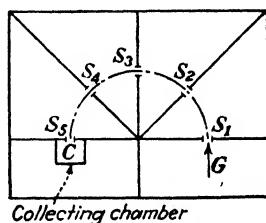
FIG. 9-5. Apparatus for measuring collision probabilities. A magnetic field perpendicular to the paper causes the electrons (which do *not* make collisions) to move in a circular arc from the electron gun *G* to the collector *C*.

from those electrons which have not made a collision in this distance. The gas pressure is then changed to $p_2$, and the new current $i_2$ to the collector is recorded. Then,

$$\frac{i_1}{i_2} = \frac{\epsilon^{-p_1 P_c x}}{\epsilon^{-p_2 P_c x}} = \epsilon^{(p_2 - p_1)P_c x} \tag{9-7}$$

from which $P_c$ can be calculated, since all the other quantities in the equation are known.

The experimental technique involved in making these measurements and the results of some of these experiments, some of which are illustrated in Fig. 9-4, can be found in the review of this subject by Brode.[15] The type of apparatus used by many investigators is shown schematically in Fig. 9-5. The electrons leaving the electron gun *G* are directed to travel in a circular path by means of a transverse magnetic field. The gas container is partitioned, as shown, so that only those electrons which pass through the slits $S_1, S_2, \ldots, S_5$, enter the collecting chamber *C*. If an electron has

made a collision with a gas atom, so that it is deflected through even a small angle, it will strike one of the baffles and will not be collected. Furthermore, if the electron has made an inelastic collision, it will suffer a loss of energy and the radius of its resulting path, for the given magnetic-field intensity, will be smaller (in accordance with the results of Sec. 2-11). Because of the altered radius of curvature, these electrons will likewise strike the baffles. Therefore any collision, elastic or inelastic, removes the colliding electron from the incident beam. The distance $x$ in Eq. (9-7) for such an apparatus is the total length of the curved path (shown as the broken semicircle in Fig. 9-5). The experimentally obtained results for the probability of collision in Hg, A, He, and Ne are shown in Fig. 9-4.

It is observed that $P_c$ depends very markedly upon the velocity with which the electrons are traveling. In view of the foregoing discussion, this was to be expected. A collision can no longer be visualized as the impact of a "point" electron with a solid spherical atom. Instead the term "collision" now refers to that type of encounter in which a change of direction and energy takes place. For charged particles, the deflection occurs because the impinging electron enters the electric field of the nucleus and the surrounding cloud of electrons. If the scattered electron emerges with practically no loss of energy, an elastic collision has taken place. However, if the initial energy of the electron is greater than the critical potentials of the gas, then excitation or even ionization may take place. The total collision probability $P_c$, as measured, includes both elastic and inelastic encounters, since the apparatus shown in Fig. 9-5 does not discriminate between the different collision processes.

The shapes of the curves of $P_c$ vs. energy of incident electron vary markedly from element to element. However, those atoms in the same column of the periodic table have been found to exhibit similar collision probabilities. The inert gases (Ne, A, Kr, Xe, He) offer very little hindrance to very slowly moving electrons, and $P_c$ increases to a maximum and falls off as the energy of the electrons is increased. This behavior is known as the *Ramsauer effect*. The alkalies (Na, K, Rb, Cs) follow curves that are somewhat similar to those of the inert gases, except that the maximum is extremely sharp and occurs at a much smaller value of energy (about 2 volts). Also, the probability of collision for the alkalies is roughly ten times as great as those illustrated in Fig. 9-4. The elements Zn and Cd have curves similar to that of Hg. They exhibit no maximum but decrease monotonically with velocity. Unlike most other elements, the probability of collision for slowly moving electrons in mercury is very high. The theoretical explanation of these curves can be given in terms of wave mechanics.[16]

It was shown in Sec. 8-6 that the probability of collision $P_c$, on the assumption that the molecules were rigid spheres, was equal to $\sigma N$, the

total projected area of all of the atoms in 1 cm³ of the gas, which is maintained at a pressure of 1 mm Hg and at 0°C. It is found convenient to retain this relationship, even though the kinetic-theory concept of a gas is not valid. Thus, by *definition*, the collision cross section $\sigma$ of an atom is such that $P_c \equiv \sigma N$. As a consequence of this definition, the curves of Fig. 9-4 are frequently referred to as the "collision cross-section curves." They yield direct information regarding the variation of mean free path $l_0$ with electronic speed, since $l_0 = 1/P_c$.

**9-15. Ionization Probabilities.**[15] A question of the utmost importance in discharge devices is the following: How many of the collisions represented by $P_c$ cause ionization? The experimentally determined answer is shown in Fig. 9-6 for those elements for which $P_c$ has been given in Fig. 9-4. The ordinates represent the probability of ionization $P_i$, which is defined as the number of ionizing collisions made per centimeter of path through a gas at 1 mm Hg pressure and 0°C. This quantity is also known as the *cross section for ionization.*

It is seen that the probability of ionization is zero for potentials below the ionization energy of the gas, as is to be expected. The curves then rise approximately linearly, reach maxima, and then fall slowly. This means that if an electron has an energy that is much larger than the



Fig. 9-6. The probability of ionization by electron bombardment of Hg, A, Ne, and He as a function of the energy of the impinging electrons. (*R. B. Brode.*)

ionization energy the chance of its making an ionizing collision is very small.

It should be further noted that the value of $P_c$ for a given electron energy is larger than the corresponding value of $P_i$. This is to be expected, since $P_c$ includes exciting and elastic collisions as well as the ionizing collisions.

**PROBLEMS**

**9-1.** Show that the time for one revolution of the electron in the hydrogen atom in a circular path about the nucleus is

$$T = \frac{m^{\frac{1}{2}}e^2}{4\sqrt{2}\epsilon_0(-W)^{\frac{3}{2}}} \quad \text{sec}$$

where the symbols are defined in Sec. 9-1.

**9-2.** Bohr postulated that the stationary states are determined by the condition that *the angular momentum must be an integral multiple of $h/2\pi$*, where $h$ is Planck's constant. For the hydrogen atom show that (in mks units)

*a.* The possible radii are given by

$$r = \frac{h^2 \epsilon_0 n^2}{\pi m e^2} \quad \text{meters}$$

where $n$ is any integer but not zero. For the ground state ($n = 1$) show that the radius is 0.53 A.

*b.* The energy levels are given by

$$W_n = -\frac{m e^4}{8 h^2 \epsilon_0{}^2} \frac{1}{n^2} \quad \text{joules}$$

*c.* The reciprocal of the wave length (called the *wave number*) of the spectral lines is given by

$$\frac{1}{\lambda} = R \left( \frac{1}{n_2{}^2} - \frac{1}{n_1{}^2} \right) \quad \text{waves/m}$$

where $n_1$ and $n_2$ are integers with $n_1$ greater than $n_2$ and $R = m e^4/8\epsilon_0{}^2 h^3 c = 1.10 \times 10^7/m$ is called the *Rydberg constant*.

If $n_2 = 1$, this formula gives a series of lines in the ultraviolet called the *Lyman series*. If $n_2 = 2$, the formula gives a series of lines in the visible, called the *Balmer series*. Similarly the series for $n_2 = 3$ is called the *Paschen series*. These predicted lines are observed in the hydrogen spectrum.

**9-3.** Show that Eq. (9-5) follows from Eq. (9-4).

**9-4.** A photon of wave length 1,400 A is absorbed by cold mercury vapor, and two other photons are emitted. If one of these is the 1,850-A line, what is the wave length of the second photon?

**9-5.** Cold mercury vapor is bombarded with radiation from a mercury arc, and as a result the fluorescent lines 2,537 A and 4,078 A appear. What wave length must have been present in the bombarding radiation?

**9-6.** The six lowest energy levels of hydrogen are 0, 10.19, 12.07, 12.73, 13.04, and 13.20 ev. If cold hydrogen vapor absorbs the ultraviolet 972-A line, what possible fluorescent lines may appear?

**9-7.** The seven lowest energy levels of sodium vapor are 0, 2.10, 3.19, 3.60, 3.75, 4.10, and 4.26 ev. A photon of wave length 3,300 A is absorbed by an atom of the vapor, and three other photons are emitted.

*a.* If one of these is the 11,380-A line, what are the wave lengths of the other two photons?

*b.* Between what energy states do the transitions take place in order to produce these lines?

**9-8.** What might happen if cold mercury vapor is bombarded with

*a.* One 5.00-ev photon?

*b.* One 5.00-ev electron?

*c.* One 5.46-ev photon?

**9-9.** *a.* With what speed must an electron be traveling in a sodium-vapor lamp in order to excite the yellow line whose wave length is 5,893 A?

*b.* Could electrons with this speed excite the 2,537-A line of Hg?

**9-10.** *a.* What is the minimum speed with which an electron must be traveling in order that a collision between it and an unexcited neon atom may result in ionization of this atom?

*b.* What is the minimum frequency that a photon can have and still be able to cause photoionization of a neon atom?

**9-11.** An X-ray tube is essentially a high-voltage diode. The electrons from the hot filament are accelerated by the plate supply voltage so that they fall upon the anode

with considerable energy. They are thus able to effect transitions among the tightly bound electrons of the atoms in the solid of which the target (the anode) is constructed.

a. What is the minimum voltage that must be applied across the tube in order to produce X rays having a wave length of 0.5 A?

b. What is the minimum wave length in the spectrum of an X-ray tube across which is maintained 60 kv?

**9-12.** An electron possessing a kinetic energy of 18 ev is moving in Hg vapor. If it collides with an atom that is in its lowest metastable state, causing the atom to become ionized, can the electron induce ionization in a subsequent collision?

**9-13.** An electron, after falling through a potential of 10 volts, collides with a mercury atom that is in its lowest metastable state. As a result of the impact the atom is elevated to its 7.73-volt level. What is the energy in joules of the impinging electron after the collision? Assume that the kinetic energy of the atom is unaffected by the collision.

**9-14.** Argon resonance radiation falls upon sodium vapor. If a photon ionizes an unexcited sodium atom, with what speed is the electron ejected?

**9-15.** A metastable neon atom possessing 16.6 volts energy "collides" with an unexcited argon atom and ionizes this atom. If the atoms are at rest before and after the impact, calculate the energy with which the electron is emitted.

**9-16.** Using the kinetic-theory values of the radii of Hg, A, and Ne (Table 8-1), calculate the probability of collision for these atoms. Compare these values with the experimentally determined probabilities of Fig. 9-4.

**9-17.** A low-voltage gas-filled cathode-ray tube contains argon gas at room temperature. The cathode-to-anode voltage is 100 volts, and the distance from the anode to the screen is 20 cm. What must the pressure of the argon gas be if 10 per cent of the electrons in the beam are to make collisions in their passage from the anode to the screen?

**9-18.** The pressure of the argon in a cathode-ray tube is 0.01 mm Hg. The length of the tube is 20 cm. The beam current is 50 $\mu$a. The accelerating voltage is 400 volts. The tube operates at room temperature.

a. Find the total number of ions formed per second over the entire length of the tube.

b. How many electrons per second do *not* make ionizing collisions on their way to the screen?

**REFERENCES**

1. RUTHERFORD, E., *Phil. Mag.*, **21**, 669, 1911.
2. BOHR, N., *ibid.*, **26**, 1, 1913.
3. For a discussion of the various experimental methods that have been used to determine the mean life of excited states, see DARROW, K. K., "Electrical Phenomena in Gases," pp. 125-132, The Williams & Wilkins Company, Baltimore, 1932.
4. BLEAKNEY, W., *Phys. Rev.*, **34**, 157, 1929; **35**, 139, 1930.
5. FRANCK, J., and G. HERTZ, *Deut. physik. Gesell. verh.*, **16**, 12, 1914.
6. HARNWELL, G. P., and J. J. LIVINGOOD, "Experimental Atomic Physics," p. 319, McGraw-Hill Book Company, Inc., New York, 1933.
7. DAVIS, B., and F. S. GOUCHER, *Phys. Rev.*, **10**, 101, 1917.
8. WOOD, R. W., "Physical Optics," The Macmillan Company, New York, 1934.
9. ZEMANSKY, M. W., *Phys. Rev.*, **29**, 513, 1927.
10. VARNEY, R. N., L. B. LOEB, and W. R. HASELTINE, *Phil. Mag.*, **29**, 379, 1940.
11. KIRCHSTEIN, B., *Z. Physik*, **60**, 184, 1930.
12. CARIO, G., *ibid.*, **10**, 185, 1922.
    CARIO, G., and J. FRANCK, *ibid.*, **17**, 202, 1923.
13. DARROW, K. K., "Electrical Phenomena in Gases," pp. 123, 297, The Williams & Wilkins Company, Baltimore, 1932.

14. COMPTON, K. T., and I. LANGMUIR, *Revs. Modern Phys.*, **2**, 123, 1930.
15. BRODE, R. B., *ibid.*, **5**, 257, 1933.
16. ALLIS, W. P., and P. M. MORSE, *Z. Physik*, **70**, 567, 1931.


### General References

For a treatment of critical potentials, see COMPTON, K. T., and F. L. MOHLER: *Natl. Res. Council Bull.* 48, 1924.

DARROW, K. K.: "Electrical Phenomena in Gases," The Williams & Wilkins Company, Baltimore, 1932.

DOW, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

ENGEL, A. V., and M. STEENBECK: "Elektrische Gasenladungen," Vol. 1, Verlag Julius Springer, Berlin, 1932.

FRANCK, J., and P. JORDON: "Anregung von Quantensprüngen durch Stösse," *Handbuch der Physik (Geiger Scheel)*, **23**, 641, 1926.

HULL, A. W.: *Elec. Eng.*, **53**, 1435, 1934.

LOEB, L. B.: "Atomic Structure," John Wiley & Sons, Inc., New York, 1938.

LOEB, L. B.: "Fundamental Processes of Electrical Discharge in Gases," John Wiley & Sons, Inc., New York, 1939.

MITCHELL, A. C. G., and M. W. ZEMANSKY: "Resonance Radiation and Excited Atoms," Cambridge University Press, London, 1934.

MOTT, N. F., and H. S. W. MASSEY: "The Theory of Atomic Collisions," Oxford University Press, New York, 1933.

RAMSAUER, C., and R. KOLLATH: "Der Wirkungsquerschnitt von Gasmolekülen gegenüber langsamen Elektronen und langsamen Ionen," *Handbuch der Physik (Geiger Scheel)*, **22**(2), 243, 1933

RICHTMYER, F. K., and E. H. KENNARD: "Introduction to Modern Physics," McGraw-Hill Book Company, Inc., New York, 1947.

RUARK, A. R., and H. C. UREY: "Atoms, Molecules, and Quanta," McGraw-Hill Book Company, Inc., New York, 1930.

SEMAT, H., "Introduction to Atomic Physics," Rinehart & Company, Inc., New York, 1946.

SLEPIAN, J.: "Conduction of Electricity in Gases," Westinghouse Electric & Manufacturing Company, Educational Department, Course 38, 1933.

TONKS, L.: *Elec. Eng.*, **53**, 239, 1934.

WHITE, H. E.: "Introduction to Atomic Spectra," McGraw-Hill Book Company, Inc., New York, 1934.

# CHAPTER 10

# ELECTRICAL DISCHARGES IN GASES

IN ORDER to study any of the individual processes that can occur in a gaseous-discharge tube, an experimental arrangement is usually employed in which only the process under survey is important. This permits the particular characteristics of interest to be investigated without the presence of complicating secondary phenomena. Under ordinary circumstances, the discharges usually encountered in practice (for example, arc and glow discharges) involve the simultaneous existence of several fundamental processes. Therefore, it is not always possible to give a complete explanation of the observed phenomena. However, the characteristics of the discharges to be considered in this chapter will be discussed wherever possible in the light of the fundamental processes already considered in Chap. 9.

**10-1. Electrical Character of Discharges.** The most convenient way of distinguishing among the various possible electrical discharges in gases is to study the volt-ampere characteristics of these discharges. Consider a tube containing two cold plane-parallel electrodes between which is contained a fixed quantity of gas at a low pressure (say, a few millimeters Hg). An adjustable source of potential in series with a resistor is connected across the tube. The current through the tube is observed as a function of the voltage across it, as either the magnitude of the voltage or the magnitude of the resistor is varied.

FIG. 10-1. Volt-ampere characteristics of a non-self-maintained discharge.

The general form of the characteristic so obtained is given in Fig. 10-1.

The current is found to vary gradually with variations in potential from the point $O$ to the point $A$. Further increases in potential in the region from $A$ to $B$ results in no further increase in current through the tube, and the horizontal portion of the curve from $A$ to $B$ is obtained. As the voltage is increased beyond the value corresponding to $B$, the current is found to increase rapidly to the point $C$, somewhat as shown. For reasons which will appear in the next section, this discharge is called *a non-self-maintained, a field-intensified, or a Townsend discharge*. If an attempt is made to ob-

257

tain the characteristic beyond point $C$, the entire character of the discharge changes suddenly. The voltage is found to decrease rapidly. The gas in the tube suddenly begins to glow, the color of the luminous region being a function of the gas or gases contained in the tube. Also, the current through the tube rises very rapidly, the magnitude of the current that flows through the tube being determined by the magnitude of the potential and by the size of the current-limiting resistor in the external circuit. The gas in the tube is said to have *broken down*, and the voltage at $C$ is the *sparking potential*. The discharge becomes a glow or an arc. The theory of the Townsend discharge will now be given. Conditions after breakdown are studied in later sections.

**10-2. Non-self-maintained Discharge.** A gas is always in a state of partial ionization because of the action of a number of natural ionizing sources, as explained in Sec. 9-10. If, therefore, a voltage is applied to two electrodes sealed in a tube containing a gas, a few ions will drift to the cathode and some electrons will be transported to the anode. This will result in the flow of a minute current. If it is assumed, as is generally true, that the rate of production of the ions by the ionizing agent is a constant, then the current is independent of voltage over a range of voltages. This is so because the ions and free electrons are swept out of the field immediately upon formation. This is the region $AB$ of Fig. 10-1. The external radiating agents produce saturation current densities of the order of $10^{-22}$ amp/m$^2$.

In order to explain the rapidly rising portion of the characteristic, the region $BC$ of the figure, suppose that the potential $E_i$ at the point $B$ represents the ionization potential of the gas. Then in the region beyond $B$ the potential through which the electron falls from the time of its liberation by the external ionizing agent until it collides with a gas molecule in its path a short distance away may be sufficient to give it an energy adequate to cause ionization of the molecule. That is, the potential per mean free path of the electron is sufficient to cause ionization. If it ionizes the molecule, it will liberate, say, one electron. Now two electrons exist to cause further ionization. It is then evident that this process, which is cumulative, may result in very large currents in the tube. An analytical representation of this process is possible.

Suppose that $n_0$ electrons are ejected from the cathode per second (see Fig. 10-2). If $n$ denotes the number of electrons per second at any distance $x$ from the cathode, then $dn$ new electrons per second will be formed in the distance $dx$ owing to the ionization that occurs. If $\alpha$ *represents the number of new electrons (or ions) formed by one electron in traveling a distance of* 1 m *through the gas, then*

$$dn = \alpha n \, dx$$

The parameter $\alpha$ is frequently called the *first Townsend coefficient*. This leads, upon integration, to

$$n = n_0 \epsilon^{\alpha x} \qquad (10\text{-}1)$$

*provided that $\alpha$ is independent of $x$.* Thus, for every electron produced at the cathode, $\epsilon^{\alpha d}$ electrons reach the anode a distance $d$ away. The progeny of $\epsilon^{\alpha d}$ electrons produced by one electron is called an *electron avalanche*.

By multiplying both sides of Eq. (10-1) by the electronic charge, the expression is transformed into an equation for the current. It is

$$I_b = I_0 \epsilon^{\alpha d} \qquad (10\text{-}2)$$

where $I_b$ is the current to the collector and $I_0$ represents the current at the cathode resulting from the effects of the external ionizing agents. If $I_0$
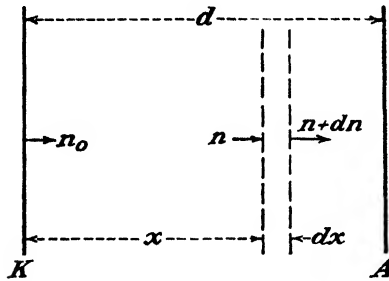


FIG. 10-2. Because of ionization, $dn$ new electrons are formed in the distance $dx$.
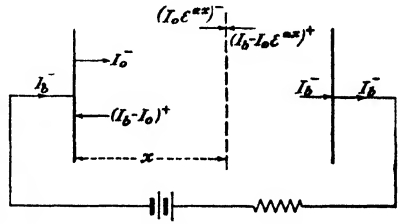
FIG. 10-3. The current in a discharge tube consists of electrons traveling toward the anode and positive ions traveling toward the cathode.

drops to zero, so also does $I_b$. In other words, if the external excitation is removed the discharge stops. This is why the discharge is termed *non-self-maintained*.

In a discharge the current is made up of electrons traveling toward the anode and positive ions moving toward the cathode. The *net* current at any plane must be a constant because otherwise there would be a continuous piling up (or perhaps a depletion) of charge with time, which is not possible if a steady-state condition has been reached. In Fig. 10-3 a positive superscript is used for positive ions and a negative superscript for electrons. The arrows indicate the direction of travel of the particles. The current *everywhere* is $I_b$. Note, in particular, that this current at the anode is due entirely to electrons which have entered the collector from the discharge. These electrons flow in the external circuit to the junction of the cathode and the gas. Here a certain number of them per second neutralize the positive-ion current $(I_b - I_0)$. The remainder, $I_0$, is the electron current entering the discharge under the influence of the external excitation.

**10-3. The Parameters $\alpha$ and $\eta$.** Because of the small values of $I_0$ resulting from the natural ionizing agents, it is difficult to obtain accurate measurements of the currents in the region $BC$ of Fig. 10-1. It is possible, however, to investigate this theory and that of the next section by using a tube that is provided with a photosensitive cathode, as the current in such a device (of the order of microamperes) may readily be measured. Except for the fact that the initial electrons are produced photoelectrically instead of by natural ionizing agents, the conditions of gas amplification are the same. The volt-ampere curve of an argon-filled phototube (General Electric PJ 23) is given in Fig. 15-9. Its shape is exactly that of the curve of Fig. 10-1.

The number of electrons formed per electron per meter of path (the value of $\alpha$) will depend both upon the number of collisions that the incident electron makes in traversing this unit path and also upon the energy that the incident electron possesses when it collides with the gas molecules. But the number of collisions made by an electron per meter of its path, which is the inverse of the electronic mean free path, will be directly proportional to the pressure of the gas. Also, the energy that an impinging electron possesses at the time of collision with a molecule will depend both upon the electric-field intensity and upon the mean free path. That is, the energy per mean free path is given by the product $\mathcal{E}l$. Since the mean free path varies inversely with the pressure, then $\mathcal{E}l$ may be written proportional to $\mathcal{E}/p$. Thus $\alpha$ is proportional to $p$ and also to some function of $\mathcal{E}/p$. Mathematically, these results may be expressed in the form

$$\frac{\alpha}{p} = f\left(\frac{\mathcal{E}}{p}\right) \tag{10-3}$$

where $f(\mathcal{E}/p)$ is some undetermined function which remains constant when the ratio $\mathcal{E}/p$ is constant. Consequently, $\alpha$ will be independent of $x$ if $\mathcal{E}$ and $p$ do not depend upon $x$.

In order to obtain data for the investigation of these matters, it is necessary that the experiments be performed with a tube that is provided with plane-parallel electrodes the spacing between which may be varied.[1] The pressure of the gas $p$ is maintained constant, and the applied potential is increased in proportion to $d$ so that the electric-field intensity $\mathcal{E}$ remains constant. (The current density is kept sufficiently low so that the space charge is negligible.) Under these conditions $\alpha$ is independent of $x$, and Eq. (10-2) should be valid. A plot of $\log I_b$ vs. $d$ from these experiments is found to be linear, this equation being thus verified. The slope of this straight line gives $\alpha$ for the fixed value of $p$ and $\mathcal{E}$. The pressure or the electric-field intensity is then varied, and a new value of $\alpha$ is obtained from the slope of the new logarithmic plot. In this manner, $\alpha$ is determined as a function of $p$ and $\mathcal{E}$. The data are expressed by plotting $\alpha/p$ as a function

of $\mathcal{E}/p$.   The results so obtained lie on a smooth curve as illustrated in Fig. 10-4.*   This verifies the functional relationship (10-3).

A theoretical calculation of the shape of the function $f(\mathcal{E}/p)$ is extremely difficult since it depends upon such factors as the energy distribution of the electrons and the probability of ionization, neither of which is accurately known.   Furthermore, these terms cannot be predicted with any
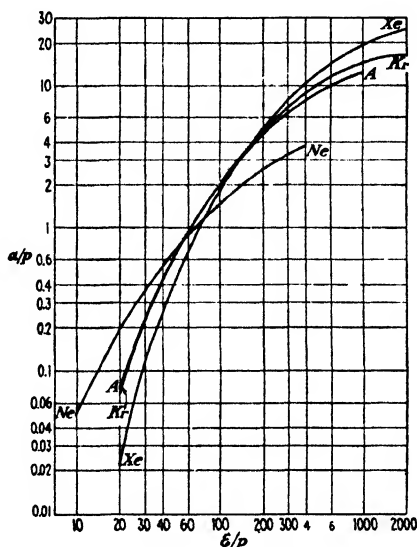
Fig. 10-4.   Values of $\alpha/p$, 1/cm $\times$ mm for neon, argon, krypton, and xenon as a function of $\mathcal{E}/p$, volts/cm $\times$ mm.   (*After A. A. Kruithof, Physica*, **7**, 519, 1940.)

certainty.[2,3]   An approximation which can be made to fit the data over a limited range by the proper choice of the parameters $A$ and $B$ for each gas is [4]

$$\frac{\alpha}{p} = A\,\epsilon^{-Bp/\mathcal{E}} \qquad (10\text{-}4)$$

The *ionization coefficient* $\eta$ *is defined as the number of ions formed by one electron in falling through one volt.*   Since $\alpha$ is the number of ions formed by an electron in traveling a distance of 1 m and $\mathcal{E}$ is the electric-field intensity in volts per meter, then, clearly,

$$\eta = \frac{\alpha}{\mathcal{E}} \qquad (10\text{-}5)$$

* In the literature values of $\alpha$ and $\mathcal{E}$ are given in terms of the centimeter as the unit of length and the millimeter of Hg as the unit of pressure.   This practice has been followed in Figs. 10-4, 10-5, 10-7, and 10-8.

and from Eq. (10-3)

$$\eta = \frac{p}{\varepsilon} f\left(\frac{\varepsilon}{p}\right) \tag{10-6}$$

so that $\eta$ is a function of $\varepsilon/p$.

The current is given by Eq. (10-2), or

$$I_b = I_0 \epsilon^{\eta \varepsilon d} = I_0 \epsilon^{\eta E} \tag{10-7}$$

where $E = \varepsilon l$ is the applied voltage.

This equation is based on the assumption that $I_b = I_0$ at $d = 0$ or $E = 0$. Actually the current remains equal to $I_0$ until a voltage $E_0$ ap-
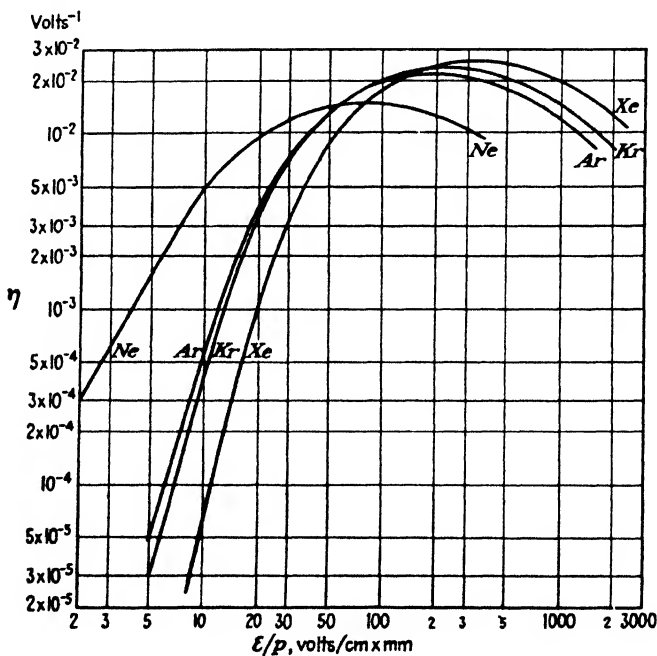


Fig. 10-5. The ionization coefficient $\eta = \alpha/\varepsilon$ for neon, argon, krypton, and xenon. (*After A. A. Kruithof, Physica, 7, 519, 1940.*)

proximately equal to the ionization potential is reached (point $B$ of Fig. 10-1). Hence Eq. (10-7) should more properly be modified to

$$I_b = I_0 \epsilon^{\eta(E - E_0)} \tag{10-8}$$

where $E_0$ is a constant approximately equal to the ionization potential.

Curves of $\eta$ as a function of $\varepsilon/p$ are given in Fig. 10-5, and they exhibit a maximum. Thus at a given $\varepsilon$ there should be an optimum pressure for a gas photocell (corresponding to the peak in Fig. 10-5). This is Stoletow's

law and is verified by experiment.[5]  For an argon-filled tube operating at an electric-field intensity of the order of 100 volts/cm, the maximum response is very flat and occurs at pressures between approximately 1 and 0.2 mm Hg.  This is the correct order of magnitude for the pressure in a gas photocell.

**10-4. Breakdown.**  If the potential across the phototube is made too high, the gas in the tube will "break down" and will begin to glow.  The current in the tube will rise to a high value and will be limited principally by the resistance in the external circuit.  If, following breakdown, the light source is removed so that the photocathode is no longer illuminated, the current nevertheless continues.  Such discharges are called *self-sustained* or *self-maintained discharges*.  Although external agencies may assist in starting the discharge, once it has been initiated it will maintain itself without any external source.

For breakdown to take place there must be some mechanism of electron production in the tube which has not yet been mentioned.  The most likely process is that of electron emission at



Fig. 10-6.  When positive ions strike the cathode, they liberate electrons, $I_s$ (in addition to those liberated from the cathode by the external source, $I_e$).

the cathode due to the positive-ion bombardment (often called *secondary emission* by positive ions).  This electron current is designated by $I_s$, the electron current due to the external excitation by $I_e$, and the total electron current at the cathode by $I_0$.  Thus

$$I_0 = I_e + I_s \qquad (10\text{-}9)$$

If $I_b$ is the plate current, then by the principle outlined in the preceding section, the total current anywhere in the tube is $I_b$.  Since the total electron current at the cathode is $I_0$, then the ion current at this electrode must be $I_b - I_0$.  The situation is as pictured in Fig. 10-6, which is identical with Fig. 10-3 except that the total electron current $I_0$ at the cathode now consists of two terms.  Hence, Eq. (10-2) is still valid provided only that $I_0$ is given by Eq. (10-9).

*If on the average $\gamma$ electrons are liberated from the cathode by one positive ion,* then

$$I_s = \gamma(I_b - I_0) \qquad (10\text{-}10)$$

By combining Eqs. (10-2), (10-9), and (10-10) there results

$$I_b = \frac{I_e \epsilon^{\alpha d}}{1 + \gamma - \gamma \epsilon^{\alpha d}} \qquad (10\text{-}11)$$

Since the production of electrons by positive-ion bombardment of the cathode is a rather inefficient process, then $\gamma$ is a small quantity. Representative values [6] range from 0.20 to 0.001, meaning that on the average somewhere between 5 and 1,000 positive ions must strike the cathode before a single electron is liberated. Figure 10-7 shows curves of $\gamma$ as a function of $\mathcal{E}/p$. The shape and average height of the curves are strongly de-
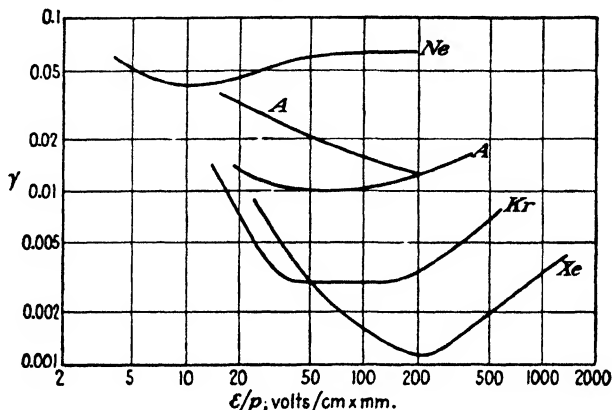


FIG. 10-7. The coefficient $\gamma$ for neon, argon, krypton, and xenon, with a copper cathode. The two curves for $A$ correspond to two different states of the cathode. (*After A. A. Kruithof, Physica*, **7**, 519, 1940.)

pendent upon the cathode material and the state of the cathode. Note that $\gamma$ is a much less sensitive function of $\mathcal{E}/p$ than is $\eta$.

Equation (10-11) predicts a more rapid increase of current with electrode separation than the simple exponential expression (10-2). This results from the additional electrons that are liberated from the cathode by the positive-ion bombardment. A plot of $\log I_b$ vs. $d$ is found to be a straight line for small values of $d$ but rises more rapidly for large values of $d$. The values of $\alpha$ and $\gamma$ are chosen to give the best agreement between the experimental results and those calculated from Eq. (10-11).[7] An approximate value of $\alpha$ is obtained from the straight-line portion of the plot, as outlined in the last section. It is to be noted that since the number of secondary electrons emitted from the cathode depends upon the energy of the impinging ions, then $\gamma$ should be a function of the energy gained by the ion in its last free path before striking the cathode. Hence, $\gamma$ should be a function of $\mathcal{E}l_m$, or $\mathcal{E}/p$, and should not depend upon $\mathcal{E}$ or $p$ individually. This conclusion is verified by experiment.

Mathematically, it follows that the current will increase without limit when the denominator of Eq. (10-11) becomes zero. That is, breakdown occurs when

$$\gamma \epsilon^{\alpha d} = 1 + \gamma \qquad (10\text{-}12)$$

Physically, of course, the current can never increase without limit; but rather, as the left-hand term of this expression approaches $1 + \gamma$, an enormous increase in the current will occur, and the transition from the non-self-maintained to the self-maintained discharge takes place.

The existence of a self-maintained discharge requires that the number of ions which are produced by an electron in moving from the cathode to the anode must regenerate one electron when they strike the cathode. This criterion is actually expressed by Eq. (10-12). To verify this, it is noted that if $n_0$ electrons per second leave the cathode, then these produce $(n_0 \epsilon^{ad} - n_0)$ positive ions per second in their passage to the anode. Therefore, when one electron leaves the cathode, it produces $(\epsilon^{ad} - 1)$ positive ions in the body of the gas. These positive ions eject, in turn, by the $\gamma$ process at the cathode, $\gamma(\epsilon^{ad} - 1)$ electrons. Evidently, if this quantity just equals unity, then the one electron originally emitted from the cathode has caused another electron to leave the cathode. This new electron will likewise cause the production of a third electron, and so on indefinitely. In other words, the discharge will no longer require an external agent for its maintenance. The criterion for this condition is

$$\gamma(\epsilon^{ad} - 1) = 1$$

which is seen to be equivalent to Eq. (10-12).

In this theory, which is due to J. J. Thomson, it has been assumed that the additional electrons were produced by the action of the positive-ion bombardment of the cathode.[8] However, many other types of electron-producing collision processes are possible, as discussed in the preceding chapter. Thus one may expect that a certain amount of ionization will result from collisions between the ions moving toward the cathode and the neutral gas molecules with which they may collide. However, experiment shows that this process is so inefficient as to be completely negligible. Electrons may arise from the presence of metastable atoms in the region of the cathode resulting from ionization by impact between several of these atoms. There may be photoemission from the cathode due to radiation from excited gas atoms. Further, photoionization of the gas molecules may produce some electrons. For a detailed discussion of the theories of electron multiplication before breakdown, see Ref. 8.

A method of separating $\gamma$ into its component processes [9] is based upon the fact that these individual mechanisms require different times. Thus with reasonable values of electric field, spacing, and pressure the formative times are approximately as follows: Electrons released by ionization reach the anode in $10^{-8}$ sec. Photoemission takes $10^{-8}$ sec. Ions drifting under the influence of the applied field reach the cathode in $10^{-6}$ sec, whereas metastable atoms, which must arrive at the cathode by diffusion, take $10^{-3}$ sec. The transient current resulting from the liberation of a very short (0.2 $\mu$sec) pulse of photoelectrons at the cathode is observed experi-

mentally. This technique should yield very valuable information with respect to the fundamental processes involved. At the time of this writing (1950), this microsecond Townsend discharge is being actively investigated at the Bell Telephone Laboratories.

**10-5. Paschen's Law.** Many terms are used synonymously with the term "breakdown voltage." Some of them are "sparking," "ignition," "starting," or "striking" potential. The factors which determine the sparking potential will now be investigated.

If the coefficient $\eta$ is used in place of $\alpha$ and if the constant $E_0$ is introduced as in Eq. (10-8), then Eq. (10-11) takes the form

$$I_b = \frac{I_0 \epsilon^{\eta(E - E_0)}}{1 + \gamma - \gamma \epsilon^{\eta(E - E_0)}} \tag{10-13}$$

The condition for sparking is obtained, as before, by setting the denominator of the above equation equal to zero when $E$ equals the sparking voltage $E_s$. Thus

$$\gamma \epsilon^{\eta(E_s - E_0)} = 1 + \gamma \tag{10-14}$$

Since $\gamma$ and $\eta$ are functions of $\mathcal{E}/p$, this equation predicts that $E_s$ is a function of the ratio $\mathcal{E}/p$. Furthermore, since

$$\frac{\mathcal{E}}{p} = \frac{\mathcal{E}d}{pd} = \frac{E_s}{pd}$$

then Eq. (10-14) contains only two parameters, namely, $E_s$ and the product $pd$. Thus, one is led to the conclusion that the breakdown voltage is a function only of the *product* of the pressure and the interelectrode spacing and does not depend upon these two parameters separately. This conclusion is verified by experiment. Thus, if the pressure is doubled but the interelectrode spacing is halved, the breakdown voltage remains unchanged. This is known as *Paschen's law.* Expressed mathematically, Paschen's law states that

$$E_s = \varphi(pd) \tag{10-15}$$

Clearly, from the above discussion, *the sparking voltage also depends upon the cathode material and the type of gas.* These conclusions are verified by experiment as can be seen from Fig. 10-8. In curves $a$ the effect of the cathode material is in evidence because three different types of cathodes are used with the same gas. Also, the fact that $E_s$ is a function only of $pd$ and not $p$ or $d$ separately is demonstrated because curves for the same gas and the same cathode coincide (roughly) even if different distances are used. In curves $b$ the effect of different gases with the same cathode is demonstrated. Curve $c$ shows breakdown in air for a very wide range of $pd$.

For electrodes of a given area, the volume of gas contained between them is proportional to the separation $d$. Also, since the concentration is proportional to the pressure, then the product $pd$ is proportional to the number

of molecules between the electrodes.   Hence, Paschen's law states that the ignition voltage depends only upon the total number of molecules of gas between the cathode and the anode, for a given cathode material and a given gas.



| Tube No. | Gap spacing (cm) | Surface element |
|---|---|---|
| 1 | 0.094 | Ba |
| 2 | 0.086 | Ba |
| 3 | 0.210 | Ba |
| 4 | 0.130 | Mg |
| 5 | 0.279 | Mg |
| 6 | 0.178 | Mg |
| 7 | 0.114 | Al |
| 8 | 0.091 | Al |
| 9 | 0.206 | Al |
| 10 | 0.110 | Al |
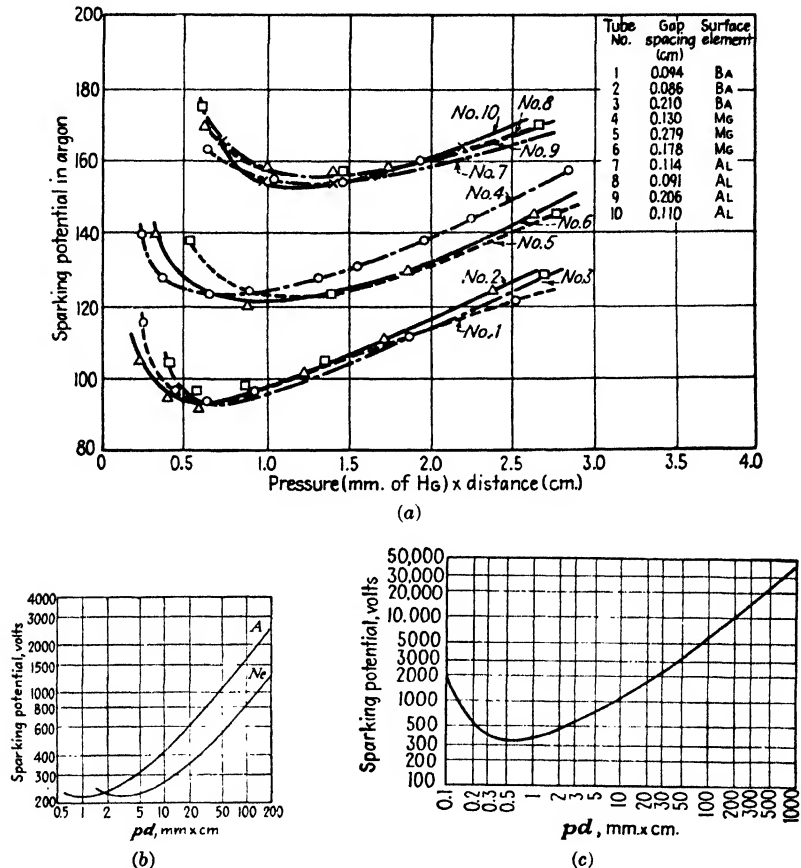
Fig. 10-8.   Sparking potential between plane-parallel electrodes as a function of the product of pressure times distance.   (a) Barium, magnesium, and aluminum in argon (with distance as a parameter).   (After H. Jacobs and A. P. LaRocque, J. Applied Phys., **18**, 199, 1947.)   (b) Argon and neon with a copper cathode.   (After A. A. Kruithof, Physica, **7**, 519, 1940.)   (c) Air with a brass cathode.   (After M. J. Dryvesteyn and F. M. Penning, Revs. Modern Phys., **12**, 87, 1940.)

The experimentally determined curves of sparking potential vs. the product $pd$ show a minimum, the voltage being high for very low and very high values of $pd$.   Refer to Fig. 10-8c, which is for plane-parallel electrodes in air.   The minimum value of the product $pd$ is 0.6 mm × cm.

Note that regardless of the pressure or spacing, it is impossible to cause breakdown to occur between parallel electrodes in air at voltages less than 350 volts.

The reason for the existence of a minimum in the sparking-potential curve is easy to discover. Consider, for example, that the spacing is fixed and that $p$ is varied. At very low pressures, there are so few molecules present that enough secondary electrons can be produced only if the energies of the impinging electrons are high. This means that a high voltage must be applied. On the other hand, if the pressure is high, then the number of collisions is so large that the energy gained by each electron per mean free path is small unless the applied potential is high. For ionization to take place, the energy per mean free path must exceed a certain minimum amount (the ionization potential of the gas), and so a high potential will be necessary. Between these two extremes of pressure will be a pressure at which a minimum sparking potential will be necessary.

Although the foregoing discussion has been confined to the case of plane-parallel electrodes, the discussion is generally true for systems with electrodes of any shape. This is so because the energy of the electrons per mean free path determines the ionizing abilities of the electrons and the ions and therefore the breakdown potential. A generalization of Paschen's law is based upon this fact. Thus, if all the linear dimensions of a discharge system are multiplied by a factor $a$ and if the pressure is multiplied by the factor $1/a$, the breakdown voltage remains unchanged. This follows, of course, from the fact that the energy per mean free path remains unchanged. Also, since the mean free path $l$ varies inversely as the pressure, the breakdown voltage is a function only of the ratio $\mathcal{E}/p$. In fact, the quantities $\alpha$, $\gamma$, and $E_s$ are usually plotted as functions of $\mathcal{E}/p$ in the literature.

**10-6. Breakdown by Streamer Propagation.**[11] * Loeb feels that the Townsend theory discussed in Sec. 10-4 is inadequate at the higher pressures and longer gap lengths, for many reasons, four of which follow:

1. On the basis of the transit time of the positive ions in such fields, and for the usual gap lengths, the time lag in sparking should be $10^{-5}$ sec or longer, whereas it is found that the formative time lag of sparks at atmospheric pressure is of the order of $10^{-7}$ sec or less at moderate overvoltages (1 to 2 per cent).

2. At high values of $pd$ the cathode seems to play no important role in the breakdown mechanism since the same sparking voltage is obtained with different cathodes.

3. There are gaseous discharges (like the lightning stroke or the positive-point corona) where the cathode is not involved at all.

* The authors are indebted to Professor L. B. Loeb for reading and correcting this section and, in particular, for adding the last two paragraphs. It should be noted that this streamer theory is not generally accepted. See, for example, Llewellyn-Jones, F. and A. B. Parker *Nature*, **165**, 960, 1950.

4. It is observed that at high pressures sparks occur along very narrow filamentary channels which may be branched and irregular in shape.

The Townsend mechanism cannot account for such sparks. A qualitative mechanism of sparking, known as the *streamer theory*, which is in accord with experiment was arrived at independently by Raether and by
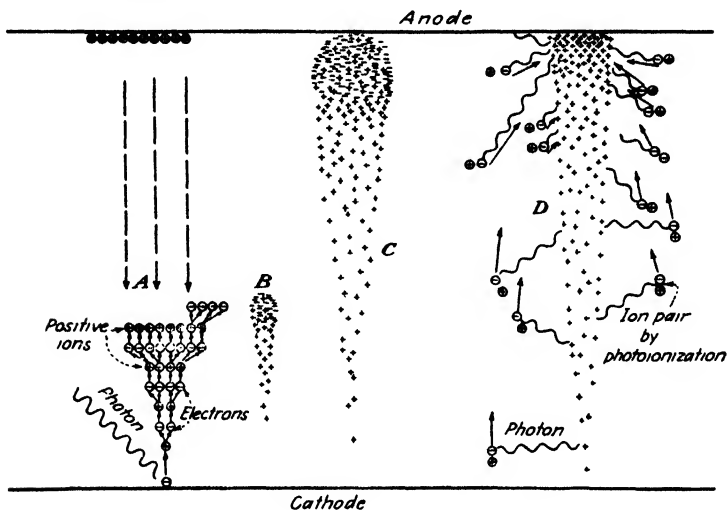


FIG. 10-9. A pictorial representation of the advance of an electron avalanche in a gas discharge. (*After L. B. Loeb and J. M. Meek, "The Mechanism of the Electric Spark," Stanford University Press, Stanford University, Calif., 1941.*) (*A*) Electron multiplication. (*B*) Development and structure of the avalanche. Positive ions lag behind electrons at tip. (*C*) Avalanche has crossed the gap. (*D*) Later history of the avalanche. Electrons have disappeared into the anode.

Loeb. A semiempirical quantitative criterion based on this theory was given by Meek for the positive streamer spark.

To understand this streamer theory, consider a plane-parallel gap at atmospheric pressure, 1 cm in length, between which is applied the sparking voltage. Suppose that the cathode is illuminated by ultraviolet light of such an intensity that one electron per microsecond per square centimeter leaves the cathode. As before, in traversing a distance $x$, each electron produces an electron avalanche of $\epsilon^{\alpha x}$ electrons. The $\epsilon^{\alpha x}$ positive ions that are produced during the avalanche formation remain virtually stationary during the $10^{-7}$ sec required for the avalanche to advance across the gap. Owing to the random diffuse movement of the electrons during the process of ionization by electron bombardment, the tip of the avalanche spreads as the avalanche advances. The general features of the process are

shown schematically in Fig. 10-9. In traversing the 1-cm path under the conditions specified, approximately, $2.4 \times 10^7$ positive ions are produced. However, such a distribution of ions does not make a conducting filament of charges across the gap, and the avalanche that has crossed the gap does not in itself constitute a breakdown.

In addition to the positive ions that are formed during the electron avalanche, from four to ten times this number of excited atoms and molecules also result. These excited atoms emit ultraviolet radiation, the emission time being approximately $10^{-8}$ sec after excitation. The ultraviolet radiation is absorbed in the gas and leads to additional photoionization.
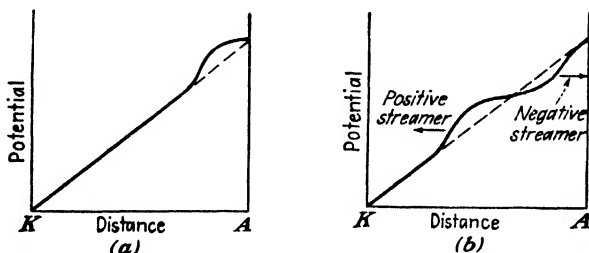


FIG. 10-10. (a) The potential distribution due to a streamer at the anode. The positive ions cause the potential to rise more rapidly than the space-charge-free straight line (shown broken). (b) The mid-gap streamer distribution.

These new photoelectrons, which arise almost instantaneously in the whole gap, will begin to ionize cumulatively, producing other avalanches. However, these are usually smaller and occur later than the parent avalanche. Moreover, those at remote distances from the primary avalanche ordinarily do not contribute to the breakdown.

Those photoelectron avalanches that are produced near the positive-ion channel and particularly near the anode are in a combined applied and space-charge field of the positive ions. This space-charge field is quite large, owing to the large space-charge density, and (slightly above the threshold) is comparable with the applied sparking field. This field distorts the applied field along the avalanche axis as shown in Fig. 10-10a. As a result, these photoelectrons are drawn into the positive space-charge region, which now becomes a conducting region which starts at the anode. The positive ions that are left behind by these electrons will extend the space charge toward the cathode. These new electrons also create photons, which produce more electrons, to continue this process. In this way, the positive space charge develops toward the cathode from the anode as a self-propagating positive space-charge *streamer*. Since this propagating positive space charge depends on the photoionization in the gas, with very much enhanced fields, the velocity of propagation of this space-charge streamer is larger than the

avalanche and increases as the streamer progresses. The general aspects of streamer formation are illustrated in Fig. 10-11.

The streamer produces a filamentary region of intense space-charge distortion, a steep gradient existing at the cathode end of the streamer tip. As this advances toward the cathode, the photoelectron avalanches produced by the radiation at the cathode begin to produce intense ionization here. The positive ions created there bombard the cathode and increase
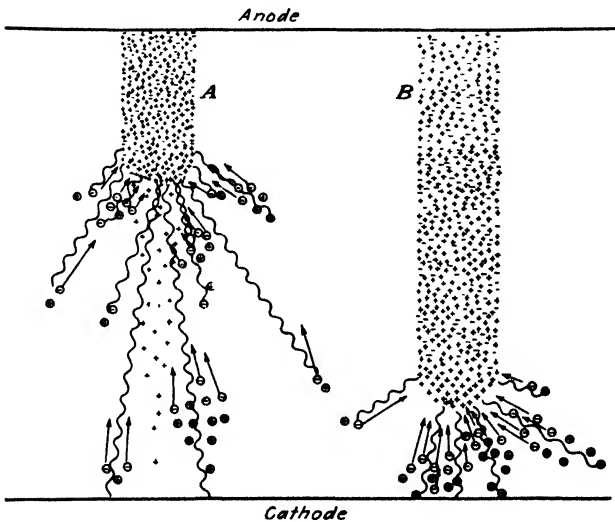


FIG. 10-11. The pictorial representation of streamer formation in a gas discharge. (*After L. B. Loeb and J. M. Meek, "The Mechanism of the Electric Spark," Stanford University Press, Stanford University, Calif., 1941.*) (*A*) Channel of plasma which is advancing toward cathode. (*B*) The streamer approaches cathode.

the electron emission. Thus as the space-charge streamer approaches the cathode, a cathode spot is forming. When the streamer reaches the cathode, a conducting filament bridges the gap, and because of the high field distribution on junction, there is a rush of electrons toward the anode.

Now, owing to the high-voltage wave front that passes along the ionized conducting channel to the anode, additional intense ionization of the order of $10^{18}$ to $10^{20}$ electrons per cubic meter in the streamer channel occurs, and the channel is rendered highly conducting. It is this intense increase in ionization by the potential wave that gives the highly conducting channel (with as many as $\frac{1}{3}$ the molecules ionized) the characteristic of the spark. The velocity of propagation of this returning wave of ionization along the ionized channel is high, and reaches $10^8$ m/sec in the return stroke in a lightning discharge. This process produces the brilliant and noisy phase of the spark.

It should be noted that only about 1 in $10^8$ avalanches leads to streamer formation with subsequent sparking at the threshold. Overvoltage increases this markedly and also reduces the formative time lag. These formative time lags amount to from $0.5 \times 10^{-8}$ to $10^{-7}$ sec for the time for the avalanche to cross a 1-cm gap, and a smaller but roughly comparable time for the crossing of the anode streamer. Of course, if the applied voltage is less than the breakdown voltage, the avalanche will leave the cathode and cross the gap. However, no streamers will form, and the gap will not break down.

According to Meek's equation, the breakdown voltage should be a function of the distance as well as the product $pd$. However, this deviation from Paschen's law is not great. For example, for $pd = 760$ mm $\times$ cm, a variation in sparking voltage of less than 5 per cent is predicted by this theory for variations in $d$ from 0.1 to 10 cm.

If the pressure of the gas is reduced, the corresponding density of photoionization that is produced will become too small for the streamer mechanism to operate. Under these circumstances, the streamer mechanism is replaced by the more efficient diffuse Townsend mechanism. This transition point for air has not been clearly established but it is less than 80 mm $\times$ cm. Thus, over most of Fig. 10-8, the breakdown is due to the $\gamma$ process, rather than the streamer mechanism.

The foregoing describes the streamer mechanism of spark breakdown, and this original picture following Loeb and Meek permits of a simple introduction. Very recent experiments on spark breakdown by Fisher and Bederson and by B. Gaenger indicate that the very short formative time lags of $10^{-7}$ sec which led to the streamer theory occur only at overvoltages of the order of 2 per cent and above. Carrying time-lag measurements down to of the order of 0.1 per cent overvoltage, the spark lag continuously increased to 100 $\mu$sec or more. Increasing overvoltage decreased the time lags continuously. This extended from pressures of the order of 80 mm and a 1-cm gap on up. The uniformity of the process and the filamentary character of the spark leave no doubt about these being streamer breakdowns and that these extend well below the $pd = 200$ mm $\times$ cm originally postulated. The tentative explanation is that the electrical-image force field of the avalanche just crossing the gap at streamer-forming magnitudes has its field too weak to give the cathode-directed streamer. In fact, Raether never observed a cathode-directed streamer starting right at the anode. Thus, the streamer process must come from avalanches so dense that the streamers start nearer midgap. This requires some overvoltage. At the threshold it is believed by Loeb, and Fisher and Bederson, that a low-grade diffuse self-sustaining Townsend discharge functioning primarily by photoelectric effect at the cathode sets in. Such discharges under the field strengths and pressures normally considered in air can

build up a space-charge distortion of positive ions that will lead to breakdown by means of streamer processes occurring in midgap.

The streamer breakdown with midgap streamers according to Raether's pictures consists of two streamers, the cathode-directed positive streamer of Loeb and Meek and a simultaneous anode-directed electron cloud streamer as indicated in Fig. 10-10$b$. For streamers starting fairly close but not at the anode, the electron-cloud negative anode-directed streamer starts first. For streamers starting from nearer midgap, the cathode-directed positive streamer starts first. For positive points the streamer can start right at, or very close to, the surface because of the high fields. Negative points require higher potentials and produce anode-directed streamers, which explain the negative pilot streamer in most lightning discharges. Raether developed the quantitative theory for the negative anode-directed streamer, and Meek did the same for the positive cathode-streamer. The theories are similar but both are inaccurate and neither includes the factors for photoionization needed in a complete theory. Townsend type self-sustaining discharges thus precede breakdown at threshold for streamers, and when pressures are low and fields high, constitute the diffuse glowlike sparks. These have time lags ranging from milliseconds or more to tens of microseconds.

**10-7. Corona and Brush Discharges.**[12] Suppose that the two electrodes to which the potential is applied are not parallel plates and are well separated in space. Under these circumstances, the field at these electrodes may become sufficiently high for breakdown to occur in the neighborhood of the electrodes long before a spark will pass between the electrodes across the gap space. The localized discharge is accompanied by the emission of some light, by audible noises of a "sizzling" character, and by the propagation of some energy which appears as radio interference.

If the discharge occurs at atmospheric pressure, the luminosity is limited in extent. This is known as a "corona" discharge and is explained by the streamer theory. At lower pressures the luminosity extends farther into the gap and is called a "brush" discharge.

**10-8. Glow Discharge.** Now return to the consideration of a low-pressure discharge *after* breakdown. This discharge, which is known as the *glow discharge*, is visually characterized by brightly but differently colored luminous regions in the gas. Electrically, it is characterized both by a *low current density* and by a *high voltage drop*. This maintaining voltage is considerably higher than the ionizing potential of the gas present, but lower than the sparking potential. The volt-ampere characteristic of this portion of the discharge is given in Fig. 10-12. The region $FK$ has been separated into two portions $FH$ and $HK$ in order to call attention to a condition which is apparent by visual inspection of the tube while under operation. This condition is the following: Only a small part of the cath-

ode is covered with glow for low values of total current. In fact, as the current through the tube increases, the portion of the cathode surface that is covered with glow increases linearly with the current. This indicates that the current density remains constant. This constant value is known as the *normal current density*. Normal current densities are of the order of 1 amp/m$^2$ at a pressure of 1 mm Hg. The region *FH*, which is characterized by a fairly high though more or less constant potential, is known as the *normal glow* discharge.
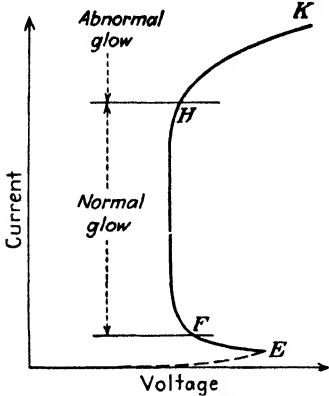


Fig. 10-12.   Volt-ampere characteristic of a glow discharge.

The variation of potential from cathode to anode has the general form shown in Fig. 10-13. The region adjacent to the anode and extending over the major portion of the tube is called the *plasma*, because it contains approximately equal concentrations of positive and negative charge. The formation of the plasma is discussed in detail in Sec. 10-15. It is sufficient for the present to point out that since the net charge density is zero, the potential must vary linearly with distance (according to Poisson's equation). Actually the variation is often found to be very small,
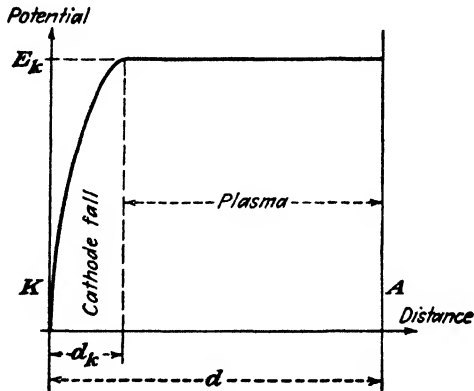


Fig. 10-13.   The essential potential variation in a glow.

and in what follows the plasma will be taken as an equipotential region at the anode voltage. Almost all of the voltage across the tube appears in the small distance between the cathode and the boundary of the plasma, called the *cathode-fall* region.

**10-9. Conditions at the Cathode.** The cathode fall, which represents the potential through which the positive ions "fall" in passing to the cathode, is the most important part of the discharge. It is found experimentally that the magnitude of this voltage is of the order of (but somewhat less than) the *minimum* value of the breakdown voltage for the given gas and the given cathode material.[13] There appears to be no theoretical explanation for this coincidence. Another important experimentally determined feature is that the cathode fall is substantially independent of the pressure. This means that the distance $d_k$ covered by the cathode fall adjusts itself so as to yield the correct value of $pd_k$ at which the minimum breakdown voltage occurs. That is to say, Paschen's law applies here, with $d$ replaced by $d_k$. This is verified by observing that as the pressure is varied the cathode-fall potential remains constant, and $d_k$ varies inversely as the pressure. Also, the criterion for the establishment of a self-maintained discharge [Eq. (10-12)], with $d_k$ replacing $d$, has been successfully applied to the cathode fall of a glow discharge.

Normal values for cathode-fall voltages range between about 59 volts (a potassium surface and helium gas) and 350 volts. The presence of a low-work-function coating on the cathode tends to give a low cathode fall with any gas.[14] Also, the use of one of the inert gases (He, Ne, A, etc.) results in a low cathode fall with any cathode material. Of course, the presence of small amounts of impurities in either the gas or the cathode will result in a considerably modified value of cathode fall. It is the characteristics of a substantially constant cathode fall over a rather wide range in currents that account for the use of glow-discharge tubes as ballast tubes.

It has been noted that the discharge is maintained by electron emission by positive-ion bombardment at the cathode. Since it takes a large number of positive ions to produce one electron by this $\gamma$ process, the cathode fall must be a region containing many positive ions. If the small electron current at the cathode is neglected, then a positive-ion current only exists in this region. The conditions in this region are then precisely analogous to those which prevail in a vacuum tube under the conditions of a negative space charge resulting from an excess of electrons. Consequently, Eq. (7-15), which relates the current density (in this case a positive-ion current density) with the potential, must be valid. Thus

$$ J = K \frac{E_k{}^{\frac{3}{2}}}{d_k{}^2} \qquad (10\text{-}16) $$

where $E_k$ is the cathode-fall potential and $d_k$ is the distance from the cathode to the edge of the negative glow.

This equation contains an explanation of the "normal" region of the glow discharge. Suppose, for example, that the current through a glow

tube is measured for a given applied potential and external series resistor. According to the foregoing discussion $E_k$ is approximately the minimum breakdown voltage, and $d_k$ is the distance at which this minimum occurs for the given pressure. This fixes both $E_k$ and $d_k$ in Eq. (10-16), and thus $J$ may be calculated. Consequently, the area of the glow $I/J$ is known. However, as the magnitude of the external resistor is varied, the magnitude of the current is changed and $E_k$, $d_k$, and so $J$ remain constant, whence the area of the glow on the cathode increases directly with the magnitude of the current. This accounts for the "normal" discharge region and for the normal current density that exists in such discharges.

If $d_k$ varies inversely as $p$, then it follows from Eq. (10-16) that *J varies directly as the square of the pressure.* This has been verified experimentally.

Once the entire cathode is covered with glow, then further increases in current can be obtained only by a corresponding increase in the current density. This requires an increase in $E_k$ and a decrease in $d_k$. This region is called the "abnormal" glow and is the region $HK$ of Fig. 10-12.

It is not possible to maintain a glow discharge simply by applying the breakdown voltage between the electrodes, if the spacing is too small for a given pressure. Thus, if sparking occurs to the left of the minimum of the Paschen's curves (Fig. 10-8), then the breakdown voltage will be equal to the cathode-fall voltage, since this fall will occupy the entire tube. In order to be able to draw any current from the tube, a voltage in excess of the breakdown potential must be applied.

No adequate quantitative theory of the glow discharge has been given, and the above description certainly is not intended as such. However, the discussion is useful because it shows in a semiquantitative way approximately what takes place in the cathodic part of the discharge.

**10-10. Details of the Glow Discharge.** The results of a very careful examination of the discharge both visually and electrically are illustrated in the diagrams of Fig. 10-14. These diagrams show the variations of light intensity, potential, field strength, space-charge densities, and current density as a function of the distance between the electrodes. The various portions of the glow discharge will be considered in some detail. The reader can best follow the discussion by making constant reference to Fig. 10-14.

Very close to the cathode, there is a narrow dark region which is known as the *Aston dark space*. The velvety glow adjacent to this, which is often brightly but differently colored, is known as the *cathode glow*. (Both these regions are often missing in a discharge.) The *Crookes dark space*, which emits very little or no light (it is faintly blue in air), extends outward for some distance from the cathode glow. This region is also known as the *cathode*, or *Hittorf, dark space*. Adjacent to the Crookes dark space is a luminous region known as the *negative glow*, which starts quite abruptly

and gradually fades into the region known as the *Faraday dark space*. The Faraday dark space merges into the luminous *positive column*, or *plasma*, which is pink in air. This terminates in the *anode glow*, which is separated from the anode by a narrow *anode dark space*.
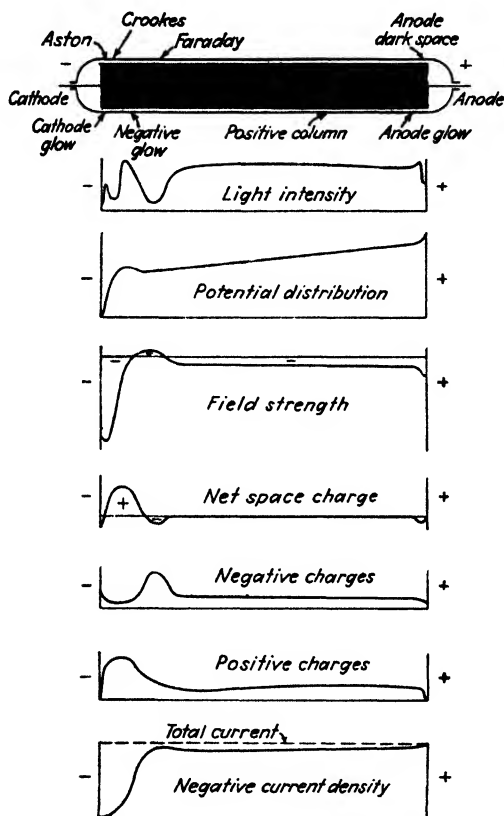


Fig. 10-14. Electrical and visual characteristics of a glow discharge. (*Reprinted by permission from "Fundamental Processes of Electrical Discharges in Gases" by L. B. Loeb, published by John Wiley & Sons, Inc.*)

*The Cathode Region.* Since the glow is a self-maintained discharge, it is necessary that the electrons regenerate themselves. This is done through the process of inducing sufficient cumulative ionization so that the ions, in addition to being neutralized at the cathode as a result of the bombardment, liberate enough other electrons to maintain the discharge. These electrons are emitted with essentially zero initial velocities. This accounts for the small negative charge density right at the cathode. Since these

electrons do not possess enough energy to cause excitation, this will be a dark region. This is the Aston dark space.

The mechanism of the cathode glow is not clearly understood. It has been suggested that after passing through the Aston dark space the electrons emitted from the cathode gain enough energy to cause excitation and so give rise to the cathode glow. This theory is offset, however, by the fact that not enough electrons leave the cathode to give rise to as much light as is observed. This is so because the current at the cathode is almost entirely positive-ion current, the positive ions traveling through the cathode-fall potential toward the cathode. Furthermore, the spectrum of the emitted light does not correspond to the light expected from the first excitation potential of the gas. It is possible that the cathode glow results largely from the recombination of the positive ions at the surface of the cathode.

The electrons constituting the electron current at the cathode resulting from the positive-ion bombardment leave the cathode with very small velocities but are accelerated very rapidly through the cathode-fall space. Although the number of these electrons at the cathode is small, the number of electrons increases rapidly with the distance from the cathode, since at each ionizing collision a new electron is generated. The electric-field intensity is very high at distances close to the cathode, and so the new electrons quickly gain energy from the field. However, because of their large energy, the probability of excitation and ionization (see Fig. 9-6) will be rather small. Consequently, there is very little light emitted from the region of the cathode-fall space. This follows from the fact that there are few electrons in the region near the cathode, where the collision probability is high; and in the rest of the cathode fall, where the electron density increases, the probability is low.

So many electrons are being produced at the edge of the cathode-fall space that the net charge density drops to zero and may even become negative. As a result, the field is in such a direction as to slow the electrons down. Furthermore, the electrons formed near the anode side of the cathode fall travel but a short distance before reaching the negative glow and so do not gain much energy from the field. Thus a large number of relatively slow-moving electrons exist in this region, and so a great deal of excitation will take place here. Because of this, the intense negative glow on the cathode side of the discharge starts rather abruptly. The spectrum of this light contains many blue lines, which represent high-energy transitions. This is to be expected since some of the colliding electrons producing excitation and ionization have acquired large energies in the cathode-fall region.

*The Plasma Region.* The field in the negative glow is in such a direction as to slow down even the fastest electrons, and they thereby lose their

excitation and ionization abilities.    Thus, in passing through the negative-glow region, the intensity of the luminous region decreases.    This corresponds to the beginning of the Faraday dark space, which does not have particularly marked boundaries.    In this region the charge is negative, as illustrated, and consists of electrons that are drawn from the negative glow.    Since only a few positive ions exist in this region, the negative charge tends to increase the field between the dark space and the positive anode.    This field increases until the electrons that exist in this field are given sufficient energy to produce ionization by collision, the luminosity resulting therefrom marking the beginning of the *positive column*, or *plasma*

The field in the positive column, or plasma, remains substantially con-stant owing to the existence of equal concentrations of electrons and posi tive ions, so that ionization by collision will occur along its whole length. and a uniform region of light results.    If an accumulation of positive ions should occur at any point in the plasma, the variation of the field may be sufficient to cause the discharge to be dark at that point.    As in the case of the Faraday dark space, this loss of ionizing power will result in the existence of an accumulation of negative charge in the dark space with a consequent increase in the field.    Thus a series of bright and dark stria tions may exist in the positive column, a condition that frequently arises. It is found that striations exist mainly in mixed or impure gases.

*The Anode Region.*    That the anode region plays a relatively insignificant part in the maintenance of a glow discharge is made evident by the follow-ing experiment: At a given pressure, a discharge is established in a tube that is provided with a movable anode.    As the anode is brought closer to the cathode, it is found that no changes occur at the cathode, although the voltage across the entire tube decreases by the small voltage difference across the plasma.    This continues until the anode passes through the plasma, the Faraday dark space, and the negative glow.    Once the anode reaches the region of the cathode dark space, the voltage across the tube rises.    The reason for this is clear.    Once the anode is in the cathode dark space, the tube will be operating to the left of the minimum point on the $E_s$ vs. $pd$ curve of Fig. 10-8, and the potential across the tube must rise in order to maintain the discharge.

The anode fall of potential may be positive, zero, or negative.    In any case, it seldom exceeds several volts.    If the current demanded by the circuit is larger than the random electron current at the surface of the anode, the anode fall must be positive, as indicated in Fig. 10-14.    Positive ions in the immediate neighborhood of the anode will be repelled under these conditions, and there will be an electron boundary, or *sheath*, separating the plasma from the anode.    If, on the other hand, the random electron current is greater than the anode current, the anode fall must be negative and a concentration of positive charge exists outside the anode.

That the conditions are as here discussed has been experimentally verified by Pupp.[15] His investigations showed that the anode fall can be changed from positive to negative by controlling the positive-ion concentration in the neighborhood of the anode. This was accomplished by placing an auxiliary artificial ion source near the anode.

**10-11. Cathode Sputtering.** The bombardment of the cathode surface by the positive ions as they are neutralized by the electrons liberated from the cathode is often intense enough to cause the material of the surface to wear away gradually. Whether this sputtering occurs as a result of evaporation or by a destructive process that actually tears small portions of the cathode material away is unknown. The result of this positive-ion bombardment is to wear away the electrodes and thus limit their life. Since this sputtered metal forms a deposit on the glass walls, it tends both to decrease the luminosity of the tube owing to the opacity of the film and to increase the chance for electrical leakage along the inner surface of the glass envelope.

There is a removal of molecules from the gas because of ion penetration of the cathode. This is known as gas *cleanup*. To a smaller extent the sputtering of the cathode material during its deposition on the glass walls also takes with it some of the gas of the tube. Cleanup will tend to alter the active pressure of the gas and so may affect the operation of the discharge.



**10-12. Arcs.[2,16]** Upon increasing the current through a discharge tube containing a gas beyond the region where the normal current density prevails, the abnormal glow sets in. This region is characterized by an increasing current density and an increasing tube drop. Eventually, a sudden transition takes place in which the voltage drop decreases markedly, the current density increases until the current concentrates itself in a small spot on the cathode, and the discharge becomes an arc. The current through the system is controlled wholly by the external circuit.

Fig. 10-15. Volt-ampere characteristics of a gaseous discharge. (*After M. J. Druyvesteyn and F. M. Penning, Revs. Modern Phys.*, **12**, 87, 1940.)

Figure 10-15 shows the volt-ampere characteristic starting with a non-self-sustained Townsend discharge and progressing through the glow to the arc. The tremendous current range covered should be noted (a logarithmic current scale is used).

Electric arcs are generally associated with the flow of large currents at low voltage or with very high current densities at the cathode, and with a volt-ampere characteristic that has a negative slope. The mechanism of initiating the arc need not be that of a transition from a glow by increasing the voltage. Another common method of starting an arc is to separate two metals which are in contact through which a current flows. Examples are the opening of a switch, the starting of a carbon arc, or the ignition of a simple mercury-arc rectifier. A third method is to send a current pulse through a semiconductor in contact with the cathode. This is the mechanism used in the ignitron.

To the engineer, arc discharges are the most important type of gaseous conductors. This is because arcs convey currents of hundreds or thousands of amperes at volt drops from about 10 to 30 volts.

The dividing line between an arc and a glow discharge is indistinct, and they have many features in common. Each discharge has associated with it the cathode fall, the plasma, and the anode fall. The discharges differ in respect to the mechanism by which the electrons are supplied from the cathode. In the glow discharge, the electrons are emitted from the cathode principally by positive-ion bombardment of the cathode. In the arc discharge, the emission of the electrons from the cathode occurs through the operation of a supplementary mechanism other than by positive-ion bombardment. The production of electrons is due to a hot cathode or a high field at the cathode surface. This leads to the following classification of arcs:

1. *Thermionic Arcs.* The cathode is heated to high temperatures *by the discharge*, the thermionic-emission current being of the same order as the self-maintained arc current.

2. *Externally Heated Arcs.* These are non-self-maintained arcs in which an externally heated thermionic cathode supplies the requisite arc current.

3. *Low-boiling-point Cathode Arcs.* The electron release occurs by the mechanism of high-field emission from a relatively cold cathode surface. The thermionic emission is negligible in this case.

Common examples of arcs of type 1 are the carbon-arc lamp and the "sun lamp" with tungsten electrodes. Arcs of types 2 and 3 are the most prevalent in engineering practice. The externally heated arcs (type 2) find wide use in gas-filled thermionic diodes and triodes. These units have moderate current capacities (up to perhaps 25 amp). The high-field emission arcs (type 3) are extremely important. The mercury-pool-cathode arcs used in high-current tank rectifiers presumably belong in this class. Arcs of type 3 also play a significant role in circuit breakers.

The temperature at the surface of liquid mercury at the cathode spot (type 3 arc) is difficult to measure directly,[17] for this temperature is determined not only by the heat conduction but also by the convection of the

liquid mercury. The temperature may be estimated from spectroscopic data. In addition to the Hg lines emitted from the cathode spot, a continuous spectrum is also emitted. This continuous spectrum has been attributed to the high temperature of the liquid.[18] The temperature would be about 2000°C in that case. It seems more probable, however, that this continuous spectrum is emitted by the highly ionized mercury vapor in the cathode spot and should not be attributed to radiation from the liquid surface. The temperature of the liquid surface, based upon the rate of evaporation from the cathode spot and upon the current density, is found to lie between 220° and 300°C. It seems probable, therefore, that the temperature of the cathode spot of a mercury arc is so low that thermionic emission from the cathode is entirely negligible.

Langmuir [18] suggested that the electronic emission in the mercury arc results from the presence of a very high electric field at the surface of the cathode, the mechanism of the emission being the usual autoelectronic effect. The necessary high fields will arise in this case from the presence of the cathode fall at the small distance over which this fall exists. The thickness of the cathode-fall space is often less than the mean free path of the mercury ions. It should be pointed out, however, that this suggested theory has not been generally accepted by all research workers in the field.[19]

The arcs of type 3 include those which are maintained between electrodes that would normally vaporize at the temperatures necessary for thermionic emission. It would also include those arcs which are formed from glows in times that are far too short for the cathode to heat up and liberate electrons thermionically.[20] The most probable mechanism of these arcs is also high-field emission. The necessary high field is perhaps obtained because of the surface irregularities caused by vaporization of the low-melting-point cathode. However, this theory is still speculative.

**10-13. Static Volt-ampere Arc Curves.** The general form of the volt-ampere characteristic of a self-maintained arc is given in Fig. 10-16. The volt-ampere characteristic is a falling one and may be expressed, in general, by a relation of the form

$$E_b = A + B(I_b)^{-n} \qquad (10\text{-}17)$$
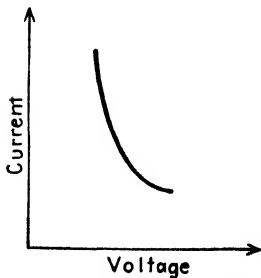


Fig. 10-16. Volt-ampere characteristic of an arc.

where $A$ and $B$ are functions of the arc length and $n$ is a quantity proportional to the boiling temperature of the anode material. This expression was empirically established by Nottingham [21] and applies for values of current between 1 and 10 amp. For values of current greater than 10 amp, the second term in the expression is small, so that the volt drop across the arc is practically a constant and independent of the current.

It is interesting that the boiling temperature of the anode material rather than the cathode materials sets a limit to the possible temperature of the discharge. Nevertheless this equation contains no explanation of the essential phenomena of the arc itself.

For an externally heated cathode the voltage is almost constant with current.

**10-14. A Comparison of the Arc and Glow.** A summary of the essential similarities and differences of glow and arc discharges follows.

*Similarities.* (1) Both have the same shape volt-ampere characteristic. The breakdown voltage is greater than the maintaining voltage, and the latter is essentially constant with load current. (2) Both have the same shape voltage-distance characteristic. It consists of a cathode fall, a plasma, and an anode fall.

*Differences.* (1) A glow is a low-current high-voltage discharge, whereas an arc is a high-current low-voltage discharge. (2) In the glow the electrons are supplied by the $\gamma$ mechanism at the cold cathode. In the arc there is available a more efficient electron supply, such as thermionic or high-field emission. (3) The arc has no counterpart to the normal glow region of the glow tube.

**10-15. The Formation of the Plasma.**[22] Consider a tube containing gas between parallel electrodes separated by a distance $d$. If there were no space charge the potential vs. distance curve would be the straight line 1 shown in Fig. 10-17. Assume now that the electrons liberated from the cathode (by any of the mechanisms discussed in connection with either glows or arcs) ionize some molecules. Since the electrons are much lighter than the ions, they are rapidly drawn out of the discharge, whereas the positive ions remain essentially stationary where they are formed. Hence, a positive space charge is built up. By Coulomb's law this raises the potential at every point in the interelectrode space, and curve 2 now represents the potential distribution. These curves and the discussion to fol-



Fig. 10-17. Potential-distance curves as positive space charge is formed.

low should be compared with the somewhat analogous situation in Sec. 7-1, where the potential-distance curves for a negative-space-charge diode are developed.

If the positive space charge increases further, it may become large enough for the distribution to take on the shape of curve 3, which has a maximum greater than the applied plate potential $E_b$. However, in general, such a situation is not physically possible. To demonstrate this,

refer to Fig. 10-18, where are shown the potential-energy curves for an electron corresponding to the potential curves of Fig. 10-17. (It should be recalled that potential energy equals potential times the charge on the electron—which is a negative number, so that the curves of Fig. 10-18 have the same shape as those of Fig. 10-17 but are inverted.) The potential maximum corresponds to a potential-energy minimum and hence to a potential-energy barrier for slow-moving electrons. For example, consider an electron that starts from the cathode with very little energy, travels a distance $OD$, collides with a gas molecule, and gives up an amount
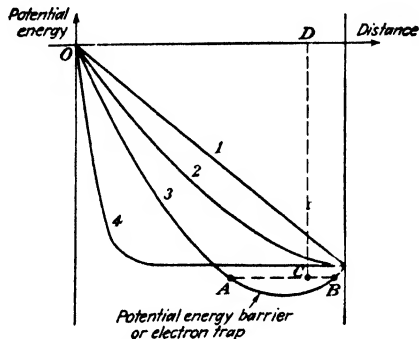


Fig. 10-18. Potential-energy–distance curves for an electron corresponding to the potentials in Fig. 10-17.

of energy $DC$. This electron will not be able to reach the anode but will collide with the barrier at point $B$. It will be "trapped" in the region between $A$ and $B$. Similarly, if the collision causes ionization, the liberated electron will have very little energy and will also be trapped in the potential maximum. Hence, these electrons will effectively neutralize positive ions. In this manner the captured electrons reduce the net space charge to zero, and the potential maximum disappears. The equilibrium distribution, curve 4 of Fig. 10-17 (and Fig. 10-18), is one in which most of the space is an approximately equipotential volume containing equal concentrations of positive and negative charge—the plasma. The necessary condition for the generation of a plasma is that the rate of production of positive ions is sufficient to produce a potential maximum.

In the case of the glow tube, it is necessary that the applied voltage be fairly high in order that the condition for a self-maintained discharge be satisfied. For an arc discharge, however, an efficient mechanism of electron production at the cathode is available, and hence the potential needs to be only large enough to produce sufficient ionization to form the plasma. For a hot-cathode mercury-vapor tube this is usually about 10 to 15 volts. However, the drop across such a tube may be as low as 5 or 6 volts. The

necessary ionization energy under these circumstances is produced with the aid of metastable states within the atom, the ionization occurring in two stages.

It is found that an arc can, in many cases, be maintained at a voltage which is below the lowest excitation level or metastable state. For example, the tube drop across a particular type 866-A mercury-vapor diode rectifier tube maintained at a temperature of 100°C was found to be less than 4 volts, whereas the lowest state in mercury vapor is the 4.66-volt level. This type of result was very baffling until Compton and Eckart [23] found, using the probe method to be described below, that an accumulation of positive charge existed in the neighborhood of the cathode. This caused a potential maximum (as indicated in curve 3, Fig. 10-17) to exist, the height of which was enough to permit ionization by electron collision, even though the total anode-cathode potential was too small to provide for such ionization.

**10-16. Plasma Characteristics.**[22] The largest portion of a discharge is the plasma. As already mentioned, the plasma is a region that contains substantially equal numbers of positive ions and electrons, so that the net charge in this region is zero. The impression should not be gained that the plasma contains only electrons and positive ions. In fact, it consists almost entirely of non-ionized gas molecules. The concentration of electrons and ions rarely exceeds several per cent of the gas concentration and may actually be only a small fraction of 1 per cent of the latter.

An electron which collides with a gas molecule gives very little kinetic energy to the molecule (because the molecular mass is so much larger than that of the electron). If ionization results from the collision, then the energy picked up from the applied field by the incident electron divides (after subtracting the ionization energy) between the two electrons. Hence the electrons in the plasma have relatively large energies, of the order of a few electron volts. The electrons now make elastic collisions with the molecules and thereby obtain a random distribution of velocities which turns out to be Maxwellian. It is thus possible to introduce the concept of the equivalent temperature of the electron gas. This is of the order of tens of thousands of degrees—corresponding to a few electron volts energy.

Because of their smaller mass the electrons will diffuse out of the plasma more rapidly than the ions. They will go to the walls of the tube containing the gas and will charge it a few volts negative with respect to the main portion of the plasma. This negative potential will repel electrons and attract positive ions, and in the equilibrium state the net current to the walls must be zero.

The region between the plasma and a boundary across which there is a large potential gradient is designated by the name *sheath*. A sheath may

form around any probe or electrode that is immersed in the plasma. The sheath that forms at the surface of an insulator which is negative with respect to the plasma (for example, the glass walls of the discharge tube) will arise from the presence of an excess number of positive ions. However, sheaths may contain either electrons or positive ions, an electron sheath forming about a probe which is positive with respect to the plasma.

Although the average energy of the electrons in the plasma is of the order of a few electron volts, nevertheless a few electrons possess energies that are high enough to cause ionization by collision, in accordance with the MB distribution function. This situation must prevail because otherwise the random velocities of the ions would cause them all to pass ultimately through the sheaths at the boundaries of the plasma. This loss of ions must be compensated for by the production of other ions in order to maintain the plasma.

If the area of the plasma is large compared with the volume, then the rate at which deionization takes place at the walls will be high. For sufficient ionization to take place in the plasma to replace these lost ions, the plasma can no longer be an equipotential region but there must be an appreciable voltage drop across it. Such a situation exists in a neon advertising sign, which consists of a long tubular glass envelope of small cross section. The drop in the plasma in such a tube is found to be 100 to 200 volts/ft.

**10-17. The Theory of Probes.** The physical quantities of interest with respect to a plasma are the following: the electron temperature $T_e$; the electron concentration $N_e$; the positive-ion concentration $N_i$; the random electron current density $J_{re}$; the random positive-ion current density $J_{ri}$; the plasma potential $E'$; the floating potential $E_f$.

Langmuir and Mott-Smith [24] developed the theory of probes and showed how these quantities could be measured. A "probe," or "sounding electrode," is an insulated conductor which is inserted into the plasma of a discharge. If a known voltage (with respect to the cathode) is applied to the probe and the current to it is measured, the form of the current variation found is that illustrated in Fig. 10-19. For large negative potentials with respect to the space or plasma potential, the saturation region $AB$ results. Since positive ions will be attracted and electrons repelled by the action of this negative potential, the saturation region $AB$ is that for which only positive ions will be collected. This is analogous to drawing saturation current from a thermionic diode; the current becomes independent of the plate voltage. This implies, of course, that no excess secondary electrons will be liberated from the probe surface as a result of the positive-ion bombardment. This assumption is generally valid for mercury ions and applies for most ions that have fallen through small potentials. The positive-ion current is plotted as a negative current.

As long as the probe is negative with respect to the plasma, the saturation positive-ion current will be collected. However, as the voltage of the probe is made more positive, all the electrons are no longer repelled, so that the difference between the positive-ion current and the electron current drifting to the probe will be collected. Hence, the true electron current $I_e$, at any voltage, is that indicated on the diagram by the curve $BC$. At a potential designated the *floating potential*, the number of positive ions collected and the number of electrons collected will be equal to each other, so that an ammeter in the external probe circuit will be zero. This corresponds to the point $F$. It follows from the discussion given in the preceding section that the potential at $F$ must be negative with respect to the plasma. As the probe potential is increased beyond the point $F$, the number of electrons that are collected will increase rapidly, since the retarding action of the probe potential with respect to the plasma potential is being decreased. The quantitative relation that exists between the voltage and the current to the probe in this region will be discussed in the next section.
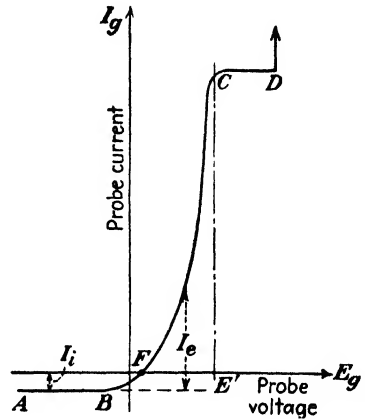


Fig. 10-19. Volt-ampere characteristic of a probe in the plasma of a discharge. This diagram is not drawn to scale since the maximum electron current is usually about 1,000 times the positive-ion current.

When the probe is at the same potential as the plasma, the probe will exert no repelling force on the electrons and the ions and so those that diffuse to the probe because of their random drift will be collected. Consequently, further increases in potential of the probe will not result in an increased electron current. Potentials more positive than $E'$ produce a field that repels the positive ions. But the random electron current is very much greater than the positive-ion current. Hence a region $CD$ which is practically horizontal exists. Thus a more or less sharp break will exist in the curve at the potential $E'$, which then represents the potential of the plasma. If the voltage of the sounding electrode is increased sufficiently, the electrons attracted to the probe may have sufficient energy to ionize the molecules in the sheath. These will neutralize the electron space charge, the probe current will rise rapidly, and the probe will become an auxiliary anode as at point $D$ in Fig. 10-19.

**10-18. Probe Measurements.** As discussed above, when the potential of the probe is negative with respect to the plasma, it will be surrounded by a sheath of positive charge which will separate it from the plasma. If

$E_r$ is the potential drop across this sheath, then only those electrons will be collected which have sufficient energy to overcome this retarding potential $E_r$. The situation is exactly analogous to that in which a retarding potential is applied between the plate and the cathode of a vacuum diode. The plate current, therefore, consists of those electrons which are emitted from the cathode with sufficient energy to overcome the retarding potential. Under the present conditions, the variation of electron current $I_e$, with variations of the sheath potential, is given by Eq. (5-54), *viz.*,

$$I_e = AJ_{re}\epsilon^{-E_r/E_T} \tag{10-18}$$

where $A$ is the area of the sheath, which is the area of the probe exposed to the discharge; $J_{re}$ is the random current density of the electrons in the plasma and corresponds exactly to $J_{th}$, the saturation thermionic current density in the Dushman equation; and $E_T$ is the voltage equivalent of the electron temperature.

The random current density corresponding to a MB distribution is given by Eq. (8-45),

$$J_r = Ne\sqrt{\frac{kT}{2\pi m}} \tag{10-19}$$

where $N$ is the concentration of the particles under consideration. It is seen that $J_r$ depends solely upon the conditions within the plasma and is independent of the probe material.

It is important to keep in mind the fact that the foregoing equations are based on the assumption that the electrons in the plasma possess a Maxwellian distribution of velocities. If the distribution is non-Maxwellian, then these equations are no longer valid and none of the foregoing discussion is valid. In order to test whether or not the assumption of a Maxwellian distribution in the plasma is correct, the logarithm of Eq. (10-18) is plotted. That is, the following expression is plotted:

$$\log_{10} I_e = \log_{10}(AJ_{re}) - 0.434\frac{E_r}{E_T} \tag{10-20}$$

Since the potential of the probe with respect to the cathode is $E_g$, then the potential across the sheath is

$$E_r = E' - E_g$$

Equation (10-20) may then be written as

$$\log_{10} I_e = \log_{10}(AJ_{re}) - 0.434\frac{E' - E_g}{E_T} \tag{10-21}$$

If a plot of $\log_{10} I_e$ vs. $E_g$ yields a straight line, then the assumption of a

Maxwellian distribution in the plasma is verified. The slope of this straight line is $0.434/E_T$, from which the equivalent electron temperature may be calculated, since $E_T = T/11,600$. If this logarithmic plot is not linear, then the distribution is not Maxwellian.

For a Maxwellian distribution, the probe potential $E'$ is that indicated by the break in the curve at the point $C$ of Fig. 10-19 or by the break in the corresponding logarithmic plot. The saturation current in the region beyond $C$ represents the random electron current to the probe. If the probe area $A$ is known, then the random electron current density $J_{re}$ is given by the ratio of the saturation current to the area of the probe. This states, in effect, that when the probe potential equals the plasma potential, i.e., when $E_e = E'$, then $E_r = 0$ and Eq. (10-18) reduces to $I_e = AJ_{re}$. Also, since $J_{re}$ and $T_e$ are both known, then Eq. (10-19) permits a calculation of the electron concentration $N_e$.

The use of a number of probes or the use of a movable probe in the plasma of a discharge reveals that the plasma potential varies only slightly and approximately linearly from point to point between the cathode and anode. This indicates that $d^2V/dx^2$ must be zero, approximately. This requires, in accordance with Poisson's law, that the net space charge be zero. If it is assumed that only singly ionized atoms are present in the plasma, and multiple ionization would not be expected in such a region, it then follows that the concentration of positive ions equals that of the electrons, whence $N_i = N_e$.

The floating potential $E_f$ (sometimes called the "wall potential") is determined experimentally by noting the voltage at which the net current to the probe is zero. At this point, the electron current equals the ion current, whence Eq. (10-18) yields

$$J_{ri} = J_{re}\epsilon^{-(E'-E_f)/E_T} \tag{10-22}$$

The region $AB$ (Fig. 10-19) gives the total random ion current $I_i$. The random positive-ion current density is, therefore, $J_{ri} = I_i/A$. Although all the terms in this equation are known, the value of $E'$ can be measured directly with the least accuracy. It is therefore desirable to use this equation to calculate this quantity.

**10-19. Sheath Thickness.** In the region $AB$ (Fig. 10-19), the electron current is zero, and the positive-ion current is a constant. What then is the consequence of variations of probe potential? The answer to this question can be found by a closer examination of the sheath that so effectively shields the plasma from the probe. This sheath is a region containing charges of one sign only (positive ions, in this case), and the conditions here differ very little from those which exist in the space-charge-limited vacuum diode. Within the range of probe voltages for which this analogy is valid, the same three-halves-power law must also be applicable. Thus,

Eq. (7-15) may be written in the form

$$x = \left[\frac{2.33 \times 10^{-6}}{J_{ri}(m_i/m_e)^{\frac{1}{2}}}\right]^{\frac{2}{3}} E_r^{\frac{1}{2}} \qquad (10\text{-}23)$$

where $x$ now represents the sheath thickness and $m_i/m_e$ is the mass ratio of ion to electron.

As repeatedly emphasized, the random ion current density $J_{ri}$ is independent of $E_g$ and hence of $E_r$. This equation clearly indicates that the effect of varying the probe potential is simply to vary the thickness of the sheath. The sheath in a luminous plasma is conspicuous by its lack of luminosity. This permits $x$ to be measured as a function of $E_r$. The results so obtained agree well with the results calculated from Eq. (10-23).

With nominal values of currents and voltage this equation yields sheath thicknesses of approximately $10^{-4}$ m. This is so small that the probe disturbs the discharge hardly at all. In other words, *as the probe voltage is varied, the anode current remains unchanged.* The practical consequences of this result are discussed in Chap. 12 in connection with the thyratron tube.

As $E_r$ is decreased, the sheath thickness also decreases. At the point where $E_g = E'$, the sheath thickness becomes zero. If the probe potential is raised above the plasma potential, positive ions are repelled and electrons are attracted so that an electron sheath results. In this case, the mass ratio in Eq. (10-23) is unity, and $J_{ri}$ must be replaced by $J_{re}$.

**10-20. Discharges in High-frequency Fields.**[25] An important class of discharges are those which occur in the very high-frequency fields that exist in certain microwave discharge tubes. The mechanism for the discharge in these tubes is completely different from that which maintains a low-frequency discharge.

As discussed at some length in Sec. 10-2, the d-c discharge mechanism is the cumulative ionization of the electrons that leave the cathode and multiply as they proceed toward the anode. Such discharges depend on the cathode. The high-frequency a-c discharge is maintained entirely from electrons which are obtained from ionization by collisions in the gas and does not depend on the electrodes. Breakdown occurs when the number of electrons that are produced by ionization by collision in the gas equals the number of electrons that are lost by diffusion in the gas.

Owing to the inelastic collisions that occur in the high-frequency discharge, the distribution of electrons is no longer Maxwellian, and the discharge possesses important differences from the d-c discharge. It is of some interest that such high-frequency discharges show a minimum in the field strength vs. pressure diagram, but this is not a Paschen minimum.

High-frequency discharges find a very important application in what is called the TR (transmit-receive) switch of a radar set.  The essentials of such a system are shown schematically in Fig. 10-20.  The TR switch is a relatively simple plane-parallel gas diode mounted into the transmission line.  The requirement is that during the interval when the high-frequency high-power generator (the magnetron) is sending out a pulse, it must not get to the receiver (since it would certainly overload the receiver).

The operation is as follows: The TR tube breaks down almost instantly under the excitation of the high voltage pulse, and the discharge essentially short-circuits the input to the receiver.  The high energy pulse is transmitted to the antenna and radiates into space.  The reflected echo is picked up by the antenna.  It is so weak that it does not have enough energy to excite the gas in the TR switch, and hence it passes down the transmission line to the receiver.
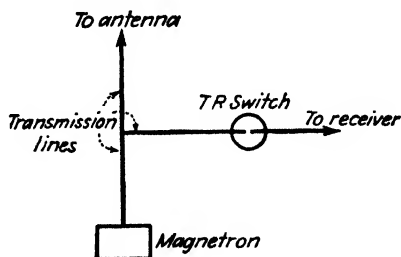


Fig. 10-20.  Illustrating the use of a TR switch in a radar set.

## PROBLEMS

**10-1.** Consider a Townsend discharge in which there is zero concentration of electrons at the cathode (no photocurrent, and the $\gamma$ emission is negligible).  There is a constant volume ionization due to an external source of $q$ ions per cubic meter per second. Show that the current density at any distance $x$ is given by

$$J = \frac{qe}{\alpha}(\epsilon^{\alpha x} - 1)$$

where $e$ is the electronic charge and $\alpha$ is the first Townsend coefficient.

**10-2.** *a.* In a gas photocell there is a pressure for which the current is a maximum at a given field intensity (or since the spacing is fixed, at a given voltage).  Show that this is the pressure at which a line through the origin is tangent to the $\alpha/p$ curve.  This corresponds to the maximum of $\eta$.

*b.* If $\alpha/p$ can be approximated by Eq. (10-4), show that the optimum pressure is given by $p = \varepsilon/B$.

**10-3.** Show that the first Townsend coefficient is given by the expression

$$\alpha = \frac{1}{l}\epsilon^{-E_i/\varepsilon l}$$

where $l$ is the mean free path, $\varepsilon$ is the electric-field intensity, and $E_i$ is the ionization potential of the gas.

The proof depends upon the following assumptions:

*a.* If an electron, having less than the ionization potential, collides with a molecule, then the collision is an elastic one.

*b.* The probability of ionization is unity if the colliding electron possesses an energy greater than $E_i$.

*c.* The survival equation (8-15) is valid.

Show that this result can be put in the form of Eq. (10-4).

**10-4.** Prove that an electron released at the cathode of a gas diode with plane-parallel electrodes will yield $\epsilon^{\alpha d}/(1-y)$ electrons at the anode. This is essentially Eq. (10-11).

PROB. 10-4.

The symbols have the meanings used in Sec. (10-4) with the added abbreviation $y = \gamma(\epsilon^{\alpha d} - 1)$.

Proceed by considering the sequence of events indicated in the sketch. One electron released at the cathode multiplies to $\epsilon^{\alpha d}$ electrons at the anode. Owing to this ionization, there are liberated $\epsilon^{\alpha d} - 1$ positive ions which travel to the cathode and there release $\gamma(\epsilon^{\alpha d} - 1) = y$ secondary electrons. These $y$ electrons multiply by ionization to $y\epsilon^{\alpha d}$ at the anode. Hence $y\epsilon^{\alpha d} - y$ positive ions travel to the cathode and liberate $\gamma(y\epsilon^{\alpha d} - y) = y^2$ secondary electrons, etc. The total number of electrons at the anode is obtained by adding the number that reach the anode owing to each of the individual events, as outlined above. The result is a geometric series with an infinite number of terms.

**10-5.** Using plane-parallel electrodes in air at 1 mm Hg pressure and a constant electric-field intensity of 160 volts/cm, the following data were taken:

| $d$, cm | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
|---|---|---|---|---|---|---|---|
| $\dfrac{I_b}{I_e}$ | 2.42 | 5.87 | 14.4 | 35.1 | 88.2 | 250 | 915 |

*a.* Verify the theory of Sec. 10-4 by finding the value of $\alpha$ and $\gamma$.

*b.* What are the breakdown separation and the breakdown voltage?

NOTE: These data are based upon the results of F. H. Sanders, *Phys. Rev.*, **44**, 1020, 1933.

**10-6.** Given two cold-cathode tubes as shown, containing the same gas at the same pressure. The cathode of tube 2 has a much larger surface than that in tube 1. Sketch the volt-ampere characteristic of each tube. Give the reasons for your expected curves.
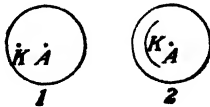
PROB. 10-6.
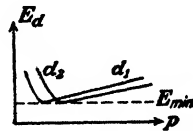
PROB. 10-7.

**10-7.** The sparking-potential curves of a gas as a function of pressure for two different separations, $d_1$ and $d_2$, of the plane-parallel electrodes are illustrated. Which is the greater distance, and why?

**10-8.** It is found that as the voltage is increased from zero, a discharge takes place between the cathode and anode through the long path 1 instead of the short path 2 of the tube sketched. Explain.
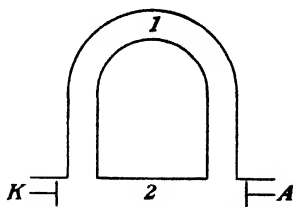


PROB. 10-8.

**10-9.** Using the tube of Prob. 10-8 at a constant voltage between $K$ and $A$, the pressure is gradually decreased. Explain why the discharge will first take the shorter path and then the longer path.

**10-10.** It is often said that a field of $3 \times 10^6$ volts/m is required for sparking in air at atmospheric pressure. Show that this is true only for a spacing of 1 cm. In particular, what is the breakdown field strength for 0.25 cm separation?

**10-11.** Given a gas at a constant pressure between plane-parallel electrodes. The breakdown curve as a function of separation of electrodes shows a minimum. Explain why this is to be expected physically.

**10-12.** Given a thermionic plane cathode and a plane-parallel collector a distance $d$ apart in an envelope containing a gas at a high pressure. The mean free path is so small that each electron makes many collisions with the gas molecules and hence may be considered to be traveling with a steady drift velocity proportional to the electric-field intensity. (In a vacuum where there is no mechanical hindrance due to collisions with molecules the electron would move with constant acceleration and not constant velocity.) Assume that the energy per mean free path is less than the ionization potential so that the tube is operating under space-charge conditions. If the drift speed $v$ is given by

$$v = -k\mathcal{E} = k\frac{dV}{dx}$$

show that the maximum current at a plate potential $E_b$ is given by

$$I_b = \frac{SkE_b{}^2}{32\pi \times 10^9 d^3}$$

where $S$ is the cathode area.

HINT: Proceed as in Sec. (7-2).

The proportionality factor between the electric-field intensity and the speed is called the *mobility*.

**10-13.** *a.* Consider a tube with plane-parallel electrodes between which there is a constant charge density $\rho$. Show that the potential at any point is given by the formula

$$V = \frac{E_b x}{d} + \left(\frac{\rho x}{2\epsilon_0}\right)(d - x)$$

where $d$ is the cathode-anode separation, $E_b$ is the anode potential, and $\epsilon_0$ is the permittivity of free space in mks units.

*b.* Show that if $\rho$ is positive there exists a potential maximum greater than the anode voltage. This corresponds to the electron trap of Fig. 10-17.

**10-14.** Show that the electric-field intensity at the cathode end of the cathode-fall region is equal to $4E_k/3d_k$, where $E_k$ and $d_k$ represent the cathode-fall potential and distance, respectively.

**10-15.** The cathode fall of potential for brass plane-parallel electrodes in air is 350 volts. Calculate the normal current density if the pressure is 1 mm Hg.

**10-16.** It is verified experimentally that the electrons in a certain mercury discharge obey the MB distribution law. The electron temperature is found to be 18,000°K. The random ion current density is 6.0 amp/m². The plasma is 6.8 volts above the floating potential. Calculate the following:

 *a.* The random electron current density.

 *b.* The electron and positive-ion concentrations.

 *c.* The thickness of the positive-ion sheath, if the probe is maintained 10 volts negative with respect to the plasma.

 *d.* The current collected by the probe if it is 6.0 volts negative with respect to the plasma. The probe area is 2.5 cm².

**10-17.** The following data were obtained with a plane probe in the plasma of a mercury discharge: *

| $E$, Voltage of Probe below That of Anode | $I$, Electron Current, Ma |
|---|---|
| 5 | 980 |
| 8 | 980 |
| 10 | 970 |
| 11 | 930 |
| 12 | 800 |
| 13 | 450 |
| 14 | 200 |
| 15 | 80 |
| 16 | 30 |
| 17 | 3.8 |
| 18 | −0.8 |
| 19 | −2.0 |
| 20 | −2.5 |
| 30 | −2.6 |

The area of the probe was 3.61 cm². The vapor pressure is 0.025 mm Hg. Verify that the MB distribution is valid by plotting these data in the appropriate manner. From this plot calculate the following:

 *a.* The random electron current density.

 *b.* The random ion current density.

 *c.* The floating potential.

 *d.* The plasma potential.

 *e.* The electron temperature.

 *f.* The electron and positive-ion concentrations.

 *g.* The ratio of ion to molecular concentration (see Fig. 11-1).

 *h.* The positive-ion sheath thickness when the probe is 7 volts negative with respect to the plasma.

 *i.* The electron sheath thickness when the probe is 7 volts positive with respect to the plasma.

* LANGMUIR, I., and H. MOTT-SMITH, JR., *Gen. Elec. Rev.*, **27**, 538, 1924.

## REFERENCES

1. KRUITHOF, A. A., *Physica*, **7**, 519, 1940.
   HUXFORD, W. S., *Phys. Rev.*, **55**, 754, 1939.
   KRUITHOF, A. A., and F. M. PENNING, *Physica*, **6**, 430, 1937.
   SANDERS, F. H., *Phys. Rev.*, **41**, 667, 1932; **44**, 1020, 1933.
2. DRUYVESTEYN, M. J., and F. M. PENNING, *Revs. Modern Phys.*, **12**, 87, 1940.
3. LOEB, L. B., "Fundamental Processes of Electrical Discharges in Gases," John Wiley & Sons, Inc., New York, 1939.
4. COBINE, J. D., "Gaseous Conductors," p. 149, McGraw-Hill Book Company, Inc., New York, 1941. See also Prob. 10-3.
5. COBINE, *op. cit.*, p. 152.
   ZWORYKIN, V. K., and E. G. RAMBERG, "Photoelectricity," John Wiley & Sons, Inc., New York, 1949.
6. COBINE, *op. cit.*, p. 159.
7. KRUITHOF, A. A., and F. M. PENNING, *Physica*, **3**, 315, 1936.
8. DRUYVESTEYN and PENNING, *loc. cit.*
   DARROW, K. K., "Electrical Phenomena in Gases," The Williams & Wilkins Company, Baltimore, 1932.
9. NEWTON, R. R., *Phys. Rev.*, **73**, 570, 1948.
10. JACOBS, H., and A. P. LaROCQUE, *J. Applied Phys.*, **18**, 199, 1947.
11. LOEB, L. B., *Revs. Modern Phys.*, **20**, 151, 1948.
    LOEB, L. B., and J. M. MEEK, "The Mechanism of the Electric Spark," Stanford University Press, Stanford University, Calif., 1941.
12. COBINE, *op. cit.*, p. 252.
13. SLEPIAN, J., and R. C. MASON, *Elec. Eng.*, **53**, 511, 1934.
14. HALE, D. H., and W. S. HUXFORD, *J. Applied Phys.*, **18**, 586, 1947.
15. PUPP, W., *Physik. Z.*, **34**, 756, 1933.
16. COBINE, *op. cit.*, p. 290.
17. SLEPIAN, J., and A. H. TOEPFER, *J. Applied Phys.*, **9**, 483, 1938.
    LUBCKE, E., *Z. tech. Physik.*, **10**, 598, 1929.
18. LANGMUIR, I., *Gen. Elec. Rev.*, **26**, 731, 1923.
19. LOEB, L. B., "Fundamental Processes of Electrical Discharges in Gases," pp. 629–636, John Wiley & Sons, Inc., New York, 1939.
20. SLEPIAN, J., *J. Franklin Inst.*, **201**, 79, 1926.
21. NOTTINGHAM, W. B., *Phys. Rev.*, **28**, 764, 1926.
22. LANGMUIR, I., *J. Franklin Inst.*, **214**, 275, 1932.
    "Industrial Electronics Reference Book," p. 48, John Wiley & Sons, Inc., New York, 1948.
23. COMPTON, K. T., and C. ECKART, *Phys. Rev.*, **25**, 139, 1925.
24. LANGMUIR, I., *J. Franklin Inst.*, **214**, 275, 1932.
    LANGMUIR, I., and H. MOTT-SMITH, JR., *Gen. Elec. Rev.*, **27**, 449, 538, 616, 762, 810, 1924.
25. MARGENAU, H., and/or L. M. HARTMAN, *Phys. Rev.*, **73**, 297, 309, 316, 326, 1948.
    MACDONALD, A. D., and S. C. BROWN, *ibid.*, **75**, 411, 1949.

### General References

BÄR, R.: "Die Glimmentladung," *Handbuch der Physik*, **14**, 171, 1927.
DARROW, K. K.: "Electrical Phenomena in Gases," The Williams & Wilkins Company, Baltimore, 1932.

Dow, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

ENGEL, A. v., and M. STEENBECK: "Elektrische Gasentladungen," Vol. 2, Verlag Julius Springer, Berlin, 1934.

HAGENBACH, A.: "Der elektrische Lichtbogen," *Handbuch der Physik*, **14**, 324, 1927.

REICH, H. J.: "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.

SLEPIAN, J.: "Conduction of Electricity in Gases," Westinghouse Electric & Manufacturing Company, Educational Department, Course 38, 1933.

THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," 3d ed., Cambridge University Press, London, 1933.

TOWNSEND, J. S.: "Electricity in Gases," Oxford University Press, New York, 1915.

# CHAPTER 11

# COMMERCIAL GAS TUBES

GAS-FILLED tubes depend for their operation on the characteristics of gaseous discharges discussed in the preceding chapter. Many different types of gas-filled tubes are available commercially, the most important of which are the two-element hot-cathode tube (the low-voltage externally heated arc tube); the two-element cold-cathode tube (the glow tube *); the two-element pool-type tube (the mercury-tank rectifier); the three-element hot-cathode tube (the thyratron *); the three-element cold-cathode tube (the grid glow tube *); and a number of single-anode multiele-ment pool-type tubes (the excitron and a variety of ignitrons). The most important use of these tubes, that of rectification, will be the subject of a detailed discussion in Chaps. 12 and 14. Several of the other applications, such as voltage regulators, light sources, control devices, and arcless switching, will be considered below.

**11-1. Hot-cathode Gas-filled Diodes.** These are low-voltage non-self-maintaining arc tubes and were discussed in essence in the previous chapter. These tubes are frequently referred to as "phanotrons." * They may be provided with an inert gas (neon, argon, etc.) at low pressure, although more commonly the gas is mercury vapor. In the latter case, a few drops of mercury are added to the tube after evacuation. The pressure in the tube is then a function of the mercury-vapor condensation temperature. The relationship between the pressure and temperature is shown in Fig. 11-1. Under normal conditions of operation of a tube of this type, the temperature of the tube will be about 25°C higher than the temperature of the surroundings (known as the "ambient" temperature). This higher temperature arises from the presence of the heated cathode and from the effect of the energy interchange in the tube itself.

The breakdown potential is usually of the order of magnitude of the ionization potential of the gas present in the tube, although it will vary with the gas pressure. At the lower pressures (which occur at low temperatures in mercury-vapor tubes) the number of collisions made by an electron in its

---

* These terms were formerly trade-marks of the General Electric Co. and The West-inghouse Electric & Manufacturing Co., but have since been relinquished for general use. The definitions and descriptions used in the text are those given in "Industrial Tube Terminology," *Electronics*, **11**, 29, September, 1938.

flight from the cathode to the anode will be relatively small. In order to provide sufficient ionization, the applied potential must be increased, and thus the probability of ionization will be higher. Consequently, high breakdown potentials, perhaps two or three times the ionization potential, may exist in low-pressure tubes.

If the gas pressure is high (at the higher temperatures in mercury-vapor tubes), the mean free path is small. Although the energy per mean free path may be small for small voltages, nevertheless higher pressures favor
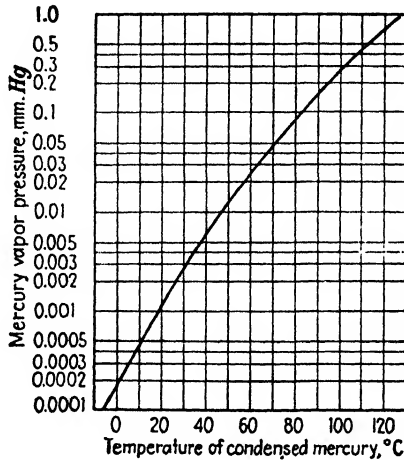


FIG. 11-1. Mercury-vapor pressure as a function of condensation temperature.

ionization by collisions of the second kind involving metastable states. The net effect is a generally falling breakdown potential with increasing gas pressures, indicating that the operating pressures are to the left of the Paschen minimum.

Once conduction has started, the tube drop will remain substantially constant and independent of the tube current at a value that is of the order of the ionizing potential of the gas. The tube drop at any pressure is generally several volts less than the breakdown potential. As discussed in Sec. 10-15, the tube drop may even be less than the potential of the first energy level of the gas.

An oscillogram of the volt-ampere characteristics of an 866-A mercury-vapor diode is given in Fig. 11-2. The substantially constant working voltage at a smaller value than the breakdown potential should be noted. If the current is decreased continuously after conduction has started, either by decreasing the applied voltage or by increasing the magnitude of the external resistance, the tube drop will fall below the voltage necessary for

maintaining the arc, and the arc will go out. This will occur when insufficient ionization takes place to maintain the discharge.

It should be emphasized that the maximum current obtainable in a phanotron is the saturation thermionic emission from the cathode enhanced by the effect of the field at the surface of the cathode (see Sec. 5-19). This has been shown to be about 1.8 times the actual thermionic-emission current.[1] The sole function of the gas in these tubes is to provide ions for the neutralization of space charge, thereby permitting the current to be obtained at much lower voltages than are necessary in vacuum tubes. Under normal operating conditions, the total drop across a mercury-filled thermionic heater gas tube will range from 10 to 15 volts. If more current is passed through the tube than the cathode can deliver, the tube drop increases to the point where the additional electrons are produced by positive-ion bombardment. However, if the tube drop in such a tube exceeds about 22 volts, the positive-ion bombardment will cause the cathode to disintegrate. This voltage is known as the *disintegration* voltage. Thus, in operation, the voltage across a mercury-vapor tube should never be permitted to approach this value. Consequently, a rising potential characteristic of a mercury-vapor tube is an indication that the cathode will not supply the electrons demanded of it.

Fig. 11-2. Volt-ampere characteristic of an 866-A gas diode. Breakdown voltage = 15 volts. Maintaining voltage = 11 volts. Ambient temperature = 25°C.

The types of oxide-coated cathode that are used in these tubes have been considered in Sec. 6-6 (see Figs. 6-6 and 6-7). It is very important in the larger tubes which are equipped with the inward-radiating cathodes that sufficient time has elapsed for the cathode to reach its operating temperature before any anode potential is applied. This heating time may extend from perhaps 30 sec to as much as 30 min, depending upon the size of the cathode. If the anode potential is applied before the cathode can supply the electrons demanded of it, the tube will begin to operate as a self-maintained discharge, with the result that intense cathode sputtering will take place. In this case, it will be the active oxide-coated surface on the cathode that will be flaked off, and the cathode will be permanently injured.

The glow in a phanotron comes from the plasma region of the discharge. The dark space near the walls is due to the positive-ion sheath which contains no electrons and hence is a region in which no ionization takes place.

A group of phanotrons is shown in Fig. 11-3. The range of voltages and currents of these tubes is given in the caption.

Fig. 11-3.   Commercial hot-cathode diodes.   (*Courtesy of General Electric Co.*)

The smallest tube is the FG-190 full-wave, gas-filled rectifier.   It is approximately 3 in. long and 1.5 in. in diameter.   The tube drop is 8 volts.   The average anode current is 1.25 amp, the peak current is 5 amp, and the surge current for 0.1 sec is 20 amp.

The largest tube is the FG-166 half-wave all-metal mercury-vapor rectifier.   It is approximately 14 in. long and 5 in. in diameter.   The tube drop is 9 volts.   The average anode current is 20, the peak current is 75 amp, and the surge current is 750 amp. The tube is capable of delivering 50 kw of power.

**11-2. Mercury-pool Rectifiers.**   A very important gas-filled tube, and the word "tube" is used in a rather broad sense, is the pool-cathode mercury-arc tube.   A basic form, which is primarily of historic interest, consists of two anodes, usually of carbon, sealed through arms in a glass envelope at the bottom of which is a pool of metallic mercury which acts as the cathode.   The two anodes are necessary in order to permit the steady operation of such a unit.   Otherwise, an auxiliary anode or an ignitor rod must be provided.   Such a glass-bulb mercury rectifier tube is illustrated in Fig. 11-4.   In these units the arc is started by tipping the bulb so that the mercury pool makes contact with the starting electrode and then setting the bulb in its upright position.   This causes the mercury connection between the cathode and the starting electrode to break, and the arc is

established. This arc forms the cathode spot on the surface of the mercury. Sufficient ionization is produced to allow the current to pass to the main electrodes, and continued operation is possible.

In this type of rectifier, the electrons enter the arc through the cathode spot, which wanders on the surface of the pool of mercury. This motion is due to the forces exerted by the bombarding ions and the reaction of the vaporizing gas. The ionized mercury ions that are formed are neutralized at the glass surface of the envelope and condense on the walls. The evaporated mercury returns to the pool under the action of gravity. It is necessary, of course, that the cathode pool contain sufficient mercury to permit continuous operation at the cathode spot.

If the arc should be broken for any reason whatsoever, recombination of the mercury ions occurs and deionization of the vapor takes place. The arc will remain out until it is again initiated by tipping the tube. It is for this reason that certain tubes are provided with "keep-alive" electrodes, a steady arc being maintained between these auxiliary electrodes at all times. These auxiliary electrodes are not absolutely necessary for tubes of the type shown in Fig. 11-4 when used in an appropriate circuit. Such keep-alive electrodes would be necessary in single-anode mercury-cathode units, since otherwise the tube would be extinguished during the negative half cycle of an applied a-c voltage. The single-anode "excitron" (see Sec. 11-7) is of this type.

Fig. 11-4. Glass pool-cathode full-wave mercury-arc rectifier.

The larger size units, which are designed to operate on polyphase a-c circuits, are mounted in steel tanks, although polyphase glass rectifier systems having power capacities as high as 100 kw have been made commercially. The steel-tank polyphase rectifiers are built in units with power capacities as high as 3,000 kw. These larger units are provided with vacuum pumps in order to maintain a low pressure in the tank. A movable starting electrode is provided in the units to initiate the arc.

It is of interest to note that polyphase tank rectifiers are rapidly becoming obsolete, and are being superseded by banks of single-anode tanks, for reasons which appear in Sec. 11-17. A photograph of such a single-anode polyphase system is given in Fig. 11-17.

**11-3. High-pressure Gas Diodes.** These diodes contain argon or a mixture of argon and mercury at a pressure of about 5 cm Hg. They contain short heavy filaments located close to heavy graphite anodes. These tubes, which are used extensively for charging automobile storage batteries, are known as "Tungar" [2] or "rectigon" tubes.

The presence of the fairly high pressure gas serves a twofold purpose. One is to provide the positive ions for reducing the space charge, in order to permit large currents. The second is to prevent the evaporation of the thorium or the oxide coating from the filament. This second effect is extremely important since the filament is operated at above-normal temperature, in order to provide the large currents from such a simple cathode structure.

Although these tubes have the advantage that fairly high currents are possible with a simple unit, they are limited in their application because the sparking potential is low at the high pressures that are used. These tubes are suitable only for use on low-voltage circuits.
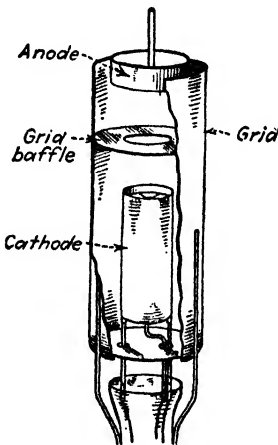


FIG. 11-5. Electrode structure of the type FG-27A (or FG-57) negative-control thyratron. (*From H. J. Reich, "Theory and Applications of Electron Tubes," McGraw-Hill Book Company, Inc.*)
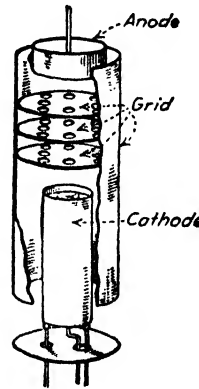
FIG. 11-6. Electrode structure of the type FG-33 positive-control thyratron. (*From H. J. Reich, "Theory and Applications of Electron Tubes," McGraw-Hill Book Company, Inc.*)

**11-4. The Thyratron.**[3] The insertion of a massive grid, so as to provide almost complete electrostatic shielding between the cathode and anode of an externally heated thermionic arc tube, permits control of the initiation of the arc by controlling the potential of the grid. The grid usually consists of a cylindrical structure which surrounds both the anode and the cathode, a baffle or a series of baffles containing small holes being inserted between the cathode and the anode. Such tubes are illustrated in Figs. 11-5 and 11-6. Because of the almost complete shielding between the cathode

and the anode, the application of a small grid potential before conduction is inaugurated is adequate to overcome the field at the cathode resulting from the application of a large anode potential. That is, a small grid voltage may neutralize the effects of a large anode potential and so prevent the arc from being initiated.

Once the arc has been initiated, the grid loses complete control over the arc, since now the grid is immersed in a plasma and is separated therefrom by the positive-ion sheath. Any variations in grid potential merely result in changing the thickness of the sheath. This behavior of the grid is to be expected in view of the theory of probes developed in Sec. 10-19. Grid control is reestablished only when the anode potential is reduced to a value below that necessary to maintain the arc. Once the arc has been extinguished by lowering the plate voltage, the grid once more becomes the controlling factor and again determines when conduction will be initiated. This control characteristic is frequently referred to as the "trigger" action of the grid.

Since the grid merely acts as a probe situated in the plasma of the arc, it may be used to investigate whether or not the electronic energy distribution in the plasma is Maxwellian. An oscillogram of the volt-ampere grid characteristic of a General Electric Co. FG-27A thyratron, with anode currents of 0.5 and 2.5 amp, is shown in Fig. 11-7. The flat positive-ion regions and the exponentially rising electron-current portions should be noted and compared with the corresponding regions of Fig. 10-19.

FIG. 11-7. Volt-ampere grid characteristics of an FG-27A thyratron for two values of anode current (0.5 and 2.5 amp). The peak grid electron current shown in the oscillogram is 450 ma.

The main function of the grid is to control the initiation of the arc. If the grid potential is more positive than that necessary for the controlling action to prevail, conduction will take place; if more negative, no conduction will occur. The curve that relates the grid ignition potential with the potential on the anode is known as the *critical grid curve*. This curve is all that need be known about a thyratron in order to determine completely its behavior in any circuit. That is, the control characteristics of the tube can be predicted from a knowledge of this one static curve.

Typical starting-characteristic curves of mercury-vapor thyratrons are given in Fig. 11-8. It will be observed that there are two distinct types of characteristic: those in which the grid potential must always be positive, and those in which the grid is generally negative, except for very low plate voltages. The physical distinction between these two types of tube lies essentially in the more complete shielding by the grid of the cathode from

the anode in the positive-control tube.    Compare the grid structures of
the FG-27A and the FG-33 tubes illustrated in Figs. 11-5 and 11-6.

As is seen from Fig. 11-8, the grid potential of an FG-33 tube must be
made positive for conduction to begin.    That is, in order that the electrons
that leave the cathode in a positive-control tube may acquire sufficient
velocity to cause ionization of the gas molecules by collision, the grid must
be made positive.    As a matter of fact, the shielding of the anode is so com-
plete that the potential of the plate exerts practically no influence on the
control action.    This means that the field at the cathode due to the anode
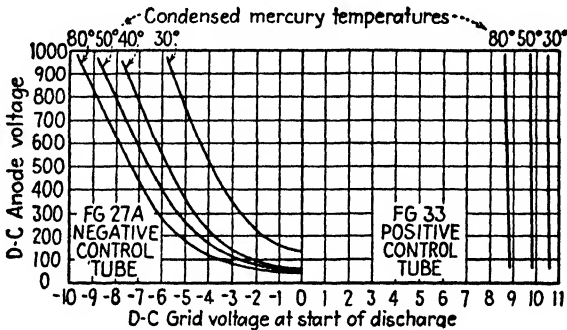


FIG. 11-8.    Critical grid characteristics of a positive- and a negative-control thyratron
with mercury condensation temperature as a parameter.    (*Courtesy of General Electric
Co.*)

potential is practically zero in the positive-control tubes, and the critical
grid voltage is independent of the plate potential.    For this reason the
grid potential must be sufficiently high to cause ionization and so conduc-
tion.

In the negative-control tube, where the shielding is far less complete,
the effect of the plate voltage is clearly seen; the higher the plate potential,
the more negative must the grid potential be in order to prevent conduction
from taking place.    For low plate voltages, positive grid voltages must be
applied before ionization by collision, and hence conduction, can begin.    If
the plate voltage is reduced still further, even below the potential necessary
for ionization, breakdown can still be obtained by making the grid suffi-
ciently positive.    Now, however, the function of the tube may be destroyed,
since the main arc may take place between the cathode and the grid, with
very little current to the plate.    The thyratron will be converted into a
gaseous diode under these circumstances with the plate acting as a dummy
electrode, the cylindrical grid now acting as the anode.

It is necessary that the grid circuit contain sufficient current-limiting
resistance in order to protect it from overload when such an arc takes place.
Although the grid structure could handle high currents, the leads are not

designed for high-current operation, and higher than rated current may overheat and crack the glass press at the point of entrance of the grid lead. It is good practice to fuse both the grid and the anode circuits.

As a specific illustration of the use of the starting characteristics, consider the circuit of Fig. 11-9 in which a thyratron is employed as a switch. In this circuit, the load may be a bell, a counter, or some other type of recording device.  The control device, which is in the grid circuit, may be a photocell circuit, a relay system of some type, or a mechanically or electrically operated switch.  If an FG-27A tube is used in conjunction with
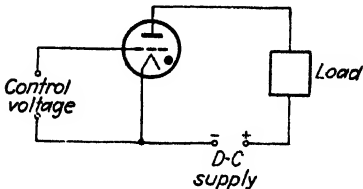


Fig. 11-9. A thyratron used as a switch. The load is not energized until the control voltage exceeds a critical value. (Heater circuit and an extinguishing circuit, if any, are not shown.)
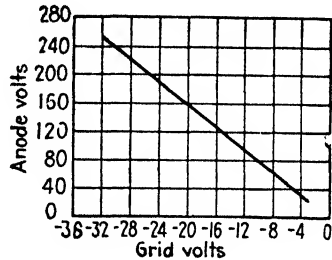
Fig. 11-10. Typical critical grid characteristic of an RCA 885 (or 884) argon-filled thyratron.

the 120-volt d-c lines, the grid voltage must be at least 3 volts negative with respect to the cathode (for a condensation temperature of 50°C) in order that no conduction occur.  If the control device in the grid circuit causes the grid voltage to become more positive than $-3$ volts, the tube will conduct and current will pass through the load.  An extinguishing circuit must be used to stop the plate current, if it is desired that the grid regain control.  Many practical applications of such circuits are possible.[4]

In addition to the mercury-vapor and gas-filled thyratrons of moderate current capacity, small argon-filled low-current-capacity tubes are available.  The shielding between the cathode and the anode is not so complete in these tubes as in the higher power units.  A typical critical grid curve for such an 885 or 884 tube is given in Fig. 11-10.  The critical grid curve is independent of temperature since the number of gas molecules in the glass envelope is constant.

A group of thyratrons is shown in Fig. 11-11.  The range of currents is given in the caption.

**11-5. The Sweep-circuit Oscillator.**  An important use of the small argon-filled thyratron is as a relaxation-type sweep-circuit generator or saw-tooth oscillator for use with cathode-ray tubes.  The diagram of a

circuit that employs such a tube to produce a saw-tooth voltage wave of the type illustrated in Fig. 3-11 is given in Fig. 11-12. To understand the operation of this circuit, assume that the capacitor $C$ is initially uncharged.
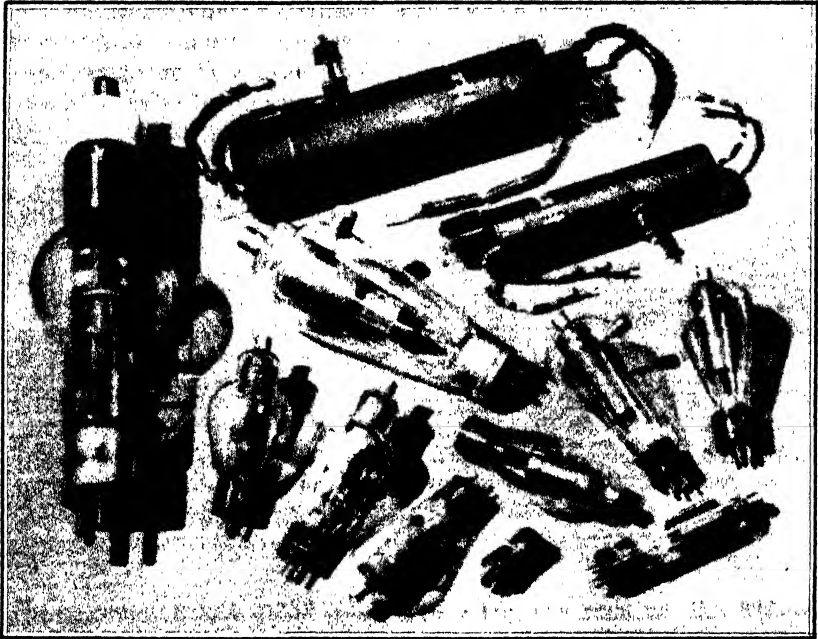


Fig. 11-11. Commercial thyratrons. (*Courtesy of General Electric Co.*)

The smallest tube is the GL-502A, which is $2\frac{5}{8}$ in. long and $1\frac{1}{32}$ in. in diameter. The peak forward voltage is 650 volts. The peak inverse voltage is 1,300 volts. The average current is 0.1 amp, the peak current is 1.0 amp, and the surge current for 0.1 sec is 10 amp.

The largest tube is the 414, which is over 15 in. long and $3\frac{1}{8}$ in. in diameter. The peak forward and the peak inverse voltage are each equal to 3,000 volts. The average current is 12.5 amp, the peak current is 100 amp, and the surge current is 1,500 amp.

After switch $S$ is closed, the voltage across the capacitor will increase exponentially according to the equation

$$e_c = E_{bb}(1 - \epsilon^{-t/RC}) \tag{11-1}$$

This curve is shown dashed in Fig. 11-13. Assume that the plate supply voltage is 250 volts and that the grid bias is 25 volts. The plate voltage corresponding to the grid voltage of 25 volts is found from Fig. 11-10 to be approximately 200 volts. Hence, the charging of the capacitor will continue only until the voltage across the capacitor reaches 200 volts. At

this time the tube will break down. Charge will leak off the capacitor very rapidly, and the tube will stop conducting when the anode current falls below that necessary to maintain ionization. The capacitor will again begin to charge through the local $RC$ circuit and the entire process will repeat itself. Figure 11-13 shows the resulting sweep voltage. If the amplitude of the swing is small compared with $E_{bb}$, then the exponential portions become almost linear and a sweep voltage proportional to time is obtained.
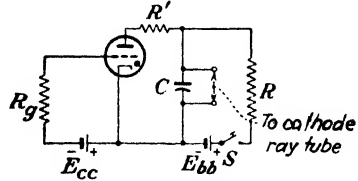


Fig. 11-12. A saw-tooth voltage generator employing an 885 thyratron. $R_g$ is the protective resistor in the grid circuit, and $R'$ is the protective resistor in the plate circuit. $R'$ should be as small as possible in order that the capacitor may discharge very quickly through the tube.

The period can be adjusted by varying $R$, $C$, or the voltages $E_{cc}$ or $E_{bb}$. The amplitude of the oscillation is determined only by $E_{cc}$ (and the maintaining voltage of the tube).

If $t'$ is the time measured from the instant when the sweep starts, then the sweep voltage is given by

$$e_c - E_e = (E_{bb} - E_e)(1 - \epsilon^{-t'/RC}) \qquad (11\text{-}2)$$

This equation is consistent with the following facts. At $t' = 0$, $e_c$ must equal the extinction voltage, $E_e$. At $t' = \infty$, $e_c$ must equal the $B$ supply voltage, $E_{bb}$. The time constant is $RC$.
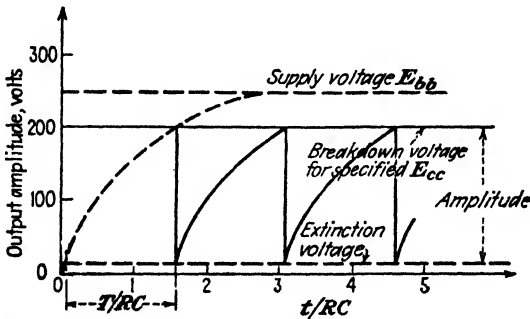


Fig. 11-13. The approximate output wave shape from the circuit of Fig. 11-12. The frequency is $f = 1/T$.

The amplitude of the sweep is given by the right-hand side of Eq. (11-2) with $t'$ equal to the period, $T$, of the oscillation. If $T/RC$ is much less than unity, the exponential term may be expanded in a power series and, keeping the first three terms of the expansion, there results

$$e_c - E_e = (E_{bb} - E_e)(t'/RC)(1 - t'/2RC) \qquad (11\text{-}3)$$

This shows that the maximum deviation from linearity is $T/2RC \times 100$ per cent.

If the charging current of the capacitor is too large, the circuit may "block," *i.e.*, cease oscillating. This arises because a steady-state d-c condition is possible in which the tube remains conducting, the capacitor stays charged at the maintaining tube drop, and the current through $R$ passes through the tube. In order to avoid this condition, the charging current must be less than that necessary to maintain the arc. If the resistor $R$ is chosen large enough so that the charging current is a fraction of a milliampere, then a steady arc will not be maintained. Under these conditions the capacitor will discharge through the tube until its voltage is less than the extinction voltage of the tube. The arc will then be extinguished, and a new charging cycle will begin.

If the capacitor is charged at a constant rate, then $e_c = I_0 t/C$, where $I_0$ is the constant current delivered to the capacitor, and a truly linear time axis results. To achieve this requires that the capacitor be charged from a constant-current unit rather than exponentially through the resistor $R$. This may readily be accomplished by replacing the resistor $R$ by a constant-current tube (a pentode). The charging rate is then controlled by controlling the tube current.

**11-6. Shield-grid Thyratrons.**[5] Before breakdown of the tube occurs, the current to the grid of a thyratron (such as the FG-27A) is a few tenths of a microampere. Although this is entirely negligible for most applications, it will introduce difficulties in circuits that require very high grid impedances. This is especially true in circuits that employ phototubes. For this reason a fourth electrode, or shield grid, has been added to the thyratron. Such a shield-grid thyratron is illustrated in Fig. 11-14. The massive cylindrical shield-grid structure encloses the cathode, control grid, and anode. This construction reduces the leakage current to a small fraction of its previous value.

Since the control grid is physically shielded from the cathode by the shield-grid baffle, the likelihood of contamination of the control grid resulting from cathode sputtering is greatly reduced. This reduces the possible effects of photoemission and thermal emission from the grid. As a result, the preignition grid current in this tube is of the order of $10^{-3}$ $\mu$a. In addition to the low grid current, this tube possesses the additional advantage that the grid-cathode electrostatic capacitance is small. This low capacitance results from the electrostatic shielding of the grid from the cathode by the shield. The capacitance is also reduced because the grid in the gas tetrode is much smaller than that in the triode. Compare the size of the control grids in Figs. 11-5 and 11-14.

The critical grid starting characteristics of such a tube are shown in

Fig. 11-15. It will be observed that these characteristics are functions of the shield-grid voltage. This feature adds to the versatility of the tube in possible applications.

The type 2050 and 2051 inert-gas tetrodes are low-current-capacity (75 to 100 ma average) shield-grid thyratrons. The shielding action of the control grid is much more complete than that in the 885 tube, and a
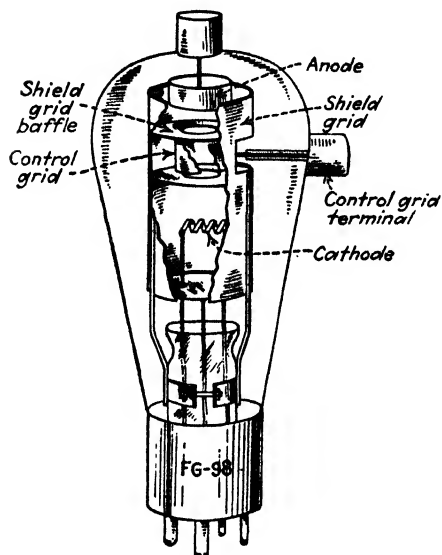


FIG. 11-14. Electrode structure of the FG-98 shield-grid thyratron. (*From H. J. Reich, "Theory and Applications of Electron Tubes," McGraw-Hill Book Company, Inc.*)
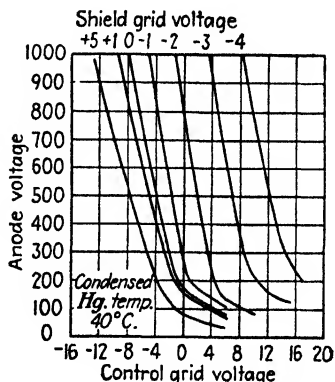
FIG. 11-15. Control characteristics of an FG-98 shield-grid thyratron.

steeper control characteristic is obtained. In fact, the critical grid characteristics are quite similar to those shown in Fig. 11-15. If the shield grid is connected to the cathode, the tetrode is converted into a very sensitive triode thyratron.

**11-7. The Excitron.**[6] As discussed in Sec. 11-2, if the arc in a pool-cathode rectifier is broken for any reason whatsoever, recombination of the mercury ions occurs and deionization of the vapor takes place. The arc will remain extinguished until it is again initiated. Of course, if an auxiliary circuit is provided in order to maintain the arc, even when the main anode is not conducting, then continued operation in the circuit is possible. The *excitron* provides such an auxiliary circuit, whereas the *ignitron* includes provision for igniting the arc at regular intervals.

The excitron is a single-anode pool-cathode rectifier tube which is provided with a holding or excitation anode, a control grid, and the main anode. A vertical section of such a tube is given in Fig. 11-16. In this tube an arc is struck to and is continuously maintained by the excitation
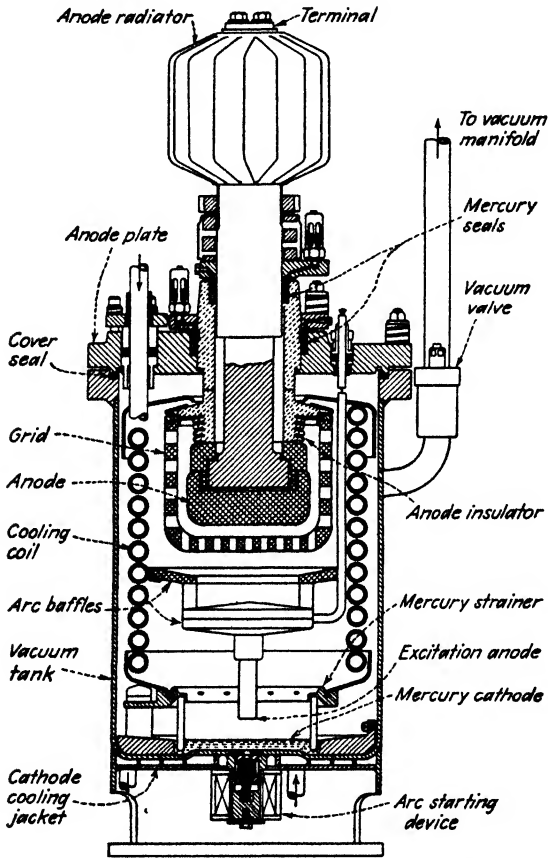


Fig. 11-16. A vertical section of an excitron. (*Courtesy of Allis-Chalmers Mfg. Co.*)

anode, which is connected in a d-c circuit. Control of the current in the main anode circuit is effected by the grid, precisely as in the thyratron. The tube possesses the control characteristics of a positive-type thyratron but likewise possesses the current capabilities of the pool cathode.

The ignition in these tubes is accomplished by means of a solenoid-operated mercury-spray splash device. The unit is placed in operation by the mercury spray, which is thrown upward to the excitation anode.

As the mercury falls away from the excitation anode, the auxiliary arc is struck.

Excitrons are designed for high-current service and are used in banks as the individual elements of a polyphase rectifier system. The photograph of Fig. 11-17 shows a six-tank excitron rectifier assembly with evacuating and cooling equipment for use in heavy electric traction work.



FIG. 11-17. A pump-evacuated six-tank excitron group. This rectifier assembly is designed for heavy-duty railway service and is rated at 1,000 kw, 600 volts. (*Courtesy of Allis-Chalmers Mfg. Co.*)

Excitrons are made in both sealed-off and pumped types by the Allis-Chalmers Mfg. Co. The largest sealed excitron made by them in 1950 carried a 400-amp d-c continuous current rating. The largest unit assembly, comprising a 12-tank rectifier for railway service, somewhat like the six-tank assembly illustrated, was rated at 3,000 kw, 600 volts.

**11-8. The Ignitron.**[7] In distinction to the excitron, the *ignitron* is a single-anode pool tank which is provided with a starting ignitor electrode which initiates the arc when it is extinguished. A drawing of such a unit is given in Fig. 11-18.

With an a-c potential applied between the anode and the pool, the arc would become extinguished once each alternate half cycle, provided that

the arc could be initiated regularly. It is the duty of the starting ignitor to provide this regular ignition once each alternate half cycle. The starting ignitor is made of a refractory material (among them being silicon carbide,
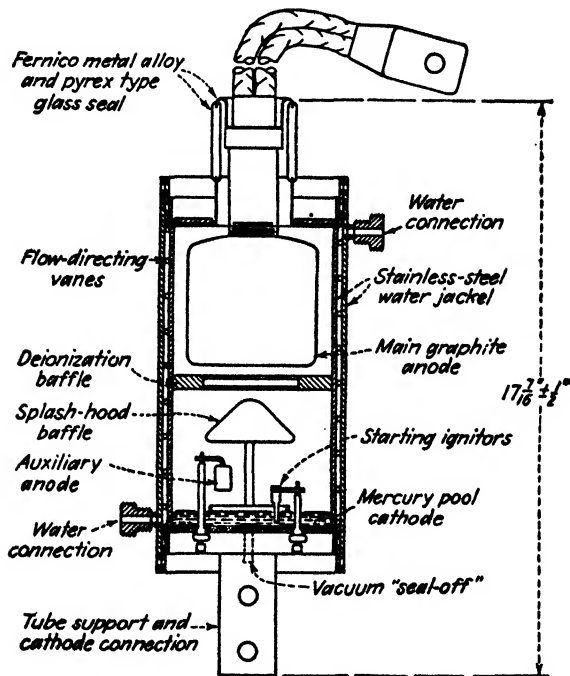


FIG. 11-18. A sealed ignitron for power rectifier service. (*Courtesy of General Electric Co.*) The type FG-238B has the following ratings:

Maximum anode current:

| | |
|---|---|
| Instantaneous | 1,200 amp |
| Continuous | 150 amp |
| 2 hr average current over any 2-min period | 225 amp |
| 1 min average current over any 1-min period | 300 amp |
| Surge, maximum duration 0.15 sec | 9,000 amp |
| Tube drop | 12.6–19.1 volts |
| Maximum inverse and forward voltage | 2,100 volts |
| Ignitor current: | |
| Maximum instantaneous allowed | 100 amp |
| Maximum average allowed | 2.0 amp |
| Time of averaging | 10 sec |
| Maximum ignition time | 100 μsec |

boron carbide, and carborundum), which projects into the mercury-pool cathode, as seen in Fig. 11-18. The ignitor has a cold resistance of 20 to 100 ohms, which, under operating conditions, decreases to 2 to 10 ohms.

There is no generally accepted theory for the mechanism of the starting. One theory [7] presumes that ignition of the arc occurs when the potential gradient in the ignitor rod in contact with the mercury surface exceeds a critical value of from $5 \times 10^3$ to $2 \times 10^4$ volts/m. The necessary potential gradient in the rod is established by passing an auxiliary current of 10 to 30 amp through it. Another theory [8] presumes that a spark occurs at the junction of the ignitor and the pool when the current which passes through the ignitor rod exceeds a critical value. Still another theory [9] considers the ignition to be a consequence of a thermal action at the junction between the ignitor and the pool. The most likely explanation is that the ignitor operates because of high field emission near the mercury-ignitor junction.

Whatever may be the actual mechanism of the ignition, once ignition begins, the tiny spark expands into an arc between the cathode and the anode. This process is accomplished in short times, usually measured in microseconds.[10] If the point in the cycle at which the current passes through the ignitor rod is controlled, then the output of the tube is likewise controlled.

Ignitrons may have two [11] or three ignitors, only one of these being used at any time. Owing to the high surge currents through the ignitor, they gradually deteriorate. With several available as stand-bys, the life of the tube is extended.

Some ignitrons are equipped with an *auxiliary* or *holding* or *excitation anode*. This auxiliary anode serves two functions. It has been found that for steady operation, especially at relatively light loads, the current surge through the ignitor should be maintained for an appreciable portion of the cycle. However, in order to prolong the life of the ignitor rods, the ignition-current pulse duration should be made as short as possible, so that heating of the ignitor rod is small. The use of the auxiliary anode permits a short excitation pulse and then serves to maintain the arc spot for the remainder of the main-anode cycle. A second reason for the auxiliary anode is that it provides another arc-conduction circuit, which, because of its relatively close spacing to the cathode, will usually conduct even in those rare instances when the main arc circuit might not fire. Owing to the usual electrical connections (see Fig. 12-41) that are used for ignitor excitation, the duty on the ignitor rod and the rectifier in the firing circuit is likely to be very severe during a misfire. Some protection from this surge is afforded by the auxiliary-anode shunt path.

Some ignitrons are provided with a grid which may be used to perform a phase-shift control function in much the same way that the grid performs in a thyratron. Thus it can be used for controlling the point in the cycle at which the tube will conduct, assuming that an arc has previously been established to the auxiliary anode by the ignitor rod. However, it has been found that the ignitor often permits control over a larger portion

of the cycle than does the grid.  The single grid which surrounds the anode and is sometimes called the *anode shield* or *shield grid* also serves to reduce the tendency for arc back.

If very high potentials are to be employed, then several grids are used. These aid in deionization, and better control characteristics are obtained. Another most important function of the grids is to allow a distribution of the potential across the tube *during the inverse cycle*.  This reduces the potential gradient, with a corresponding higher total sparking potential. The three-grid (pentode) ignitron [12] shown in Fig. 11-19 is designed for rectification or inversion service at 20,000 volts at a continuous current of 200 amp.

The ignitron has three principal applications: heavy-duty (above 25 amp) power rectification, inversion, and control switching of a-c currents, as in resistance welding.  The first two are continuous-duty operation, and the third is an intermittent application.

There is a fundamental difference between the control action in a thyratron or a grid pool tube and the action of the ignitor rod in an ignitron.  In thyratrons and grid pool tubes, the grid *prevents* the formation of an arc, whereas the ignitor *initiates* the arc.  In the former case the electrons already exist in the tube (owing to the presence of an externally heated cathode or of an arc to the keep-alive electrode), but the grid electrostatically prevents them from reaching the anode until a critical voltage is reached.  In the ignitron, on the other hand, the tube is in a nonconducting state until the ignitor circuit is energized, when conduction is forced.

**11-9. Grid Pool Tanks.**  If in a multianode mercury-pool tank each of the anodes is completely surrounded by a grid, then such a system will possess characteristics that are somewhat like those of a positive-control thyratron.  Mercury-arc rectifiers equipped with these grids are known as *grid pool tanks*.

The arc will be picked up by an anode of such a unit only when both the anode is sufficiently positive to maintain the arc and the potential of the grid surrounding that anode is greater than the critical grid starting characteristic.  This requires, of course, that an arc must be maintained in the tank either to an active anode or to an auxiliary keep-alive electrode.[6]  This device is essentially the multianode version of the excitron, although it predates the excitron.  It has been largely superseded by the ignitron and the excitron.

**11-10. Glow Tubes.**  A glow tube is a cold-cathode gas-discharge diode. It operates in the normal-glow-discharge region and so is characterized by a fairly high tube drop and a low current-carrying capacity.  The voltage drop across the tube over the operating range is fairly constant and independent of the current.  When connected in a circuit, a current-limiting resistor must be used if serious damage to the tube is to be avoided.  The
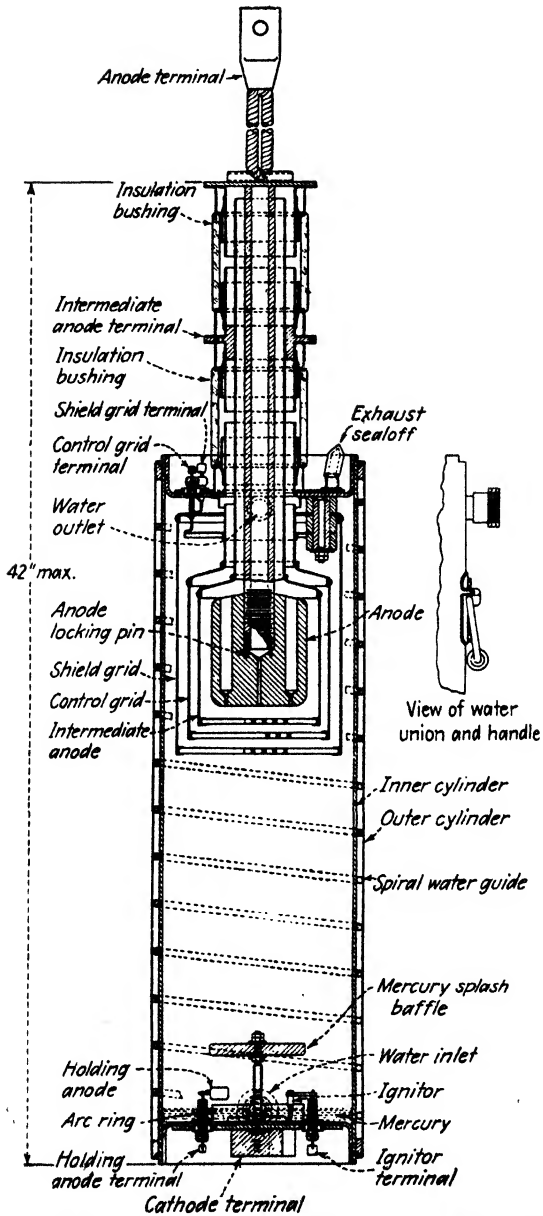
FIG. 11-19. A vertical section of GE pentode ignitron, type GL 506. (*Courtesy of General Electric Co.*)

substantially constant tube drops of such tubes are illustrated in the oscillogram of Fig. 11-20.

One commercial type of tube consists of a central anode wire which is coaxial with a cylindrical cathode, as shown in Fig. 11-21. The electrodes are of nickel, the inner surface of the cathode being oxide-coated. In order to lower the cathode fall of some tubes, some misch metal (an alloy of cerium, lanthanum, and didymium) is sputtered on the cathode. The gases commonly used are neon, argon, and helium. The tubes containing neon or helium usually contain a small amount (of the order of 1 per cent) of argon. The presence of the argon lowers the starting voltage (see Sec.
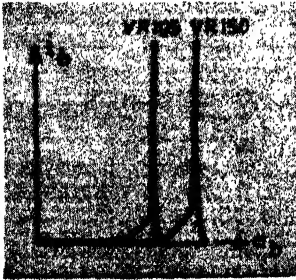


Fig. 11-20. Volt-ampere characteristics of the VR-105 and VR-150 glow tubes.
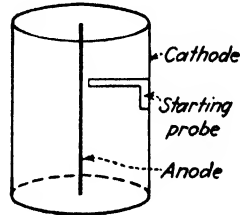


Fig. 11-21. Electrode structure in a glow tube.

9-12). The cathode fall in the normal-glow region of the discharge is determined solely by the material of the cathode and by the type of gas. Hence, the type of gas and the cathode material are the only considerations in the design of a glow tube to give a definite voltage. To date, no combination giving a stable cathode fall of less than about 60 volts has been found. For example, the types VR-75, VR-90, VR-105, and VR-150 tubes have normal output voltages of 75, 90, 105, and 150 volts, respectively, with a normal current range of from 5 to 40 ma.

The current rating of these tubes is determined by the area of the cathode and by the pressure of the gas. It should be recalled that the normal current density in the normal-glow region remains substantially constant and that the area of the cathode glow increases in proportion to the current demanded by the circuit. Hence, the operation will still occur in the normal region for a sufficiently large cathode area even though the gas pressure be varied appreciably. As a result, cleanup of the gas in these tubes plays a relatively unimportant role. However, the voltage characteristics of the tubes do vary somewhat with age mainly because of the deterioration of the cathode caused by sputtering. This effect is greatly accelerated if the tubes are overloaded.

It is observed in Fig. 11-21 that a piece of metal is attached to the cathode and is directed toward and near to the anode. It is the function of this "starting probe" to lower the breakdown voltage of the tube. Sparking takes place first to the starting probe and then spreads to the rest of the cathode. Note that the cathode is photosensitive and that the breakdown voltage will depend to some extent upon the illumination falling on the cathode.

The principal use of these glow tubes is as voltage regulators. Since the voltage remains constant (within a few per cent) over the region of the normal glow, then the output voltage will remain substantially independent of the current over this range. For example, the voltage across the load in the circuit of Fig. 11-22 will be 150 volts over a range of currents from 5 to 40 ma if a VR-150 tube is used. The difference between the supply voltage and the operating tube volt drop will appear across the resistor $R$.
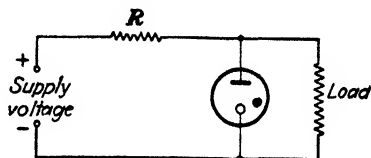


FIG. 11-22. The use of a glow tube as a voltage regulator to maintain approximately constant voltage across a load.

Glow tubes may be connected in series in order to provide a constant voltage that is the sum of the tube drops of the tubes that are used. Thus, the use of a VR-150 and a VR-105 tube in series will provide a constant 255-volt source. The supply voltage must be greater than the breakdown voltage of the tubes in order to make operation possible.

A glow lamp may be connected as in Fig. 11-12 (with the grid circuit omitted) to produce oscillations of the saw-tooth variety.

Glow lamps may also be used as rectifier tubes, although this application is not very extensive.[13] It depends upon the fact that the cathode is a specially prepared surface having a rather large area, whereas the anode wire is normally pure metal and of small area. As a result the electronic flow within the tube will be much greater from the cathode to the anode than in the reverse direction. This will result in unequal currents in each half cycle upon the application of a sinusoidal wave.

Glow tubes having smaller ratings than those considered above are available. Several such tubes having ratings ranging from $\frac{1}{4}$ to 3 watts are illustrated in Fig. 11-23. Tubes with ratings as low as $\frac{1}{15}$ watt are available. These smaller lamps are used extensively as test lamps.[14] For example, they will indicate circuit continuity; by brilliancy, whether the circuit potential is 110 volts or higher; by flicker, whether the circuit frequency is 25 cps or higher; by glow, which terminal is positive, since the negative electrode is covered with glow on a d-c source. They are also used to indicate the presence of a high-intensity radio-frequency field,

Fig. 11-23.   Photograph of several low-capacity glow lamps.   (*Courtesy of Lamp Depart-ment of General Electric Co.*)

since the gas will glow in such a field.   The tube need not be connected conductively into the circuit.   It is operating not as a normal glow tube but rather as an electrodeless discharge (see Sec. 10-20).   The tubes may also be used as pilot lights to indicate that a circuit is energized, and as a dim-light source for exit markers, for fire-station markers, or for location markers in general.
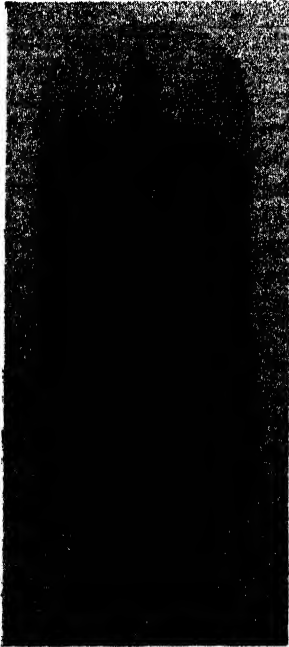


Fig. 11-24.   WE type 313C cold-cathode triode.   (*Courtesy of Western Electric Co.*)

**11-11. Cold-cathode Triodes.**[15]   Figure 11-24 shows a Western Electric 313C cold-cathode gas-filled tube containing three elements.   The cathode is a nickel surface coated with a barium and strontium mixture.   The anode is a pure nickel wire that is shielded from the other electrodes by a glass sleeve.   The bare end of this wire protrudes from the glass sleeve about $\frac{1}{2}$ in. from the other electrodes.   The third element, known as the *control anode*, or *starter anode*, is placed close to the cathode.   In the tube shown, it is identical with the cathode and placed adjacent to and coplanar with it.   The control element may be different from the cathode in structure.   In the RCA 0A4-G tube, it is a wire loop placed parallel to a plane cathode.   In other cases, it may be a metallic shield with a hole punched in it which is placed between the cathode and the anode.   This tube is referred to as a "grid glow tube," although the third element is more appropriately named the "control elec-

trode" rather than the "grid." The WE 313C contains a mixture of 99 per cent neon and 1 per cent argon at a pressure of about 40 mm Hg.

The spacing of the electrodes is such that a discharge takes place from the cathode to the control electrode at a lower voltage than is required for a discharge from the cathode to the anode. Once the control gap has been broken down, however, it is possible for the discharge to transfer to the
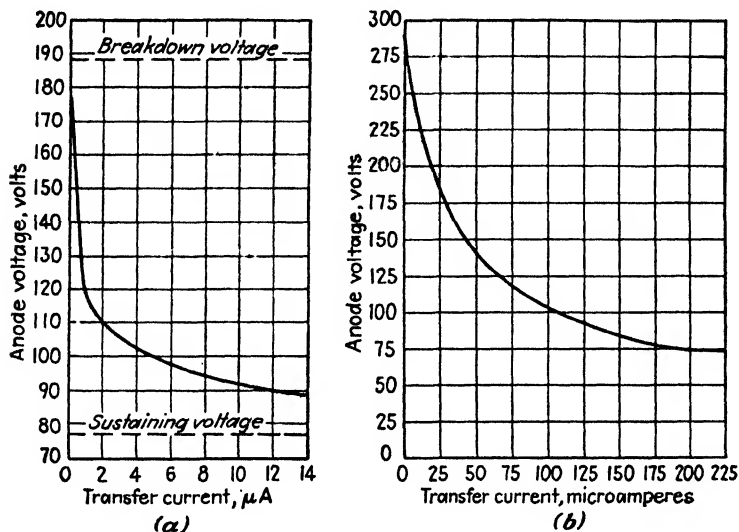


Fig. 11-25. Transfer characteristic of a cold-cathode triode. (a) WE 313C; (b) RCA OA4-G. Control-gap-maintaining voltage is 60 volts and main-gap-maintaining voltage is 75 volts for either tube.

main anode. The anode-cathode voltage that is required for this transfer to occur is a function of the transfer current, *i.e.*, the current in the cathode to control-electrode circuit. These features are evident from an inspection of the curves of Figs. 11-25*a* and *b*. Such a curve is called the "transfer" or "transition" characteristic.

For zero transfer current, which means that the control electrode is not connected in the circuit, the anode voltage is equal to the breakdown voltage between the cathode and the anode. It is noted that the required anode-cathode voltage falls rapidly as the transfer current, which provides more ionization, is increased. Regardless of the magnitude of the transfer current, however, the anode-cathode voltage can never fall below the maintaining voltage for this gap. Hence, the transfer characteristic approaches this sustaining voltage asymptotically. The transfer characteristic depends upon the type of gas, the gas pressure, and the electrode structure.

The use of the tube as a relay is illustrated in Fig. 11-26. The plate supply voltage $E_{bb2}$ is greater than the main-gap sustaining voltage but less than the main-gap breakdown voltage. The voltage $E_{bb1}$ is less than the control-gap breakdown voltage. It is now possible for a positive pulse to raise the control-anode voltage above the breakdown point so that a discharge will occur between these two electrodes. If the transfer current is sufficiently high so that $E_{bb2}$ will cause breakdown, then the discharge will transfer to the anode circuit. In this way a small input-circuit current is able to control several milliamperes in the load circuit.
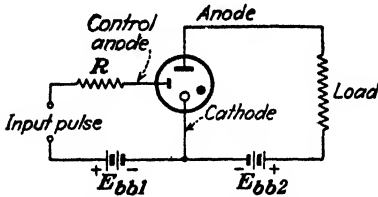


FIG. 11-26.   A cold-cathode triode as a relay.

Here, as in the case of thyratrons, once conduction between the cathode and anode has begun, the discharge cannot be controlled or extinguished by means of the control electrode. It is necessary that the anode-cathode voltage be lowered below the sustaining value for a time that is long enough to permit deionization to take place. The deionization time of these tubes is of the order of 10 millisec.

In view of the foregoing, the three-element glow tube is seen to possess many features in common with the thyratron, the excitron, and the ignitron. All tubes contain control electrodes (the control anode, the grid, or the ignitor). The discharge between the cathode and the anode cannot take place unless the control electrode is properly energized. Thus, the transfer characteristic of the glow tube is the counterpart of the critical grid breakdown characteristic of the thyratron. Further, once the discharge has been initiated between the cathode and the anode, the control element in each becomes ineffective in controlling or extinguishing the discharge. Control can be regained only if the anode voltage is reduced below the sustaining voltage for a time long enough to permit deionization to take place.

Despite the features common to these tubes, the fields of application are quite different because of the different current capacities of the tubes. The maximum current from a glow tube is approximately 100 ma, whereas that from the thyratron is about 100 amp and that from the ignitron and excitron is about 10,000 amp. The field of application of the cold-cathode triode is similar to that of the small argon-filled hot-cathode thyratron (types 884 and 2050). It possesses the advantage, however, that no cathode heating power is necessary. Among the applications in which these tubes have been used, the most important are as rectifiers in selective-ringing telephone circuits; as voltage regulators, providing two different output voltages, the anode-cathode drop or the control electrode-cathode voltage; as relaxation oscillators; and as relays or switching units.

**11-12. Current Ratings.** Gas- or vapor-filled rectifier tubes of all kinds are given *average* current ratings rather than rms current ratings. This rating specifies the maximum current that the unit may carry continuously without excessive heating of any of the parts. The time over which this average is to be taken is also specified by the manufacturer. Furthermore, units are also given peak current ratings, specifying the maximum current that the tube should be permitted to reach in each conducting cycle and averaged over various time intervals. For example, the FG-238B ignitron has the ratings listed in the caption of Fig. 11-18.

If a tube provided with a thermionic cathode is employed, several heating components must be dissipated. In addition to the constant heating power of the cathode, the tube must dissipate an instantaneous power given by the product of the instantaneous anode current and the instantaneous tube voltage. Since this voltage is substantially constant and independent of the tube current, the average power is the product of this tube drop and the *average* tube current.

The grid in the thyratron is also given average and peak ratings. These ratings should not be exceeded, for otherwise the glass press through which the leads pass may be ruptured.

**11-13. Voltage Ratings.** The function of a rectifier is to act as a synchronous switch which permits current to pass in one direction only. That is, electrons must flow only from the cathode to the anode while the anode is positive with respect to the cathode. It is necessary, therefore, that the anode remain insulating during that portion of the voltage cycle when it is negative with



FIG. 11-27. Tube drop and peak inverse voltage of an 886 gas diode as a function of temperature. (*H. C. Steiner and H. T. Maser, Proc. IRE*, **18**, 67, 1930.)

respect to the cathode. The "maximum inverse peak voltage" specifies the largest safe instantaneous value of this negative voltage. The necessity for knowing this quantity will be made evident in the next chapter when rectifying circuits are studied.

By neglecting the possibility of flashover between the electrodes through the air, conduction in the inverse direction occurs because of breakdown of the gas in the region between the cathode and the anode. This so-called "flash-back" voltage, which is nothing more than the sparking potential of the gas according to Paschen's law, is shown for an 866 tube in Fig. 11-27. For the pressures and spacing usually employed in arc discharge tubes of the externally heated types, only the portions of the sparking-potential curve to the left of the minimum sparking value comes into consideration.

The "maximum peak forward voltage" is a quantity that is significant only for gas tubes controlled by means of a grid. It specifies the largest positive potential that may be applied to the anode before the grid loses its arc-initiation ability. That is, if the plate voltage exceeds this value, breakdown will first occur between the grid and the anode. The glow that results immediately changes to an arc between the cathode and the anode. Of course, once the arc forms, a sheath forms around the grid and it becomes ineffective.

**11-14. Temperature Ratings.** The condensed-mercury temperature limits are specified for the safe and efficient operation of mercury-vapor tubes. The range usually extends from about 30° to 80°C, corresponding to vapor pressure of approximately 0.003 to 0.08 mm. The upper temperature limit is determined by the peak inverse voltage, since this voltage limit may fall below a safe value at the higher temperatures (see Fig. 11-27). The lower temperature limit is determined by several factors: As the pressure decreases, the tube drop increases (see Fig. 11-27), so that the power loss in the rectifier increases. This results in a decreased efficiency. In addition, the high tube drop may result in serious cathode disintegration.

It is desirable, therefore, especially in the large metal-tank rectifiers, that the mercury-condensation temperature be maintained in the proper range. For this reason, water cooling of the tank is usually employed.

**11-15. Deionization Time.**[16] Once the arc has been initiated in a thyratron, the grid loses its control feature, as already explained. Suppose that the anode potential is momentarily removed and is then quickly reapplied. The grid is supposed to be highly negative during this experiment. Will the tube again conduct, or will the grid, because of its large negative value, regain control and so prevent the arc from reigniting? The answer to this question is found by examining the conditions at the grid.

The grid loses control after conduction starts because of the formation of the positive-ion sheath around it, which effectively shields the electrostatic field of the grid and so prevents this field from altering conditions in the plasma. If, therefore, the positive ions in the sheath have had time enough to diffuse to the grid and recombine with electrons to form neutral molecules before the anode voltage is reapplied, then the grid regains control (see Sec. 9-13 for a discussion of recombination). The minimum time of this process is called the *deionization time.*

The deionization time $t_d$ depends upon many factors. As the gas pressure increases, $t_d$ increases, since the diffusion takes place at a slower rate. This is also affected by the increased probability of additional ion formation by electronic collisions. Also, large values of anode current result in large values of $t_d$, since the number of ions to be swept out of the sheath is correspondingly large. However, small deionization times are favored by small electrode spacings and by the presence of large surfaces at which re-

combination may occur. Furthermore, highly negative potentials on the inverse cycle favor small deionization times.

From this discussion, it is clear that the deionization time depends upon both the tube and the circuit in which the tube is used. For commercial tubes operated under rated conditions, it varies between 10 and 1,000 μsec. This time is so short that satisfactory operation of thyratrons on 60-cycle mains results. However, the deionization time may offer a serious limitation to the use of such tubes in applications at higher frequencies.

The *ionization time* is the time required for conduction to be established once the anode potential has been applied. It seldom exceeds 10 μsec and so can generally be neglected.

**11-16. Arc Back.** In multianode rectifiers, failure in the inverse direction frequently occurs because an arc may form between anodes. This unwanted initiation of an arc to an idle anode is known as *arc back, flash back*, or *backfire*. The elimination of this is one of the factors in the development of satisfactory mercury-tank rectifiers. Arc back under these circumstances results in an internal short circuit between anodes, and so of the power transformers, and may result in permanent injury to the tube unless protective devices are employed. In single-anode rectifiers, conduction in the inverse direction between the cathode and anode is likewise called "backfire." It is of interest to note that the number of backfires per unit per month are lower for a multianode rectifier than for ignitrons.[17]

The exact nature of arc-back phenomena is not completely understood. Though it had been generally accepted that arc backs occurred near the beginning of the inverse cycle, since a high concentration of positive ions exists in the region of the anode at these times, experiments have shown [18] that they may occur at any part of the inverse cycle. Also, they seem to occur at random times. That is to say, no arc backs may occur in a rectifier under normal operating conditions for some time, and then several may occur at relatively closely separated times.

The use of grids that surround each of the anodes is frequently employed in order to reduce arc back. The presence of these grids serves several useful purposes. They shield the anodes, thus preventing droplets of mercury from condensing on the anodes. They also reduce the deionization time by furnishing large surfaces for recombination. Further, the presence of the grid affects the voltage distribution in such a way as to prevent strong fields from acting on the anode surfaces. Although the use of grids enclosing the anodes does discourage arc back, it also produces an increased arc drop. This increased arc drop causes a decreased operating efficiency.

In the case of low-capacity thermionically heated arc tubes, as, for example, the types 82 and 83 tubes, each directly heated filament is completely enclosed by its respective anode, so that the arc is locally confined.

Since the plasma is not common to several anodes, the probability of arc
back is considerably reduced.

**11-17. Industrial Power Tubes.** It may be well to compare the indus-
trial power tubes, *viz.*, phanotrons, thyratrons, and tubes with pool-type
tubes. Tubes with pool cathodes require no cathode heating power and
no cathode heating time. Consequently, these units are immediately
ready for service. Large average currents (hundreds or thousands of
amperes) and extremely high peak currents are possible without any harm-
ful effects on the tube. There is no cathode sputtering as in the heater
tubes. There is, therefore, no deposition of cathode emitting material on
the anodes to increase the probability of arc back. However, the possi-
bility of mercury droplets forming on the anode and thereby causing back-
fires does exist. With proper design, however, this can be minimized.

For uncontrolled rectification, phanotrons are used. If control is de-
sired, then thyratrons are used when the average current requirement is
below about 25 amp. If the current exceeds this value pool-type tubes are
employed.

The single-anode metal sealed-off tank rectifier unit is largely supersed-
ing the multianode tank. This is principally because of the following:

1. Each unit is complete and requires no auxiliary pumping apparatus.
In the smaller sizes, only air cooling is necessary, although the larger units
are provided with a water jacket for water cooling. This type of construc-
tion is used for pool tubes with keep-alive electrodes [6] (excitron rectifiers)
as well as for ignitrons (see Figs. 11-16, 11-18, and 11-19).

2. If a polyphase system is desired, a number of single-anode tanks are
required. Since each anode is in a separate chamber, the cathode-anode
distance can be made smaller than is possible in a multianode tank with a
common pool. This reduces the length of the positive column, with a con-
sequent smaller tube drop, which produces a more efficient unit.

3. Single-anode tanks make the maintenance problem a relatively sim-
ple one. Thus, it is necessary to hold only a single unit in reserve for pur-
poses of emergency. In the case of a multianode tank, any necessary re-
pairs on the tank would require the removal of the entire rectifier from ser-
vice for extended periods.

**11-18. Gaseous Discharges as Sources of Light.**[19] Gaseous-discharge
lamps generally possess a higher luminous efficiency than lamps of the
incandescent-filament type. As a result, they are rapidly supplementing
these in many applications. The luminous efficiencies of several represent-
ative light sources are contained in Table 11-1.

In addition to the higher luminous efficiencies, the gaseous-discharge
lamps possess features characteristic of these sources. Because of this,
various types of gaseous-discharge tubes are applied to a very diversified
field of application. In the advertising field, neon, helium, and argon signs

TABLE 11-1 *

EFFICIENCIES OF REPRESENTATIVE LIGHT SOURCES

| Source | Luminous Efficiency, Lumens/Watt |
|---|---|
| Tungsten, 100-watt bulb.................. .. ......... | 16 |
| Tungsten, 1,000 watts photoflood (short life)............ | 33 |
| Sodium, 10,000 lumens........ ... ... ...  ........ | 55 † |
| Daylight, fluorescent 30 watts......       . .. ...... | 37 † |
| Green fluorescent, 30 watts........  ... ... .. ...  .... | 75 † |
| Mercury, 100 watts, type A-H4 quartz........ ........ | 35 † |
| Mercury, 1,000 watts, type A-H6 water-cooled quartz.... | 65 † |
| Mercury, 400 watts, type A-H1 glass...  ............. | 40 † |
| Low-pressure mercury, Cooper Hewitt.. ..   ...  . . . | 15–19 |

* G. A. FREEMAN, *Elec. Eng.*, **59**, 444, 1940.
† The efficiency shown is that of the lamp proper, the over-all efficiency of the lamp and auxiliary equipment is slightly lower.

are very common; sodium-vapor lamps are making considerable headway in the field of highway illumination; high-intensity mercury-vapor lamps are used for industrial lighting and floodlighting, as the intense point sources of light for projection work, for searchlights, for locomotive headlights, and for photoprinting; and fluorescent lighting is a common source of illumination for indoor lighting.

1. *"Neon" Signs.*[20] These are glow lamps in long tubular form. They contain an inert gas (not necessarily neon) at a few millimeters pressure between cold electrodes. The mechanism of the discharge is that discussed in Sec. 10-8. The cathode glow is now an insignificant part of the light-emitting region of the discharge, and most of the luminosity comes from the long plasma, or positive column, in the tube. If the tube is filled with neon, the glow is red-orange. If helium is used, the glow is yellow. And if mercury (mixed with argon and neon) is used, the glow is blue. Various other distinctive colors are obtained by using colored glass tubing. For example, a mercury-argon mixture in a yellow glass tube gives a green light.

Such tubes require a high voltage for breakdown but need only moderate voltages to maintain the discharge. It is for this reason that luminous-sign transformers are provided with magnetic leakage shunts so as to provide poor regulation. In this way, high voltages are made available at light load, and low voltages result after the discharge occurs. Under normal conditions of operation, these luminous signs require 100 to 200 volts/ft, depending upon the dimensions of the glass tubing used. A typical transformer supplies 10,000 volts on open circuit and 30 ma on short circuit.

2. *Sodium-vapor Lamps.*[21] These lamps are provided with an oxide-coated filament, and they operate, therefore, as low-voltage arcs and not as glow discharges. The bulb contains a small quantity of metallic sodium

(less than 0.5 g) which vaporizes to some extent as the temperature of the unit rises to its normal operating value (200° to 300°C). The vapor pressure of the sodium is so low, even under steady operating conditions, that neon gas is added (at a pressure of about 1.5 mm Hg) both for starting and for continuing the operation of the lamp.

To operate the lamp, the filament is first heated and then an arc is struck. At the outset of operation the color of the discharge is the distinctive red-orange that is characteristic of the neon gas. This color gives way gradually to the yellow light that is characteristic of the normal sodium discharge. The explanation of this change of color is very simple. Initially, the neon pressure is so much higher than that of sodium that most of the electrons from the hot filament collide with neon atoms. However, once the discharge is established, most of the region in the bulb will be a plasma in which a considerable number of low-energy electrons will exist. Although these electrons may make many collisions with the plentiful neon atoms, these collisions will be elastic ones because the first excitation potential of neon is 16.6 volts. Since the first excitation potential of sodium is only 2.11 volts, then the probability of sodium excitation in the plasma will be high because many electrons will possess sufficient energy for this process although they may not possess as much as 16.6 volts energy. The light emitted in the transition from the first excitation level to the normal state of sodium is $12,400/2.11 = 5,890$ A. (Spectroscopically one actually finds two lines, one having a wave length of 5,890 A, the other having a wave length of 5,896 A.) This light is in the yellow portion of the visible spectrum.

The maintenance of the arc is still dependent upon the ionization of the neon atoms in the cathode fall even when the lamp heats up to its temperature of most efficient operation. The vapor pressure of sodium at these temperatures is too low to supply a sufficient number of ions to maintain the arc. However, the efficiency of production of the 5,890- and 5,896-A lines is so much greater than that for the red neon lines that the luminous output appears to the eye to be entirely in the yellow portion of the spectrum.

The sodium-vapor lamp was made possible through the discovery of a special glass that is not attacked by the sodium at the high operating temperatures of the lamp. In order to maintain this high temperature in exposed regions, it is necessary to enclose the entire lamp in a vacuum jacket or in an unsilvered Dewar flask. The operating characteristics of a typical lamp are arc current = 5 amp, arc drop = 13 volts, filament current = 9.6 amp, filament drop = 1.8 volts. This represents a total power input to the lamp of 82 watts. The luminous efficiency is approximately 45 to 50 lumens/watt, so that the total output flux is approximately

4,000 lumens. This output depends rather critically upon the operating temperature.

3. *Mercury-vapor Lamps.*[22] The luminous efficiency of a low-pressure mercury-vapor lamp is approximately the same as that of an incandescent-filament source. However, as the pressure is increased, the luminous efficiency also increases. As a consequence, the design of these lamps has been toward higher and higher pressure. For example, the 1,000-watt type A-H6 lamp operates at a pressure of 75 atm. The entire arc is concentrated in a water-cooled quartz chamber only 1 in. long and 0.1 in. in diameter. This results in a 65,000-lumen light source, having a brightness of 30,000 candles/cm$^2$, or about one-fifth that of the sun. A gas such as neon or argon at a few centimeters pressure is used to initiate the arc. The cold electrodes are thereby raised to a temperature at which they emit electrons thermionically. The high temperature causes the mercury to vaporize, and the pressure increases to 1 atm or more.

The reason for the necessity of high pressures in order to obtain a high luminous output in a mercury-vapor lamp is found by studying the mercury spectrum. The strongest line in the spectrum is the ultraviolet line, 2,537 A. As shown in the energy-level diagram of Fig. 9-2, this line is excited when an atom falls back from its 4.88-ev level to the ground state. The lines in the visible part of the spectrum arise from transitions between higher energy levels. These high-energy transitions are favored by high current density and a high pressure because, under these conditions, the cumulative processes discussed in Chap. 9 become important. Thus, many collisions of the second kind will take place, particularly with the metastable atoms. Furthermore, for sufficiently high pressures, the resonance line 2,537 A is entirely absent from the spectrum. This results from the process of self-absorption (see Sec. 9-8). As a result, some of the energy that would appear in this ultraviolet line at the low pressures is transformed (by multiple collisions) into visible radiation. In addition, because of the high temperatures at which these arc sources operate, a considerable part of the radiation appears as a continuous visible spectrum, characteristic of an incandescent solid.

In contrast with this, the sodium-vapor lamp possesses a high efficiency at low pressures because its yellow radiation originates from transitions between the 2.11-ev level and the ground state. The sodium atom is easily excited to its resonance potential by an electronic collision, and so multiple collisions are unnecessary for the efficient production of light with this source, as is the case with the mercury radiation.

4. *Fluorescent Lamps.*[23] *Fluorescence* was defined in Sec. 3-10 as the "emission of light under the influence of external excitation." In the case of the fluorescent materials used for the screens of cathode-ray tubes, the

excitation is provided by the beam of electrons when it impinges on the screen. The color of the emitted light, *i.e.*, the frequency spectrum over which the light extends, depends not only upon the fluorescent material but also upon the mode and character of the excitation. For example, zinc orthosilicate, when used as the screen material of a cathode-ray tube, emits a bright green color when bombarded with fast-moving electrons. When used as the fluorescent material in a fluorescent lamp, it emits a bright green color when excited by the 2,537-A line of a mercury discharge, and it emits a yellow color when it absorbs radiation from a neon discharge.

The fluorescent lighting units consist generally of a coated emitter and an anode that have been sealed in the opposite ends of a glass tube 6 to 60 in. long in one type tube, or 42 to 96 in. long in another type tube, that contains mercury at low pressure. Tubes for use on a-c lines are frequently provided with emitters at each end of the glass tube, these acting alternately as cathode and anode upon potential reversal. The inner wall of these lamps is covered with a thin layer of fluorescent material. The thickness of this layer must be sufficient to absorb most of the impinging excitation radiation, although it must not be so thick that it will absorb any appreciable amount of its own light. Excitation is generally provided by the ultraviolet light contained in the spectrum of the mercury discharge. The fluorescent materials used in these lamps are particularly sensitive to the middle ultraviolet region (2,000 to 3,000 A).

In order to start a fluorescent tube, it is necessary to strike an arc in the discharge tube. This requires the application of a high voltage across electrodes. These high voltages are obtained either by the use of a step-up transformer or by means of an inductive surge produced by the associated starting device. Starting is accomplished in the usual household lighting unit by one of the circuits shown in Fig. 11-28. Refer to Fig. 11-28*a*. When the switch *S* is closed, the line current passes through the ballast unit *L* (a small inductor) and the filaments *F*. After a short time has elapsed during which the filaments have become heated, the switch *S* is opened. The interrupted current through *L* generates a high voltage across it, and in consequence a high voltage appears across the tube. Once the arc has been struck, the positive-ion bombardment of the emitter will maintain its temperature and continued operation results.

Figure 11-28*b* provides for automatic starting. The starter is a rather interesting combination of a glow-discharge tube (which operates in the abnormal-glow region) and a bimetallic element. When the unit is switched on, the starter glow discharge begins. The bimetallic element, which is part of the glow electrode, heats, so that the contacts close and the glow discharge is extinguished. This connects the filaments into the circuit, and at the same time the bimetallic element cools. After a short time, the

bimetallic unit contacts open and thereby produce the high voltage across the tube, as with the manual start. Once the lamp discharge starts, the voltage across it is not high enough to restrike the glow in the starter and it remains out.

In Fig. 11-28c is illustrated the so-called "cold-cathode" tube, which is operated from a high-voltage high-magnetic-leakage transformer without the use of a starter. The proper tube voltage is set automatically by the regulation of the transformer. The emitting surface is a large-area iron thimble-type cathode which attains a temperature of 150°C during operation.

The so-called "slimline" lamp is a hot-cathode instant-start tube. The emitter is a coated filament which is not heated externally, but is maintained at a thermionic temperature by positive ion bombardment.

Under operating conditions a constant voltage ranging from 60 to 110 volts exists across the tube, depending upon the length. This voltage drop consists of three parts, the cathode fall, the anode fall, and the fall in the positive column. The voltages at the anode and the cathode depend upon the particular type of lamp, the cold-cathode having a much higher drop than the hot-cathode tube. The gradient in the plasma is uniform, and so the voltage drop in this region is directly proportional to the tube length.

The luminous efficiency of fluorescent lamps exceeds that of the usual incandescent lamp. Furthermore, this highly efficient lighting occurs with the lamps operating at a relatively low temperature. Because the operat-



(a)-MANUAL START

(b)-AUTOMATIC START

(c)-COLD CATHODE, HIGH VOLTAGE START

Fig. 11-28. Fluorescent lamp starting by (a) manual start, (b) automatic start, (c) high-voltage start. L is the ballast unit, F are the filaments, S is the manual switch, GS is the glow switch, B is the bimetallic strip.

ing temperature of these units is about 40°C, these tubes have been said to produce "cold" light. The luminous efficiency depends very markedly on the operating temperature, being very low for either low or high tempera-

tures.   If the temperature is too low, an insufficient amount of mercury vaporizes and the vapor pressure is low.   At high temperatures the vapor pressure is above normal, and a lower percentage of radiation of wave length 2,537 A is produced in the Hg vapor.   This results in a lowered light output since, as noted above, it is this line which is responsible for most of the fluorescence.

When fluorescent tubes are used on a-c lines, the discharge reverses periodically because of the alternating potential.   This produces a stroboscopic action, or flicker, which might prove to be annoying when moving objects are present, or dangerous to life and limb when rotating machinery is present.   In order to overcome this stroboscopic action, luminaries containing two tubes are built.   The two tubes are excited by currents having a phase displacement by the use of appropriate circuit elements in the external circuit.   In this way the high point in the light cycle of one lamp occurs near the low point of the other lamp, and the flicker in the light output is greatly reduced.

The electrons in the plasma of the discharge of a fluorescent lamp perform oscillations of high frequency.[24]   This causes radio-frequency interference, which can be suppressed, however, by using radio-frequency filters in the lines.

The more common fluorescent materials that are used for lighting units, together with certain characteristics thereof, are contained in Table 11-2.   Cadmium borate, which gives a red color, is also used.

TABLE 11-2 *

FEATURES OF FLUORESCENT LIGHTING UNITS

| Gas or vapor used | Luminous material | Color | Luminous efficiency, lumens/watt |
|---|---|---|---|
| Mercury (with A and/or Ne for start) | None | Pale blue | 5 |
| | Zinc orthosilicate | Bright green | 50–60 |
| | Zinc mesodisilicate | Yellow-white | 25 |
| | Zinc-beryllium silicate (various grades) | Pink to cream white | 30–40 |
| | Calcium tungstate | Deep blue | 15 |
| | Magnesium tungstate | Light blue | 30 |
| Neon | None | Orange-red | 15 |
| | Zinc orthosilicate | Yellow | 22 |
| | Calcium tungstate | Pink | 15 |

* G. G. Isaacs, *IEEJ*, **84**, 286, 1939.

## PROBLEMS

**11-1.** Plot a curve of concentration in a mercury-vapor tube vs. mercury condensation temperature.

**11-2.** What fraction of the electrons in a mercury diode will collide with mercury molecules in their passage from the cathode to the anode, a distance of 2 cm, if the mercury condensation temperature is

    *a.* 10°C?

    *b.* 45°C?

    *c.* 100°C?

Take the molecular radius of mercury as 1.82 A.

**11-3.** A certain high-pressure argon-filled diode contains a thoriated-tungsten filament that operates at 2200°K under no load. The filament input is 18 amp at 2.2 volts. The rated average current is 6 amp, and the tube drop is 8 volts. If rated load is suddenly applied, a darkening of the filament is visibly observable. Why?

Calculate the change in temperature of the filament from no load to full load (see Prob. 6-15).

Explain clearly any approximations made in your solution.

**11-4.** *a.* A Western Electric 313C gas tube contains neon at a pressure of 40 mm Hg and room temperature. If the volume of the tube is 25 cm³, what is the mass of the gas contained in the tube?

*b.* The mercury condensation temperature in a General Electric FG-27A gas tube is 40°C. If the volume of the tube is 300 cm³, what is the mass of the mercury *vapor* in the tube?

**11-5.** An 884 argon-filled triode is to be used as a sweep-circuit relaxation oscillator in the circuit of Fig. 11-12. If $E_{cc} = -25$ volts, $E_{bb} = 400$ volts, $C = 0.0025$ μf, $R = 1.0$ megohm, and the extinction voltage is 16 volts,

    *a.* Calculate the frequency of oscillation.

    *b.* Calculate the peak amplitude of the generated oscillations.

    *c.* If the peak current rating of the tube is 0.5 amp, what is a suitable value for $R'$?

    *d.* Plot the wave shape of the sweep voltage, assuming that the capacitor discharges through the tube in zero time.

Assume that Fig. 11-10 represents the critical grid characteristic of the tube.

**11-6.** An 885 thyratron is used as a relaxation oscillator in the circuit of Fig. 11-12. The critical grid curve for the tube is that given in Fig. 11-10. The extinction voltage is 15 volts. If the maximum deviation from linearity is not to exceed 5 per cent, design a 60-cycle sweep whose amplitude is 25 volts. Specify reasonable values of $E_{bb}$, $E_{cc}$, $R$, and $C$, and give reasons for your choice.

**11-7.** The 884 thyratron is to operate as a saw-tooth oscillator with an amplitude of 100 volts at a frequency of 1,000 cps.
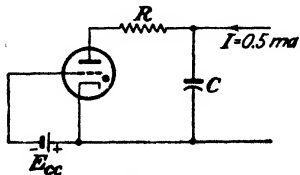


PROB. 11-7.

If the maintaining voltage is 15 volts find

*a.* $E_{cc}$.

*b.* $C$.

*c.* If the peak current rating of the tube is 0.5 amp, what is a suitable value for $R$?

**11-8.** Given two cylindrical glow-discharge tubes having cathode dimensions as shown.



PROB. 11-8.

Both tubes have the same cathode material. The current range of $T1$ is 5 to 50 ma.
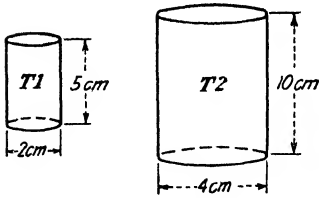
*a.* If the tubes are filled with the same gas at the same pressure, what is the maximum current rating of $T2$?

*b.* Suppose that the pressure in $T2$ is twice that in $T1$, what is the corresponding maximum current rating of $T2$?

**11-9.** *a.* A VR-150 and a VR-90 tube are to be used in series to give a constant output voltage of 240 volts. The supply voltage is 300 volts. What is the value of the current-limiting resistor that must be used if the nominal load current is 25 ma and the nominal tube current is 15 ma?

*b.* If the load suddenly decreases by 5 ma, what will be the tube current?

*c.* The load is as in (*a*). If the supply voltage suddenly changes to 310 volts, what will be the tube current?

**11-10.** *a.* A VR-105 voltage-regulator tube is used in the circuit of Fig. 11-22. The load fluctuates between 40 and 60 ma. If the supply voltage remains constant at 245 volts, find the value of $R$ so that the load voltage is maintained fixed at 105 volts. Assume that the normal operating range of the tube is 5 to 40 ma.

*b.* With $R$ set as in part *a* and with the load fixed at 50 ma, find the range over which the supply voltage may vary without affecting the output voltage.

**11-11.** A VR-105 is used in the circuit shown as a simple relaxation oscillator. $R$ and $C$ are to be chosen to yield a recurrence frequency of 1,000 cps.

*a.* Calculate and plot the wave shape of the output, when $E_{bb} = 125$ volts. (Choose the starting voltage as 115 volts, extinction voltage at 103 volts.)

*b.* Repeat when $E_{bb} = 400$ volts.

*c.* Discuss the matter of linearity on the basis of these results. In particular, determine the voltage excursion over which the saw tooth is linear to within 5 per cent in each case.
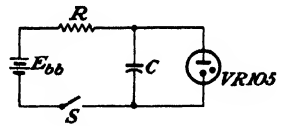


PROB. 11-11.

**11-12.** Show that the period of oscillation of a relaxation oscillator of the type illustrated in Fig. 11-12 and Prob. 11-11 is

$$T = RC \ln\left(\frac{E_{bb} - E_e}{E_{bb} - E_d}\right)$$

where $E_d$ is the voltage at which the tube starts to conduct and $E_e$ is the extinction voltage.

**11-13.** *a.* A WE 313C tube is used in the circuit of Fig. 11-26 with $E_{bb2} = 100$ volts and $E_{bb1} = 80$ volts. What is the largest value of grid-limiting resistor for which the current will transfer to the main anode? Assume that there is no input pulse.

*b.* What is the value of the load resistance needed to obtain full rated load current (10 ma)?

*c.* Repeat parts *a* and *b* for an RCA 0A4-G tube for which the rated load current is 25 ma.

**11-14.** A WE 313C tube is to be used in the circuit of Fig. 11-26. The resistor in the control anode circuit is 2 megohms, and the load consists of a 1,500-ohm relay. Assume that there is no input pulse.

*a.* What is the maximum value of $E_{bb2}$ for which the load current will not exceed the rated value of 10 ma?

*b.* What is the minimum value of $E_{bb1}$ in order that the tube operate properly?

*c.* Repeat parts *a* and *b* for the RCA 0A4-G tube for which the rated load current is 25 ma.

**11-15.** The circuit shown is a photocell-operated warning circuit:

*a.* What is the minimum photocell current which will operate the buzzer?

*b.* What is the maximum voltage across the photocell?

*c.* What is the maximum voltage across the buzzer coil?

*d.* During what fraction of a cycle does current flow through the buzzer if the photocell current is 50 $\mu$a?



PROB. 11-15.

**11-16.** *a.* In the neon spectrum, there are a number of levels grouped between 16.6 and 16.8 volts and another group between 18.3 and 18.9 volts. In what part of the spectrum are the lines emitted by transitions between these two groups of levels?

*b.* Other prominent lines in the spectrum originate on the 20.7-volt level and end on levels in the 18.3- to 18.9-volt group. To what spectral region do the emitted photons belong?

It is these transitions which are responsible for the characteristic color of a neon sign.

Explain why cold neon vapor is transparent to the red lines which give a neon sign its brilliance.
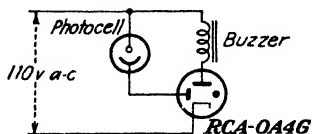
## REFERENCES

1. WHEATCROFT, E. L. E., "Gaseous Electrical Conductors," p. 189, Oxford University Press, New York, 1938.
   LANGMUIR, I., *Phys. Rev.*, **33**, 954, 1929.
2. HULL, A. W., *Gen. Elec. Rev.*, **12**, 622, 1932; *Trans. AIEE*, **47**, 753, 1928.
3. HULL, A. W., *Trans. AIEE*, **47**, 753, 1928; *Gen. Elec. Rev.*, **32**, 213, 390, 1929.
4. REICH, H. J., "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.
   HENNEY, K., "Electron Tubes in Industry," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937.
5. LIVINGSTON, O. W., and H. T. MASER, *Electronics*, **7**, 114, April, 1934.
6. MARTI, O. K., *Supplement to Electrical Engineering—Transactions, Section*, p. 927, December, 1940.
   MULDER, J. G. W., *Philips Tech. Rev.*, **1**, 65, 1936.
   WINOGRAD, H., *Trans. AIEE*, **63**, 969, 1944.
7. SLEPIAN, J., and L. R. LUDWIG, *Elec. Eng.*, **52**, 605, 1933.
8. CAGE, J. M., *Gen. Elec. Rev.*, **38**, 464, 1935.
   TOEPFER, A. H., *Elec. Eng.*, **56**, 810, 1937.
9. MIERDEL, G., *Wiss. Veröffentl. Siemens-Werken*, **15**, 36, 1936.
   KLEMPERER, H., *Electronics*, **12**, 12, December, 1939.
10. DOW, W. G., and W. H. POWERS, *Trans. AIEE*, **54**, 942, 1935.
11. HERSKIND, C. C., and E. J. REMSCHEID, *ibid.*, **65**, 632, 1946.
12. STEINER, H. C., J. L. ZEHNER, and H. E. ZUVERS, *ibid.*, **63**, 693, 1944.
13. SHAW, A. E., *Proc. IRE*, **17**, 849, 1929.
14. FERREE, H. M., *Trans. AIEE*, **60**, 8, 1941.
15. INGRAM, S. B., *Elec. Eng.*, **58**, 342, 1939.
    KNOWLES, D. D., *Elec. J.*, **27**, 232, 1930.
    KNOWLES, D. D., and S. P. SASHOFF, *Electronics*, **1**, 183, 1930.

RCA *Application Note* 91.

BAHLS, W. E., and C. H. THOMAS, *Electronics*, **11**, 14, May, 1938.

16. HULL, A. W., *Gen. Elec. Rev.*, **32**, 213, 1929.

OSTENDORF, W., *Elektrotech. Z.*, **59**, 87, 1938.

17. WARD, J. W., *Elec. Eng.*, **66**, 958, 1947.

18. PAKALA, W. E., and W. B. BATTEN, *Trans. AIEE*, **59**, 345, 1940.

KINGDON, K. H., and E. J. LAWTON, *Gen. Elec. Rev.*, **42**, 474, 1939.

EVANS, R. D., and A. J. MASLIN, *Trans. AIEE*, **64**, 303, 1945.

SLEPIAN, J., and L. R. LUDWIG, *ibid.*, **61**, 92, 1942.

WHITE, J. E., *J. Applied Phys.*, **13**, 265, 1942.

19. DUSHMAN, S., *JOSA*, **27**, 1, 1937.

COTTON, H., "Electrical Discharge Lamps," Chapman & Hall, Ltd., London, 1946.

20. MILLER, S. C., and D. G. FINK, "Neon Signs," McGraw-Hill Book Company, Inc., New York, 1935.

21. GORDON, N. T., *Gen. Elec. Rev.*, **37**, 338, 1934.

FOUND, C. G., *ibid.*, **37**, 269, 1934.

BUTTOLPH, L. J., *JOSA*, **29**, 124, 1939.

HELLER, G., *Philips Tech. Rev.*, **1**, 2, 70, 1936.

22. HELLER, *loc. cit.*

FREEMAN, G. A., *Elec. Eng.*, **59**, 444, 1940.

BUTTOLPH, L. J., *ibid.*, **55**, 1174, 1936.

MARDEN, J. W., N. C. BEESE, and G. MEISTER, *ibid.*, **55**, 1186, 1936.

23. INMAN, G. E., and R. N. THAYER, *ibid.*, **57**, 245, 1938.

FONDA, G. R., *Trans. AIEE*, **57**, 677, 1938.

MARDEN, J. W., N. C. BEESE, and G. MEISTER, *Trans. Illum. Eng. Soc. (N.Y.)*, **34**, 55, 1939.

INMAN, G. E., *ibid.*, **34**, 65, 1939.

O'DAY, A. B., and R. F. CISSELL, *ibid.*, **34**, 1165, 1939.

CLAUDE, A., *Soc. Fran. Elec. Bull.*, **9**, 309, 1939.

CLEAVER, O. P., *Elec. Eng.*, **59**, 261, 1940.

24. TONKS, L., and I. LANGMUIR, *Phys. Rev.*, **33**, 195, 1929.

### General References

DAKE, H. C., and J. DEMENT: "Fluorescent Light and Its Applications," Chemical Publishing Company, Inc., Brooklyn, 1941.

Dow, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," 3d ed., McGraw-Hill Book Company, Inc., New York, 1949.

MARTI, O. K., and H. WINOGRAD: "Mercury Arc Rectifiers—Theory and Practice," McGraw-Hill Book Company, Inc., New York, 1930.

McARTHUR, E. D.: "Electronics and Electron Tubes," John Wiley & Sons, Inc., New York, 1936.

REICH, H. J.: "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.

# CHAPTER 12

# RECTIFIERS

Rectification will be considered in a restricted sense as the conversion of alternating current into unidirectional current by means of electrical devices. Any electrical device which offers a low resistance to the current in one direction but a high resistance to the current in the opposite direction possesses the desired characteristics. An ideal rectifier is one with zero resistance in the forward direction and with infinite resistance in the inverse direction. In effect, therefore, the rectifier is a synchronized switch that opens and closes the circuit as the a-c voltage that is applied changes its polarity. A number of devices possess the requisite nonlinear properties among which are high-vacuum thermionic diodes; gas-filled and vapor-filled thermionic diodes and pool-cathode mercury arcs; metallic rectifiers; and certain crystals.

The important rectifiers for power purposes may be classified into two general groups according to their inherent characteristics. The vacuum rectifier possesses an infinite resistance on the inverse cycle, as the tube will not conduct when the plate potential is negative with respect to the cathode. On the conducting portion of the cycle, the vacuum tube is characterized by an almost constant and low value of resistance. The gas and vapor rectifier also possess an infinite resistance on the inverse cycle, but as discussed in Sec. 11-1, such tubes are characterized by a substantially constant tube drop during conduction. Owing to these differences the resulting operation in a circuit is slightly different from that incorporating vacuum diodes. A detailed discussion follows.

**12-1. Single-phase Half-wave Vacuum Rectifier.** The basic circuit for half-wave rectification is shown in Fig. 12-1. Note that the diode, which has already been treated in some detail, is indicated as a two-element device. Actually, there are either three or four external leads, depending

Fig. 12-1. Basic circuit for half-wave rectification.

upon whether the tube is of the filamentary or of the cathode-heater type. These are the plate or anode, the cathode, and two heater (or filament) leads. However, it is usually sufficient to treat it as a two-terminal device,

335

as pictured in Fig. 12-1. The complete wiring diagram for the half-wave rectifier is given in Fig. 12-7. For the sake of simplicity, a pure resistance load is indicated. Also, a transformer with zero resistance and zero leakage reactance will be presumed.

The instantaneous plate current is $i_b$, and the instantaneous voltage across the diode is $e_b$, when the instantaneous transformer secondary voltage is $e$. Evidently, by Kirchhoff's voltage law,

$$e_b = e - i_b R_l \qquad (12\text{-}1)$$

where $R_l$ is the magnitude of the load resistor. This one equation is not sufficient to determine the two unknowns in this expression, $e_b$ and $i_b$. However, a second relation between these two variables is given by the static plate characteristic of the diode (see Fig. 7-5). In Fig. 12-2 there is indicated the simultaneous solution of Eq. (12-1) and the diode plate characteristic. The straight line, which is represented by Eq. (12-1), is called the "load line." The load line passes through the points

FIG. 12-2. The method of constructing the dynamic curve from the static curve and the load line.

$$i_b = 0 \qquad e_b = e$$

$$i_b = \frac{e}{R_l} \qquad e_b = 0$$

That is, the intercept with the voltage axis is $e$ and with the current axis is $e/R_l$. The slope of this line is determined, therefore, by $R_l$.

The point of intersection $A$ of the load line and the static curve gives the current $i_A$ that will flow under these conditions. This construction determines the current in the circuit when the instantaneous transformer potential is $e$. This current is plotted vertically above $e$ at point $B$ in the diagram. If the input voltage is permitted to vary, the corresponding current will vary. Clearly, the slope of the load line does not vary since $R_l$ is fixed. Thus, when the applied potential has the value $e'$, the corresponding current is $i_{A'}$. This current is plotted vertically above $e'$ at $B'$. The resulting curve $OB'B$ that is generated as $e$ varies is called the "dynamic characteristic." An oscillogram showing the static and dynamic curves of a type 81 diode for $R_l = 1,250$ ohms is shown in Fig. 12-3.

It is to be emphasized that regardless of the shape of the static characteristic or the wave form of the input voltage, the resulting wave form of the current in the output circuit can always be found graphically from the

dynamic characteristic. The manner of this construction will be made evident below.

If the static characteristic of the tube were linear, the dynamic characteristic would also be linear. Note from the construction, however, that

there is considerably less curvature in the dynamic curve than there is in the static characteristic. It will be assumed in what follows that the dynamic curve is linear. Likewise, to an approximation that is sufficiently accurate for most engineering purposes, the static curve of the diode may be replaced by a straight line through the origin $O$. The constant ratio $e_b/i_b$ for this line is called the "dynamic tube resistance" and is designated by the symbol $r_p$.

The application of a sinusoidal voltage to the circuit of Fig. 12-1 for which the dynamic curve is linear results in a series of half-sine-wave current pulses separated by

FIG. 12-3. Oscillogram showing the static and dynamic curves of a type 81 diode (for a load resistor of 1,250 ohms).

zero current regions in the input. The manner of the construction for obtaining the input is illustrated in Fig. 12-4. Suppose that the instantaneous input voltage has the value corresponding to the point $A$. The corresponding current in the external circuit is obtained from the dynamic
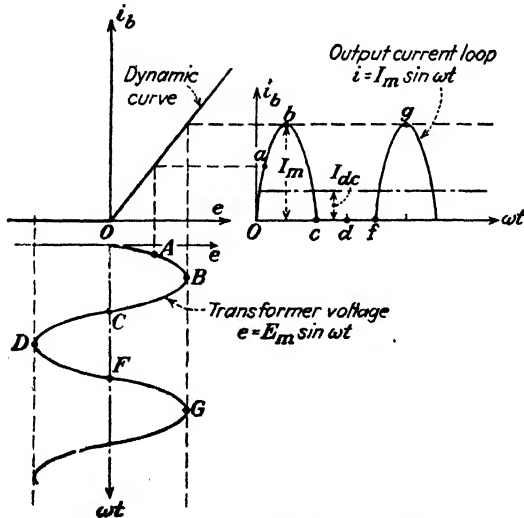


FIG. 12-4. The method of obtaining the output-current wave form from the dynamic curve for a given input-voltage wave form.

curve and is the value denoted by $a$. This "point-by-point" method permits the output-current wave form to be determined for any input-voltage wave shape, as the points $a$, $b$, $c$, ... of the current wave correspond to the points $A$, $B$, $C$, ... of the input-voltage wave.

If it is assumed that the static characteristic is linear, then the relation

$$e_b = i_b r_p$$

is valid during the conduction period. It then follows from Eq. (12-1) that

$$e = e_b + i_b R_l = i_b(r_p + R_l) = E_m \sin \omega t \qquad (12\text{-}2)$$

or

$$\left. \begin{array}{ll} i_b = \dfrac{E_m}{r_p + R_l} \sin \omega t = I_m \sin \omega t & \text{when } 0 \leq \omega t \leq \pi \\[2mm] i_b = 0 & \text{when } \pi \leq \omega t \leq 2\pi \end{array} \right\} \qquad (12\text{-}3)$$

where

$$I_m \equiv \frac{E_m}{r_p + R_l} \qquad (12\text{-}4)$$

This analysis permits one to draw an electrical circuit that is equivalent to the circuit of Fig. 12-1. This equivalent circuit is shown in Fig. 12-5. The symbol ➤ represents an ideal synchronized switch which permits current to pass through the circuit during the conduction period and which opens the circuit on the inverse cycle. The arrow points in the direction of forward current flow.



FIG. 12-5. Equivalent circuit of a single-phase half-wave rectifier.

*Example.* A d-c ammeter, an a-c ammeter, and the current coil of a wattmeter are inserted in series with the load of Fig. 12-1. The potential coil of the wattmeter is across the transformer secondary. A d-c voltmeter is placed across the diode. What do these four instruments read?

*Solution.* A d-c ammeter reads the full-cycle average current passing through it. However, by definition, the average value of a periodic function is given by the area of one cycle of the curve divided by the base. Expressed mathematically,

$$I_{dc} = \frac{1}{2\pi} \int_0^{2\pi} i_b \, d\alpha \qquad (12\text{-}5)$$

where $\alpha = \omega t$. It follows from Eq. (12-4) that

$$I_{dc} = \frac{1}{2\pi} \int_0^{\pi} I_m \sin \alpha \, d\alpha = \frac{I_m}{\pi} \qquad (12\text{-}6)$$

Note that the upper limit of the integral has been changed from $2\pi$ to $\pi$ since the instantaneous current in the interval between $\pi$ and $2\pi$ is zero and so contributes nothing to the integral.

An a-c ammeter indicates the effective or rms current passing through it. By definition, the effective or rms value squared of a periodic function of the time is given by the

area of one cycle of the curve representing the function squared divided by the base. Expressed mathematically,

$$I_{rms} = \left(\frac{1}{2\pi} \int_0^{2\pi} i_b^2 \, d\alpha\right)^{\frac{1}{2}} \tag{12-7}$$

This becomes, by use of Eq. (12-3),

$$I_{rms} = \left(\frac{1}{2\pi} \int_0^{\pi} I_m^2 \sin^2 \alpha \, d\alpha\right)^{\frac{1}{2}} = \frac{I_m}{2} \tag{12-8}$$

It should be noted that the rms value of this wave is different from the rms value of a sinusoidal wave.

A d-c voltmeter reads the average value of the voltage across its terminals. Since the voltmeter is across the tube, the instantaneous tube voltage must be plotted and the area under one cycle of this curve must be found. When the tube is conducting, it has a resistance $r_p$ and the voltage across it is $i_b r_p$. When the tube is nonconducting, the
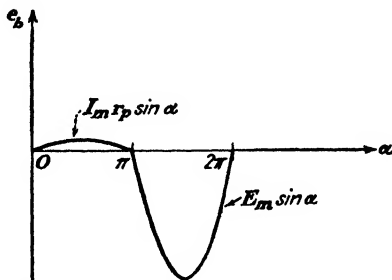


Fig. 12-6.   The tube voltage across a vacuum diode.

current is zero and from Fig. 12-1 or Eq. (12-1) it is seen that the transformer secondary voltage $e$ appears across the tube. Thus,

$$\left.\begin{array}{ll} e_b = i_b r_p = I_m r_p \sin \alpha & 0 \leq \alpha \leq \pi \\ e_b = E_m \sin \alpha & \pi \leq \alpha \leq 2\pi \end{array}\right\} \tag{12-9}$$

A plot of the tube voltage is shown in Fig. 12-6.   The reading of the voltmeter is

$$(E_{dc})_{tube} = \frac{1}{2\pi}\left(\int_0^{\pi} I_m r_p \sin \alpha \, d\alpha + \int_{\pi}^{2\pi} E_m \sin \alpha \, d\alpha\right)$$

$$= \frac{1}{\pi}(I_m r_p - E_m) = \frac{1}{\pi}[I_m r_p - I_m(r_p + R_l)]$$

where use has been made of Eq. (12-4).   Hence

$$(E_{dc})_{tube} = -\frac{I_m R_l}{\pi} \tag{12-10}$$

This result is negative, which means that if the voltmeter is to read upscale, its positive terminal must be connected to the cathode of the diode.   Since $I_{dc} = I_m/\pi$, the d-c tube voltage is seen to be equal to $-I_{dc}R_l$ or to the negative of the d-c voltage across the load resistor.   This result is evidently correct because the sum of the d-c voltages around the complete circuit must add up to zero.

It should be noted that the voltmeter reading does *not* equal the product of the d-c current $I_{dc}$ times the tube resistance $r_p$. The reason for this is that the tube is a non-linear device whose resistance is constant (and equals $r_p$) only when the plate voltage is positive. On the other hand, the d-c voltage across the load does equal the product of d-c current $I_{dc}$ times the output resistance $R_l$ because the load is a truly constant resistor.

*A wattmeter indicates the average value of the product of the instantaneous current through its current coil and the instantaneous voltage across its potential coil.* Hence the power read by the wattmeter will be

$$P_i = \frac{1}{2\pi} \int_0^{2\pi} e i_b \, d\alpha \tag{12-11}$$

This becomes, by Eq. (12-3),

$$P_i = \frac{1}{2\pi} \int_0^{2\pi} i_b^2 (r_p + R_l) \, d\alpha$$

This may be written, by virtue of Eq. (12-7), as

$$P_i = I_{\text{rms}}^2 (r_p + R_l) \tag{12-12}$$

This result could have been written down immediately by arguing physically that all the power supplied by the transformer must be used to heat the load and the tube resistance.

Equation (12-12) expresses the total power supplied to the plate circuit of the rectifier. This quantity can readily be calculated in any specific application. The total input power to the rectifier system cannot, in general, be predicted theoretically. This is because the total input power includes the transformer losses, the filament heating power, and the power losses in any auxiliary apparatus.

The above illustration indicates the general method of calculating what d-c or a-c instruments will read in any electronic circuit. It is *not* restricted to the simple diode rectifier. The wave forms, in general, may be more complicated than those of the simple diode considered above, but the method is the same. For assistance in the calculations, rough sketches of the curves are made, and the readings of the instruments are obtained from an evaluation of the area under the curve (for a d-c instrument) or the area under the squared function (for an a-c instrument). If a wattmeter reading is desired, the curve representing the current through its current coil is multiplied by the curve representing the voltage across its potential coil and the area under the product curve is then evaluated.

*The d-c power supplied to the load will be defined as the product of the reading of a d-c ammeter in the load circuit and a d-c voltmeter across the load.* Thus,

$$P_{dc} \equiv E_{dc} I_{dc} = I_{dc}^2 R_l \tag{12-13}$$

For the half-wave vacuum diode,

$$P_{dc} = \left(\frac{1}{\pi}\right)^2 \left(\frac{E_m}{r_p + R_l}\right)^2 R_l \tag{12-14}$$

It is important to note that $P_{dc}$ given in Eq. (12-14) is quite different from $P_i$ given in (12-12), or from the reading of a wattmeter placed across

the load.   In those applications which are d-c operated, as, for example, d-c machinery and electroplating, the presence of a-c components, in addition to the d-c component, may contribute to the over-all heating of the system without contributing to the useful operation.

As a measure of the efficiency of the rectification process, and also as a comparison of the various types of rectifiers to be discussed, it is convenient to define the *efficiency of rectification* (also called the "conversion efficiency" or the "theoretical efficiency") $\eta_r$ as the ratio of the d-c output power to the plate-circuit power.   This is

$$\eta_r \equiv \frac{P_{dc}}{P_i} \times 100\% = \frac{I_{dc}^2 R_l}{I_{rms}^2 (r_p + R_l)} \times 100$$

$$= \left(\frac{I_{dc}}{I_{rms}}\right)^2 \frac{100}{1 + r_p/R_l} \qquad \% \qquad (12\text{-}15)$$

By combining this with Eqs. (12-6) and (12-8), there results

$$\eta_r = \left(\frac{I_m/\pi}{I_m/2}\right)^2 \frac{100}{1 + r_p/R_l} = \frac{40.6}{1 + r_p/R_l} \qquad \% \qquad (12\text{-}16)$$

This indicates that the theoretical maximum efficiency of a single-phase half-wave circuit is 40.6 per cent.   Actually, of course, the over-all efficiency of the system, which is obtained by considering the power supplied to the entire system, is considerably less than this.

It may be easily shown that the maximum power output of such a system occurs for $R_l = r_p$.   At this load, the plate-circuit efficiency is only one-half its maximum possible value, or 20.3 per cent.

There are several features of the single-phase half-wave circuit that



Fig. 12-7.   To investigate the insulation stress in the filament transformer when one side of the load is grounded.

warrant special attention.   First, it is noticed that on the inverse cycle, *viz.*, that part of the cycle during which the tube does not conduct, the maximum potential across the tube is equal to the transformer maximum potential.   That is, the peak inverse voltage to which the tube is subjected during operation is equal to the transformer maximum value.

In order to prevent stray charges from building up on various parts of the system, it is desirable to ground either the positive or the negative side of the load.   An inspection of Fig. 12-7, with the negative terminal of the system connected to ground, indicates that the heater winding of the transformer is at a high d-c potential with respect to ground.   For exam-

ple, suppose that the output voltage is 10,000 volts. Then, the center tap of the filament-transformer secondary is 10,000 volts above ground. Since one side of the a-c power line is generally grounded, then there will be 10,000 volts between the primary and secondary windings of the filament transformer even though the primary voltage rating is only 115 volts and the secondary rating, perhaps, 5 volts. This requires that the insulation between the windings of the filament transformer be capable of withstanding this high voltage without rupture.

If the positive is grounded, then the center tap of the secondary of the filament transformer is at ground potential also. Hence, the high output voltage does not appear between the primary and secondary of the heater transformer, and it is now not necessary to use a specially built high-insulation transformer.

*Example.* Calculate the regulation and the efficiency of rectification of a type 81 diode used in the circuit of Fig. 12-1. The rated output current is 85 ma. The transformer secondary voltage is 230 volts rms. Also, find the current at which maximum power is obtained.

*Solution.* By regulation is meant the variation of d-c output voltage as a function of d-c output current. This is given by

$$E_{dc} = I_{dc}R_l$$

and, from Eqs. (12-6) and (12-4),

$$I_{dc} = \frac{I_m}{\pi} = \frac{E_m/\pi}{r_p + R_l} \tag{12-17}$$

Solve this for $I_{dc}R_l$ and combine with the first equation. The result is

$$E_{dc} = \frac{E_m}{\pi} - I_{dc}r_p \tag{12-18}$$

This result shows that $E_{dc}$ equals $E_m/\pi$ at no load and that the d-c voltage decreases linearly with the d-c output current. The larger the magnitude of the tube resistance, the greater is the rate of decrease.

From the volt-ampere curve for the type 81 tube given by the manufacturer it is found that this characteristic can be approximated by a straight line, from which a value of $r_p = 500$ ohms is calculated. Also, the peak transformer voltage is $E_m = \sqrt{2}E_{rms}$ (for a sine wave) $= (\sqrt{2})(230) = 325$ volts. Hence,

$$E_{dc} = 103.5 - 500I_{dc}$$

It follows from this expression that the d-c voltage drops from 103.5 volts at no load to 61.0 volts at the rated value of 85 ma. By definition, the percentage regulation is

$$\% \text{ regulation} = \frac{E_{\text{no load}} - E_{\text{full load}}}{E_{\text{full load}}} \times 100\%$$

If the output voltage does not vary with load, then the percentage regulation is zero. In the present illustration,

$$\% \text{ regulation} = \frac{103.5 - 61.0}{61.0} \times 100 \doteq 70\%$$

This extremely poor regulation, combined with the low efficiency and high harmonic content of the output wave, explains why half-wave high-vacuum diodes are seldom used when appreciable currents are required.

A more accurate way of estimating the effective tube resistance than that employed above using the static characteristic is to obtain a regulation plot $E_{dc}$ vs. $I_{dc}$ in the laboratory. The negative slope of the resulting straight line is $r_p$. The result so obtained includes the resistance of the transformer as well as the tube resistance. It also includes the effect of the transformer regulation.

To obtain the conversion efficiency as a function of load current, eliminate the term $R_l$ from Eq. (12-16) as was done above for the regulation. This simple substitution yields

$$\eta_r = 40.6 \left(1 - \frac{\pi r_p}{E_m} I_{dc}\right) \qquad \% \tag{12-19}$$

This shows that the plate-circuit efficiency decreases linearly with the plate current. Using the numerical values found for the type 81 tube gives

$$\eta_r = 40.6(1 - 4.82 I_{dc}) \qquad \%$$

Thus $\eta_r$ decreases from 40.6 per cent at no load to 24 per cent at full load.

The current at which maximum power is obtained from the rectifier is readily found by equating the plate-circuit efficiency $\eta_r$ to 20.3 and then solving for $I_{dc}$. This leads to the value 104 ma for this tube. Note that this exceeds the rated current. A second method of obtaining the same result is to set $R_l = r_p = 500$ in Eq. (12-17).

**12-2. Ripple Factor.** Although it is the purpose of a rectifier to convert a-c into d-c current, the simple circuit considered above does not achieve this. Nor, in fact, do any of the more complicated rectifier circuits have a truly constant output. What is accomplished is the conversion from an a-c current into a unidirectional current, periodically fluctuating components still remaining in the output wave. It is for this reason that filters are frequently used in order to decrease these a-c components. A measure of the fluctuating components is given by the *ripple factor r*, which is defined as

$$r \equiv \frac{\text{rms value of the alternating components of the wave}}{\text{average value of the wave}}$$

This may be written as

$$r \equiv \frac{I_{rms}'}{I_{dc}} = \frac{E_{rms}'}{E_{dc}} \tag{12-20}$$

where the terms $I_{rms}'$ and $E_{rms}'$ denote the rms value of the a-c components of the current and voltage, respectively.

In order to measure the ripple factor of a given rectifier system experimentally, the measurement of the ripple voltage or the ripple current in the output should be made with instruments that respond to higher than power frequencies, so that the contributions from the higher harmonic terms will be recorded. These measurements may be made using a "square-law" voltmeter of either the thermocouple or the vacuum-tube type. A

capacitor must be used in series with the input to the meter in order to "block" the d-c component. This capacitor charges up to the average value of the voltage and only the ripple components in the wave are recorded by the meter.

An analytical expression for the ripple factor, defined in Eq. (12-20), is possible. By noting that the instantaneous a-c component of current is given by

$$i' = i - I_{dc}$$

then

$$I_{rms}' \equiv \sqrt{\frac{1}{2\pi} \int_0^{2\pi} (i - I_{dc})^2 \, d\alpha} = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} (i^2 - 2I_{dc}i + I_{dc}{}^2) \, d\alpha}$$

The first term of the integral becomes simply $I_{rms}{}^2$ of the total wave. Since $\dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} i \, d\alpha$ is $I_{dc}$ by definition, then the second term under the integral sign is

$$(-2I_{dc})(I_{dc}) = -2I_{dc}{}^2$$

The rms ripple current then becomes

$$I_{rms}' = \sqrt{I_{rms}{}^2 - 2I_{dc}{}^2 + I_{dc}{}^2} = \sqrt{I_{rms}{}^2 - I_{dc}{}^2}$$

By combining these results with Eq. (12-20),

$$r = \frac{\sqrt{I_{rms}{}^2 - I_{dc}{}^2}}{I_{dc}} = \sqrt{\left(\frac{I_{rms}}{I_{dc}}\right)^2 - 1} \tag{12-21}$$

This result is independent of the current wave shape and is *not* restricted to a half-wave vacuum diode. In the case of the half-wave single-phase rectifier, the ratio

$$\frac{I_{rms}}{I_{dc}} = \frac{I_m/2}{I_m/\pi} = \frac{\pi}{2} = 1.57$$

from Eqs. (12-6) and (12-8). Hence

$$r = \sqrt{1.57^2 - 1} = 1.21 \tag{12-22}$$

This result indicates that the rms ripple voltage exceeds the d-c output voltage. This shows that the single-phase half-wave rectifier is a relatively poor device for converting a-c into d-c current.

**12-3. Single-phase Full-wave Vacuum Rectifier.** The circuit of the single-phase full-wave rectifier is shown in Fig. 12-8. This circuit is seen to comprise two half-wave circuits which are so connected that conduction takes place through one tube during one half of the power cycle and through the other tube during the second half of the power cycle.

The current to the load, which is the sum of these two currents, has the form shown in the oscillograms of Fig. 12-9. The d-c and rms values of
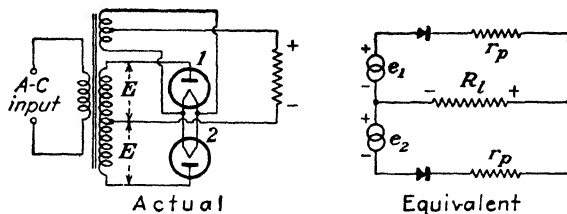


FIG. 12-8. Schematic wiring diagram and equivalent circuit of a full-wave vacuum rectifier.

the load current in such a system are readily found, from the definitions (12-5) and (12-7), to be

$$I_{dc} = \frac{2I_m}{\pi} \qquad I_{rms} = \frac{I_m}{\sqrt{2}} \qquad (12\text{-}23)$$

where $I_m$ is the maximum value of the current wave. The d-c output power is, therefore,

$$P_{dc} = I_{dc}{}^2 R_l = \left(\frac{2}{\pi}\right)^2 \frac{E_m{}^2 R_l}{(r_p + R_l)^2} \qquad (12\text{-}24)$$

It is noted, by comparing Eq. (12-23) with Eq. (12-6), that the d-c current supplied to the load for the full-wave connection is twice that for the half-
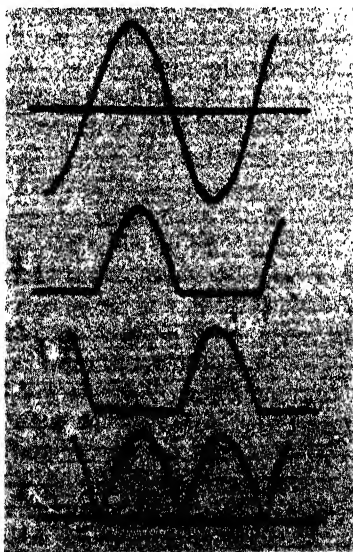


FIG. 12-9. The transformer voltage, the individual tube currents, and the load-current wave forms in a single-phase full-wave rectifier.

wave connection, whence the power delivered to the load is larger by a factor of 4 in the full-wave circuit. However, the power depends upon the circuit constants and parameters in the same way as for the half-wave circuit.

A little thought should convince the reader that the input power supplied to the plate circuit in the full-wave case is given by the same expression as for the half-wave case, *viz.*,

$$P_i = I_{rms}^2(r_p + R_l) \tag{12-25}$$

The efficiency of rectification of the rectifier is then easily found to be

$$\eta_r = \frac{P_{dc}}{P_i} \times 100 = \frac{81.2}{1 + r_p/R_l} \quad \% \tag{12-26}$$

This expression shows a theoretical maximum twice that obtained for the half-wave circuit.

The required current ratio that appears in the expression for the ripple factor is

$$\frac{I_{rms}}{I_{dc}} = \frac{I_m/\sqrt{2}}{2I_m/\pi} = 1.11$$

The ripple factor for the full-wave circuit is, from Eq. (12-21),

$$r = \sqrt{1.11^2 - 1} = 0.482 \tag{12-27}$$

A comparison of this value with the value given by Eq. (12-22) for the half-wave circuit shows that the ripple factor has dropped from 1.21 in the half-wave case to 0.482 in the present case. Clearly, therefore, the full-wave circuit gives more efficient rectification with a larger fraction of the a-c input power being converted into d-c power.

The d-c output voltage is given by

$$E_{dc} = \frac{2E_m}{\pi} - I_{dc}r_p \tag{12-28}$$

Full-wave rectification may be accomplished through the use of two separate tubes or by the use of tubes that are so constructed as to contain both elements within a single envelope, such as the type 80, 82, and 83 tubes. In either case, a center-tapped transformer must be used.

Let us consider the diagram of Fig. 12-8 from the point of view of peak inverse voltages. Suppose that tube 1 is conducting, whence tube 2 is in the nonconducting state. Except for the $i_b r_p$ drop in tube 1, the peak potential between the transformer mid-point and the cathode is $E_m$, the transformer maximum value. This is the voltage that appears across the load $R_l$. But the potential difference between the mid-point and the anode of tube 2 is also $E_m$, so that the potential between the cathode and anode

of tube 2 is $2E_m$.  Hence, the peak inverse voltage of each diode in a full-wave system is twice the transformer maximum voltage to center tap if the tube drop is neglected.

The conditions imposed on the insulation between the primary and the secondary of the filament heating transformer are the same as those which exist in the case of the half-wave circuit.  Thus, if the negative of the system is grounded, it is necessary to use a filament transformer having interwinding insulation that will withstand the full d-c potential.  If the positive of the system is grounded, this insulation stress will not exist.

**12-4.  Circuits with Gas Diodes.**  Although the general characteristics of the circuits for diode operation remain substantially unchanged when the



*(a)*                                      *(b)*
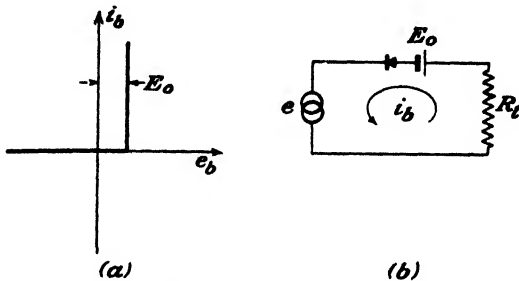
FIG. 12-10.  (a) Idealized voltage-ampere characteristic of a gas diode.  The breakdown, maintaining, and extinction voltages are all assumed to be equal to $E_0$. (b) The equivalent circuit for a half-wave rectifier using a gas diode.

high-vacuum diodes are replaced by gas-filled or vapor-filled tubes, certain differences do exist.

A gas tube will not conduct until the potential difference between the anode and the cathode reaches the breakdown value.  During conduction, a constant voltage drop will exist between the cathode and the anode.  When the anode potential falls below the extinction voltage, conduction ceases.  The situation is illustrated in Fig. 12-10; (a) gives the idealized volt-ampere characteristic and (b) gives the equivalent circuit.  The wave forms are indicated in Fig. 12-11.  Actually, the breakdown and the extinction potentials are not exactly the same, and as a result, Fig. 12-11 is only approximate.  A more exact indication of the situation is illustrated in the oscillograms of Fig. 12-12.

In oscillogram $a$, the peak alternating transformer voltage was adjusted to be $5.6\sqrt{2} \doteq 7.92$ volts, which is only slightly greater than the breakdown voltage (7.8 volts) of the particular type 83 tube used.  It is seen that conduction does not start until the transformer voltage has almost reached its peak value.  Since the extinction voltage is less than the breakdown value, conduction continues until the applied voltage falls below the
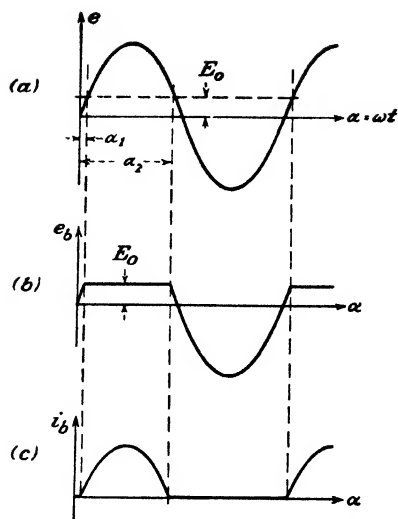
FIG. 12-11. The wave forms for a half-wave circuit using an ideal gas diode. (a) Input voltage. (b) Voltage across the tube. (c) Plate current.
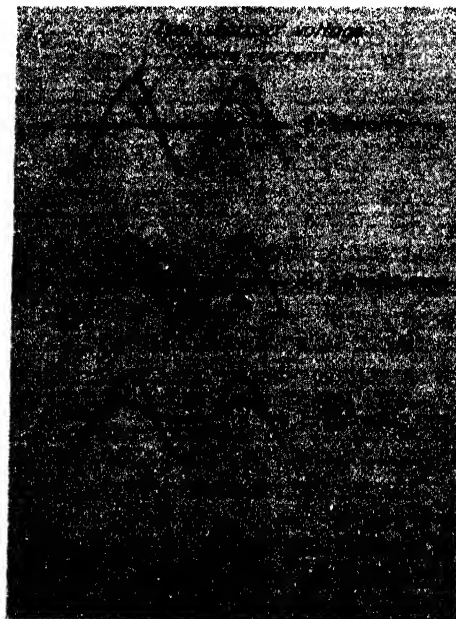


FIG. 12-12. Oscillograms which show that the point at which conduction begins in a gas-filled tube depends upon the relative values of the tube breakdown voltage and the magnitude of the applied voltage. (Type 83 tube—breakdown voltage = 7.8 volts.)

348

extinction value. This accounts for the small pulse of current that is delivered near the positive peak of each cycle. For oscillogram $b$, the peak of the applied voltage was adjusted to $9\sqrt{2} = 12.7$ volts. The angle at which conduction starts under these circumstances is obtained from the expression $7.8 = 12.7 \sin \alpha_1$, from which it is found that $\alpha_1 = 38°$. For oscillogram $c$ the peak applied voltage was $23.5\sqrt{2} = 33.3$ volts, from which it is found that the angle at which conduction starts is $\alpha_1 = 13.5°$. In oscillogram $d$, the transformer voltage was 123 volts rms, and the conduction starts almost immediately at the beginning of the cycle. Since, in most practical cases, the peak alternating voltage is very much larger than the tube drop, conduction continues for almost the entire half cycle.

A limitation to the use of gaseous diodes of the mercury-vapor type lies in the fact that the electrons in the plasma of the discharge perform oscillations of high frequencies.[1] This high-frequency disturbance may prove annoying in certain cases, although it may be suppressed by connecting radio-frequency chokes in series with each anode lead

*Example.* Calculate the regulation and efficiency of rectification of a half-wave circuit using a gas-filled rectifier tube. Assume that the constant tube drop during conduction $E_0$ is small compared with the transformer secondary peak value $E_m$.

*Solution.* The instantaneous load voltage $e_l$ equals the instantaneous transformer voltage less the constant tube drop. Hence, during conduction,

$$e_l = E_m \sin \alpha - E_0$$

The corresponding expression for the current is

$$i_b = \frac{E_m \sin \alpha - E_0}{R_l}$$

The d-c voltage is found by taking the average value of $e_l$. Thus

$$E_{dc} = \frac{1}{2\pi} \int_{\alpha_1}^{\alpha_2} (E_m \sin \alpha - E_0)\, d\alpha$$

where $\alpha_1$ is the angle at which the tube fires and $\alpha_2$ is the angle at which conduction ceases. However, if $E_m \gg E_0$, little error will be made by assuming that $\alpha_1 = 0$ and $\alpha_2 = \pi$. By changing the limits as indicated, and carrying out the indicated integration, there results

$$E_{dc} = \frac{E_m}{\pi} - \frac{E_0}{2} = \frac{E_m}{\pi}\left(1 - \frac{\pi}{2}\frac{E_0}{E_m}\right) \tag{12-29}$$

This equation does not contain the load current. This means, of course, that $E_{dc}$ remains constant, independent of the load current. Thus perfect regulation is indicated.

To calculate the efficiency of rectification, it is necessary to know the input power to the plate circuit. This is given by

$$P_i = \frac{1}{2\pi} \int_0^\pi e i_b\, d\alpha = \frac{1}{2\pi} \int_0^\pi (E_m \sin \alpha)\left(\frac{E_m \sin \alpha - E_0}{R_l}\right) d\alpha$$

where the limits have again been taken as 0 and $\pi$ instead of $\alpha_1$ and $\alpha_2$. This expression reduces to

$$P_i = \frac{E_m^2}{4R_l}\left(1 - \frac{4}{\pi}\frac{E_0}{E_m}\right) \tag{12-30}$$

The d-c power to the load is

$$P_{dc} = E_{dc}I_{dc} = \frac{E_{dc}^2}{R_l}$$

$$= \frac{E_m^2}{\pi^2 R_l}\left(1 - \frac{\pi}{2}\frac{E_0}{E_m}\right)^2 \tag{12-31}$$

The efficiency of rectification is, therefore,

$$\eta_r = \frac{P_{dc}}{P_i} = \frac{4}{\pi^2}\frac{\left(1 - \frac{\pi}{2}\frac{E_0}{E_m}\right)^2}{1 - \frac{4}{\pi}\frac{E_0}{E_m}} \tag{12-32}$$

This expression shows that $\eta_r$ is independent of the load current. If the numerator of Eq. (12-32) is divided by the denominator, and if all powers of $E_0/E_m$ higher than the first are neglected, this may be written in the form

$$\eta_r = 40.6\left(1 - 1.87\frac{E_0}{E_m}\right) \quad \% \tag{12-33}$$

To the same approximation as that of this problem, the ripple factor is given by

$$r = 1.21\left(1 + 0.5\frac{E_0}{E_m}\right) \tag{12-34}$$

The foregoing show that the result of using gas tubes in a rectifier circuit is to yield a slightly lower d-c output voltage and efficiency of rectification than the maximum possible values with vacuum diodes. However, the values are substantially constant and independent of the load current, and under normal operation these values are generally higher than actually exist when vacuum diodes are used. The ripple voltage with gas tubes is slightly higher than with vacuum diodes.

If the tube drop is very small compared with $E_m$, the foregoing expressions reduce to

$$E_{dc} = \frac{E_m}{\pi} \qquad \eta_r = 40.6\% \qquad r = 1.21$$

These are exactly the values that would be obtained from the corresponding equations for the vacuum diode, if the plate resistance $r_p$ were set equal to zero.

**12-5. Other Full-wave Circuits.** A variety of other rectifier circuits find extensive use. Among these are the bridge circuit, several voltage-doubling circuits, and a number of voltage-multiplying circuits. The bridge circuit finds application not only for power circuits but also as a rectifying system in rectifier-type a-c meters for use over a fairly wide range of frequencies.

The rectifiers for power application employ thermionic diodes of both the vacuum and gas types, whereas those for instrument use are ordinarily copper-oxide or crystal rectifiers. The essentials of the bridge circuit are shown in Fig. 12-13.

In order to understand the action of the bridge circuit, it is necessary only to note that two tubes conduct simultaneously. For example, during that portion of the cycle when the transformer polarity is that indicated in Fig. 12-13, tubes 1 and 3 are conducting, and current passes from the positive to the negative end of the load. The conduction path is shown on the
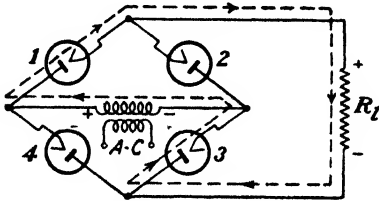


FIG. 12-13. Single-phase full-wave bridge circuit.
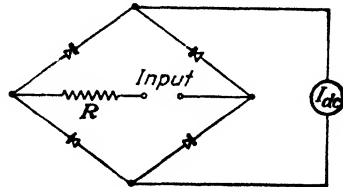
FIG. 12-14. The rectifier voltmeter. This is the single-phase full-wave bridge circuit of Fig. 12-13 with copper-oxide elements replacing the tubes, with a d-c ammeter replacing the load, and with the a-c voltage to be measured replacing the input transformer.

figure. During the next half cycle, the transformer voltage reverses its polarity, and tubes 2 and 4 send current through the load in the same direction as that during the previous half cycle.

The principal features of the bridge circuit are the following: The currents drawn in both the primary and the secondary of the supply transformer are sinusoidal, and therefore a smaller transformer may be used than for the full-wave circuit of the same output; a transformer without a center tap is used; and each tube has only transformer voltage across it on the inverse cycle. The bridge circuit is thus suitable for high-voltage applications. For example, if the output is 10,000 volts, then the peak inverse voltage across each tube is 10,000 volts. However, if a full-wave circuit were used, then the peak inverse voltage would be 20,000 volts. The transformers supplying the heaters of the tubes must be properly insulated for the high voltage.

The rectifier meter which is illustrated in Fig. 12-14 is essentially a bridge-rectifier system, except that copper-oxide elements replace the tubes, and, of course, no transformer is required. Instead, the voltage to be measured is applied through a multiplier resistor $R$, to two corners of

the bridge, a d-c milliammeter being used as an indicating instrument across the other two corners. Since the d-c milliammeter reads average values of current, the meter scale is calibrated to give rms values when a sinusoidal voltage is applied to the input terminals. As a result, this instrument will not read correctly when used with wave forms which contain appreciable harmonics.

A common voltage-doubling circuit, which delivers a d-c voltage approximately equal to twice the transformer maximum voltage at no load, is shown in Fig. 12-15. This circuit is operated by alternately charging each of the two capacitors to the transformer peak voltage $E_m$, current being continually drained from the capacitors through the load. The capacitors also act to smooth out the ripple in the output.
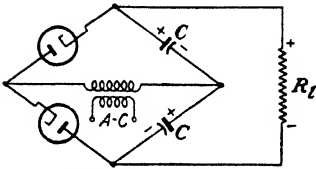


FIG. 12-15. The bridge voltage-doubler circuit. This is the single-phase full-wave bridge circuit of Fig. 12-13 with two capacitors replacing two tubes.

This circuit is characterized by poor regulation unless very large capacitors are used. The inverse voltage across the tubes during the nonconducting cycle is twice the transformer voltage. If ordinary rectifier tubes are used, two separate filament transformers or one transformer with windings well insulated from each other must be used. The latter difficulty can be avoided in low-voltage systems by using special tubes, for example, the 25Z5, which is equipped with separated, indirectly heated cathodes, each being surrounded by its own anode. The heaters are connected in series internally but are well insulated from the cathodes. The action of this circuit will be better understood after the action of the capacitor filter is studied in Chap. 13.

Circuits for obtaining $n$fold multiplication,[2] $n$ even or odd, are given in Probs. 12-15 and 12-16.

**12-6. Controlled Rectifiers.** A number of applications exist which require a controlled amount of current. These include electric welding operations, lighting control installations in theaters, motor speed control, and a variety of other industrial control applications. It is possible to vary the amount of current supplied to the load either by controlling the transformer secondary voltage or by inserting a controlling resistor in the output circuit. Neither of these methods is desirable. The first method may require expensive auxiliary equipment, and the second is characterized by poor efficiency. The development of thyratrons, ignitrons, and excitrons has made control a relatively inexpensive process.

It was pointed out in Sec. 11-4 that the presence of the massive grid structure between the cathode and the anode of the thyratron to provide complete electrostatic shielding between these electrodes permits control of the initiation of the arc by controlling the grid potential. With an

applied a-c potential to the anode, the arc is extinguished once each alternate half cycle provided that the arc is initiated regularly. The average rectified current can be varied over wide limits by controlling the point in each half cycle at which the arc initiation occurs.

In order to analyze the action of a thyratron in a controlled rectifier circuit, use is made of the critical grid breakdown characteristic of Fig.
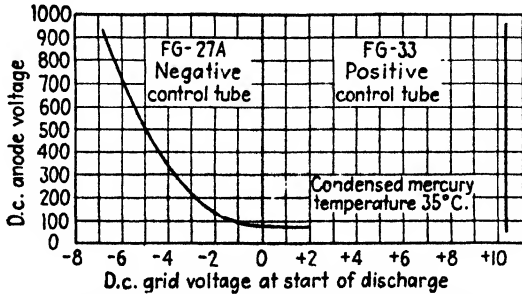


FIG. 12-16. The starting characteristics of a negative- and a positive-control thyratron.

12-16. This curve is sufficient to predict the behavior of the thyratron in the control circuit. This curve gives the minimum grid potential required for conduction to occur for each value of plate potential. Thus, if a sinusoidal plate voltage is applied to the tube, the potential of the grid just to permit conduction at each point in the cycle is found from the critical grid curve. The important curves are illustrated in Fig. 12-17. In this figure
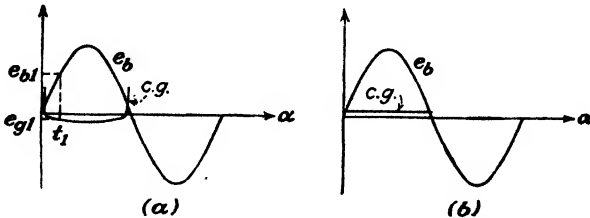


FIG. 12-17. The sinusoidal plate supply voltage and the corresponding critical grid (cg) curve for (a) a negative-control tube; (b) a positive-control tube.

the sine waves represent the plate potential as a function of time. On these same curves are drawn the critical grid (cg) voltages corresponding to these values of plate voltage. The critical grid curves for the negative-control tube in Fig. 12-17a and for the positive-control tube in Fig. 12-17b are obtained from the critical grid breakdown curves of Fig. 12-16, which are plotted for 35°C (the approximate operating temperature in air).

The critical grid breakdown curve is drawn only for positive anode voltages, since conduction does not take place if the anode is negative. For the positive tube the critical grid curve is a straight line parallel to the

time axis, indicating that the breakdown voltage is substantially independent of the anode potential. For the negative tube, a point-by-point plot is required to draw the critical grid curve, as in Fig. 12-17a. This is obtained from the critical grid starting curve as follows: Corresponding to any time $t_1$ in the positive half cycle, the plate voltage $e_{b1}$ is obtained from the impressed voltage curve. Then corresponding to this value $e_{b1}$, the critical
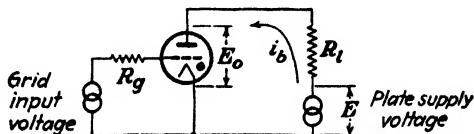


Fig. 12-18. A thyratron circuit with a-c plate and grid excitation.

grid point $e_{g1}$ (the minimum value of grid potential at which conduction will just take place) is read off the curve of Fig. 12-16. This point is plotted on Fig. 12-17a at the point corresponding to the time $t_1$. (The grid-voltage values are so small that it is difficult to draw the critical grid curve to scale.)

Suppose that the circuit is so arranged that the grid potential exceeds the critical grid breakdown value at some angle, say $\varphi$. Conduction will start at this point in the cycle. The voltage drop across the tube during
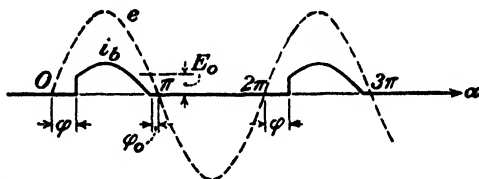


Fig. 12-19. The wave shape of the load current $i_b$ in a thyratron. Conduction starts at the angle $\varphi$ and stops at $\pi - \varphi_0$ in each cycle. The broken curve $e$ is the impressed plate voltage.

conduction of the thyratron, like that of any gas tube, remains substantially constant at a low value that is independent of the current. This tube drop is of the order of 10 to 15 volts. If the tube drop after conduction has begun is denoted by $E_0$, the current that flows through a pure resistance plate load $R_l$ during the time that the tube is conducting is given by

$$i_b = \frac{E_m \sin \omega t - E_0}{R_l} \tag{12-35}$$

where $E_m$ is the maximum value of the applied potential. This expression is seen to follow directly from the circuit of Fig. 12-18.

The resulting form of the output current is illustrated in Fig. 12-19. The breakdown or ionization time is so small [3] that it need not be taken into

consideration in most applications. This accounts for the vertical rise in the plate-current curve at the angle $\varphi$. The current is seen to rise abruptly at the point corresponding to the angle $\varphi$ and then follows the sine variation given in Eq. (12-35) until the supply voltage $e$ falls below $E_0$ at the
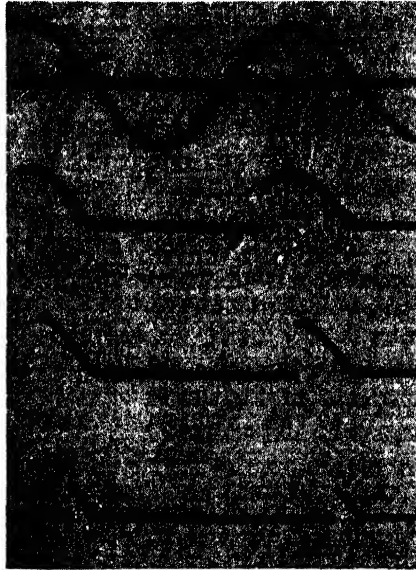


Fig. 12-20. Oscillogram of plate currents for an FG-27A thyratron. The angles at which conduction began were adjusted to approximately 135 (bottom curve), 90, and 45 deg, respectively. The top curve is the impressed plate voltage.

phase $\pi - \varphi_0$. The current will remain zero until the phase $\varphi$ is again reached in the next cycle. Such results are illustrated in the oscillograms of Fig. 12-20.

The average rectified current (the value read on a d-c ammeter) will be

$$I_{dc} = \frac{1}{2\pi} \int_{\varphi}^{\pi - \varphi_0} i_b \, d\alpha = \frac{E_m}{2\pi R_l} \int_{\varphi}^{\pi - \varphi_0} \left( \sin \alpha - \frac{E_0}{E_m} \right) d\alpha$$

which integrates to

$$I_{dc} = \frac{E_m}{2\pi R_l} \left[ \cos \varphi + \cos \varphi_0 - \frac{E_0}{E_m} (\pi - \varphi_0 - \varphi) \right] \qquad (12\text{-}36)$$

where $\alpha = \omega t$ and where $\varphi_0$ is the smallest angle defined by the relation

$$E_0 = E_m \sin \varphi_0 \qquad (12\text{-}37)$$

If the ratio $E_0/E_m$ is very small, then $\varphi_0$ may be taken as zero and Eq. (12-36) reduces to the form

$$I_{dc} \doteq \frac{E_m}{2\pi R_l}(1 + \cos \varphi) \tag{12-38}$$

The limits of variation of the angle $\varphi$ in Eq. (12-38) are from 0 to $\pi$.

This analysis shows that the average rectified current can be controlled by varying the position at which the grid potential exceeds the critical grid starting value. The maximum current is obtained when the arc is initiated at the beginning of each cycle; and the minimum current is obtained when the grid potential is applied at a point where no conduction occurs, *viz.*, at the end of the positive half cycle.

The voltages across the thyratron for the conditions illustrated in Fig. 12-19 are shown in Fig. 12-21. The applied
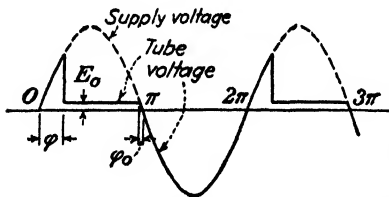


Fig. 12-21. The wave shape of the anode voltage $e_b$ of a thyratron. During conduction, between the angles $\varphi$ and $\pi - \varphi_0$, $e_b$ is a constant and equals the tube drop $E_0$. During the nonconducting portion of each cycle $e_b$ equals the plate supply voltage.



Fig. 12-22. Oscillograms of anode voltage for an FG-27A thyratron. The angles at which conduction began were adjusted to approximately 45 (top curve), 90, and 135 deg, respectively. The bottom curve is the impressed plate voltage.

voltage appears across the tube until conduction begins. After breakdown the tube voltage is a constant equal to $E_0$. When the applied voltage falls to such a low value that the tube is extinguished, then the tube voltage is again equal to the applied voltage. Figure 12-22 shows oscillograms of the tube voltage for various starting angles.

The reading of a d-c voltmeter placed across the tube will be

$$E_{dc} = \frac{1}{2\pi} \int_0^{2\pi} e_b \, d\alpha$$

$$= \frac{1}{2\pi} \left( \int_0^{\varphi} E_m \sin \alpha \, d\alpha + \int_{\varphi}^{\pi - \varphi_0} E_0 \, d\alpha + \int_{\pi - \varphi_0}^{2\pi} E_m \sin \alpha \, d\alpha \right)$$

This integrates to

$$E_{dc} = \frac{E_0}{2\pi} (\pi - \varphi_0 - \varphi) - \frac{E_m}{2\pi} (\cos \varphi + \cos \varphi_0) \qquad (12\text{-}39)$$

If $E_m \gg E_0$, this reduces to

$$E_{dc} \doteq -\frac{E_m}{2\pi} (1 + \cos \varphi) \qquad (12\text{-}40)$$

The appearance of the negative sign means that the cathode is more positive than the plate for most of the cycle. This is in agreement with Fig. 12-21. It is to be emphasized that a d-c voltmeter does not read $E_0$, if an a-c potential is applied to the plate. If the voltmeter is to read upscale, it must be connected with its positive terminal at the cathode.

It should be noted that the d-c load voltage, is the negative of the d-c tube voltage [compare Eqs. (12-38) and (12-40)]. This follows from the fact that the sum of the d-c voltages around the circuit is zero.

*Example.* A thyratron is connected according to Fig. 12-18 and supplies power to a 200-ohm resistor load from a 230-volt source of supply. If the grid voltage is adjusted so that conduction starts 60 deg after the start of each cycle, calculate the readings of the following meters: (1) an a-c ammeter in series with the load, (2) an a-c voltmeter across the tube, (3) a wattmeter inserted in the circuit so as to read the total power delivered by the a-c supply. Assume that the tube drop during conduction equals 15 volts.

*Solution.* Since $E_m \gg E_0$, little error will be made by assuming that conduction

Fig. 12-23. The wave shapes of the tube current and tube voltage in a thyratron. Conduction begins at 60 deg and (since $E_m \gg E_0$) ceases at 180 deg.

continues until the end of each positive half cycle. The instantaneous current through the tube and the voltage across the tube will have the forms shown in Fig. 12-23.

1. The instantaneous current during the conduction is

$$i_b = \frac{230\sqrt{2}\sin\alpha - 15}{200} \left.\begin{matrix}\\\\\end{matrix}\right\}$$
$$i_b = 1.625 \sin\alpha - 0.075 \left.\begin{matrix}\\\\\end{matrix}\right\} \quad \text{for } \frac{\pi}{3} \leqq \alpha \leqq \pi$$

An a-c ammeter reads the rms value of the current wave.   For the wave sketched,

$$I_{rms} = \sqrt{\frac{1}{2\pi}\int_{\pi/3}^{\pi}(1.625\sin\alpha - 0.075)^2\, d\alpha}$$

$$= \sqrt{\frac{1}{2\pi}\int_{\pi/3}^{\pi}(2.65\sin^2\alpha - 0.244\sin\alpha + 0.00564)\, d\alpha}$$

$$= \sqrt{0.533 - 0.058 + 0.002} = 0.69 \text{ amp}$$

If the tube drop is neglected, then

$$I_{rms} = \sqrt{\frac{1}{2\pi}\int_{\pi/3}^{\pi}(1.625\sin\alpha)^2\, d\alpha} = \sqrt{0.533} = 0.73 \text{ amp}$$

The limits of integration are from 60 to 180 deg, the current being zero outside of this range.

2. The a-c voltmeter reads the rms value of the voltage wave sketched.   It is noted that between 0 and $\pi/3$ the tube voltage equals the line voltage; between $\pi/3$ and $\pi$, it is constant and equal to $E_0$; and between $\pi$ and $2\pi$, it again equals the line voltage. Thus,

$$E_{rms} = \sqrt{\frac{1}{2\pi}\left[\int_0^{\pi/3}(230\sqrt{2}\sin\alpha)^2\, d\alpha + \int_{\pi/3}^{\pi}15^2\, d\alpha + \int_{\pi}^{2\pi}(230\sqrt{2}\sin\alpha)^2\, d\alpha\right]}$$
$$= 178 \text{ volts}$$

If the tube drop is neglected, then very little error is introduced.

3. The instantaneous power from the a-c supply is the product of the instantaneous line current and the instantaneous line voltage.   The wattmeter will read the average value of this product.   Hence

$$P = \frac{1}{2\pi}\int_{\pi/3}^{\pi}(1.625\sin\alpha - 0.075)(230\sqrt{2}\sin\alpha)\, d\alpha = 101 \text{ watts}$$

The integration extends only from $\pi/3$ to $\pi$, for there can be no power when the current is zero.   If the tube drop is neglected, the calculated power equals 107 watts.

**12-7. Phase-shift Control.**   In the phase-shift method of control, the point in each half cycle at which conduction will take place is varied by changing the phase angle between the plate and grid potentials.   These conditions are illustrated in Fig. 12-24.   In this figure the grid voltage $E_c$ lags the plate voltage $E_b$ by an angle $\theta$, as indicated by the sinors that represent the sine waves.   An examination of this figure shows that at the time corresponding to the phase $\varphi$ the grid voltage just equals the critical breakdown value, so that conduction starts at this point in the cycle.   The arc will be extinguished, of course, when the plate potential falls to a

value too small to maintain conduction. These conditions are illustrated in the oscillogram of Fig. 12-25.

It should be noted that the curve marked $e_b$ in Fig. 12-24 (and also in Fig. 12-17) is really the plate supply voltage $e$ and not the anode voltage.
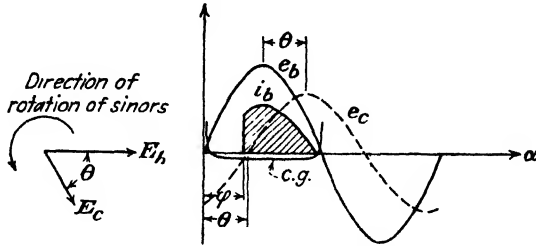


FIG. 12-24.  Phase-shift control of a thyratron.  The grid voltage $e_c$ lags the plate voltage $e_b$ by an angle $\theta$.   Conduction begins at the angle $\varphi$ at which the grid-voltage curve intersects the critical grid curve.  The sinors which generate the plate and grid voltages are indicated to the left.

However, before conduction begins in each cycle, $e = e_b$, and hence the foregoing method of determining the angle $\varphi$ at which conduction begins is valid.

If the magnitude of the applied grid voltage is large compared with the critical grid voltage (appropriate to the plate potential at any phase), the



FIG. 12-25.  Oscillogram of the applied (120-volt rms) grid and plate voltages of an FG-27A thyratron and also of the resultant plate current.  Note that conduction begins very close to the point at which the grid voltage crosses the zero voltage axis.

angle $\varphi$ is approximately equal to the angle $\theta$.   *Under these circumstances the critical grid curve may be assumed to coincide with the zero voltage axis.* Furthermore, if the maximum value of the plate voltage is much larger than the tube drop, then Eq. (12-38) will give the dependence of d-c load current upon the phase angle for all values of $\varphi$ for which the grid voltage

*lags* the plate voltage. When the grid voltage *leads* the plate voltage by *any* angle, an inspection of Fig. 12-24 reveals that conduction will occur very nearly at the beginning of each cycle. The maximum possible rectified current is obtained under these conditions. These results are illustrated in Fig. 12-26.

It is observed from this curve that the current is very small for an angle slightly less than 180 deg. For an angle slightly larger than 180 deg, the plate current suddenly rises to its full value. Since, when the grid is 180 deg leading, the tube will be on, and, when it is 180 deg lagging, the tube will be off, the 180-deg point is a critical one. This critical point may be
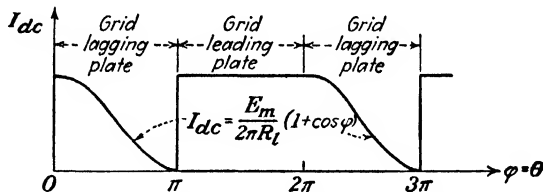


FIG. 12-26. The average load current as a function of the angle between the grid and plate voltages. The tube drop has been neglected, and the critical grid characteristic has been assumed to coincide with the zero voltage axis.

obtained in the laboratory by using a phase-shifting system that allows a shift through the 180-deg point.

For those cases where a small plate potential or a small grid potential may be employed, the generalization that $\varphi \doteq \theta$ will no longer be valid. Such cases can, however, be analyzed by drawing the plate, grid, and critical grid curves to scale in the manner shown in Fig. 12-24. Under some extreme conditions, it may be found that $\varphi$ differs considerably from $\theta$. Also, if the tube drop is not negligible compared with the peak plate voltage, the d-c plate current will be given by Eq. (12-36) and not by the simplified expression (12-38). However, the approximations introduced above, *viz.*, the neglect of the tube drop and the assumption that the critical grid curve coincides with the zero voltage axis, are valid in many practical problems.

**12-8. Phase-shifting Circuits.** A number of methods exist by which the phase between the plate and grid voltages may be varied. A direct way is to employ a polyphase phase shifter. These devices are essentially wound-rotor induction motors, generally with unity transformation ratio, the rotor of which may be set at any fixed desired angle with respect to the stator. The primary is excited with a three-phase supply. This causes a rotating magnetic field to be set up, the phase of the emfs induced in the secondary windings being a function of the angular position of the rotor. A so-called "three-phase selsyn" may be used for this purpose in exactly this way.

The more customary method of shifting the phase is through the use of simple phase-shifting networks. The phase-shifting network of Fig. 12-27 is frequently employed.[4] The cathode return $K$ is connected to the mid-point of a single-phase system. The phase between the grid and plate voltages is controlled by means of the two impedors $Z_1$ and $Z_2$, arranged as shown.

In order to analyze this phase-shifting network, it is convenient to draw a sinor diagram of this circuit. This is done for the case where

$$Z_1 = R \text{ (a resistor)}$$

and

$$Z_2 = \omega L \text{ (an inductor)}$$

The voltage notation conventions used here are explained in Appendix VIII. The reader is urged to study these carefully before proceeding further; otherwise, the diagrams which follow may not be clear to him.



FIG. 12-27. A simple network which allows the phase of the grid voltage to be shifted with respect to the phase of the plate voltage.

The first step in the analysis is to label the cathode, plate, and grid with the letters $K$, $P$, and $G$, respectively. Other letters ($a$, $b$, and $h$) are also added at the junction points in the network. Conditions are then considered *before conduction begins in each cycle*. The plate current, the grid current, and the cathode current are all zero before conduction begins, and the
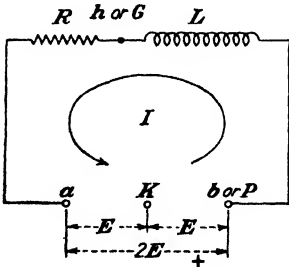


FIG. 12-28. Before conduction begins in each cycle the circuit of Fig. 12-27 reduces to this network.
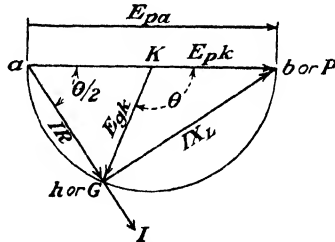


FIG. 12-29 Circle diagram for the circuit of Fig. 12-27 for conditions before conduction starts in each cycle. The diagram is used to determine the phase angle $\theta$ between the plate and grid voltages.

tube may be removed from the circuit without any effect. The simple network of Fig. 12-28 is obtained. Before conduction, there is no voltage drop in the load resistor $R_l$, and points $b$ and $P$ are at the same potential as indicated on the diagram. Similarly, $h$ and $G$ are at the same potential.

The sinor diagram is given in Fig. 12-29. The voltage drop $\mathbf{E}_{pa}$ from $P$ to $a$ is drawn along the horizontal axis. Since the cathode is connected

to the mid-point of the system, then the point $K$ is located midway between the end points $a$ and $P$ of the sinor $\mathbf{E}_{pa}$. Since the circuit is inductive, the current $\mathbf{I}$ lags the voltage $\mathbf{E}_{pa}$. The $IR$ drop from $G$ to $a$ is in phase with the current, and the $IX_L$ drop leads the current by 90 deg. The sum of these two drops must, of course, equal the line voltage drop $\mathbf{E}_{pa}$. The angle $aGP$ is a right angle. As either $R$ or $L$ is varied, the point $G$ has as its locus a semicircle whose diameter is the voltage drop $E_{pa} = 2E$ or whose radius is $E$. This is referred to as a "circle diagram."

Since the points $K$ and $G$ are already located on the diagram, then the sinor $\mathbf{E}_{gk}$ can be drawn between them. The sinor diagram shows that the grid voltage $\mathbf{E}_{gk}$ lags the plate voltage $\mathbf{E}_{pk}$ by the angle $\theta$. As either $R$ or $L$ is varied, this angle changes but the magnitudes of both the plate and the grid voltages remain constant and equal to the input voltage $E$.

With the system of Fig. 12-27, the load current will decrease as the resistance decreases, provided that $X_L$ is constant. This result is evident from the following: If $R$ decreases, $I$ increases, $IX_L$ will increase, and the angle $\theta$ increases. From Fig. 12-26, it is seen that $I_{dc}$ decreases as $\theta$ increases. The phase angle can be obtained from Fig. 12-29. The result is seen upon inspection to be

$$\tan\frac{\theta}{2} = \frac{Z_2 I}{Z_1 I} = \frac{Z_2}{Z_1} \tag{12-41}$$

where

$$Z_2 = X_L = \omega L$$

and

$$Z_1 = R$$

If $Z_1$ and $Z_2$ are interchanged, then a sinor diagram similar to the foregoing for the new conditions indicates that now $\theta$ becomes an angle of *lead* of the grid voltage over the plate voltage. For this case there can be no control over the magnitude of the rectified current, $I_{dc}$ remaining constant and independent of $\theta$, as indicated in Fig. 12-26.

A similar analysis shows that for the case where $Z_1 = 1/\omega C$ and $Z_2 = R$, the angle $\theta$ is also given by Eq. (12-41). Here with $X_C$ held constant and $R$ adjustable, the rectified plate current decreases as $R$ increases. The possible arrangements for $L$ and $C$ constant, with an adjustable $R$, are tabulated.

| $Z_1$ | $Z_2$ | Control |
|:---:|:---:|---|
| $R$ | $X_L$ | $I_{dc}$ increases as $R$ increases |
| $X_L$ | $R$ | No control; maximum conduction |
| $R$ | $X_C$ | No control; maximum conduction |
| $X_C$ | $R$ | $I_{dc}$ decreases as $R$ increases |

The quantity $R$ that appears in the phase-shift circuit need not necessarily be an ohmic resistor but may be a thermionic tube whose resistance is varied by changing one of the control voltages; it may be a photocell whose resistance varies with illumination; or it may be any other resistor. Of course, $R$ could be maintained fixed in any of these circuits, and the phase angle could be controlled by varying $L$ or $C$.

Numerous other phase-shifting networks exist. The following illustration shows another such circuit:

*Example.* Two similar capacitors $C$ and two similar resistors $R$ are employed in the phase-shifting network sketched in Fig. 12-30. Show that phase-shift control of the thyratron is obtained by varying $C$. In particular, if $R = 10^6$ ohms and $C = 0.001$ μf, what will be the d-c plate current through the 200-ohm plate load?
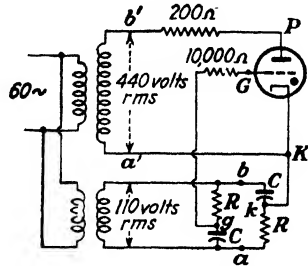


FIG. 12-30. A phase-shifting network for thyraton control.

*Solution.* Since the plate and grid voltages are large, little error is made by neglecting the tube drop and by assuming that the critical grid curve coincides with the zero voltage axis.

The first step in the analysis is to label the cathode, plate, and grid with the letters $K$, $P$, and $G$, respectively. Other letters are also added at the junction points in the network as shown. Before conduction begins in each cycle, the tube represents an open circuit and the network of Fig. 12-30 reduces to that shown in Fig. 12-31.

To construct the sinor diagram of the control network, the grid transformer secondary voltage sinor $E_{ba}$ is taken, for convenience, along the horizontal axis in Fig. 12-32.
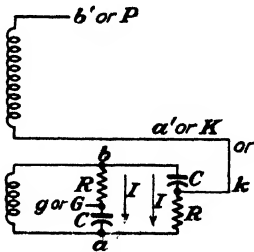


FIG. 12-31. The equivalent circuit of the network in Fig. 12-30 before conduction begins in each cycle.
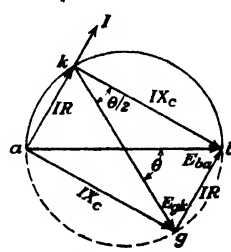


FIG. 12-32. Circle diagram for the network of Fig. 12-30.

The currents through both parallel branches $bga$ and $bka$ are equal, since each circuit contains the same elements $R$ and $C$. This current is denoted by I, and it leads the voltage $E_{ba}$, since each branch is capacitive. The voltage drop from point $k$ to point $a$ equals $IR$ and is in phase with I. Similarly, the voltage drop from $b$ to $k$ is $E_{bk} = IX_C$ and lags I by 90 deg, since $X_C$ is a capacitive reactance. Since the sinors $E_{ka}$ and $E_{bk}$ are mutually perpendicular, this locates the point $k$ on a semicircle with $E_{ba}$ as the diameter and drawn above the horizontal axis.

By employing a similar analysis, the point g is located on a semicircle with $E_{ba}$ as diameter but drawn below the horizontal axis.

This is shown by the broken semicircle in Fig. 12-32. The grid voltage $E_{gk}$ is seen to be equal in magnitude to the secondary grid transformer voltage and is therefore independent of the particular values of $R$ and $C$. The angle between $E_{gk}$ and $E_{ba}$ is $\theta$ and represents the angle between the grid and plate voltages. This follows from the fact that $E_{b'a'} = E_{pk}$ (before conduction starts in each cycle) and $E_{ba}$ is either in phase with or 180 deg out of phase with $E_{b'a'}$.

When $E_{ba}$ is 180 deg out of phase with $E_{b'a'}$, the angle $\theta$ is one of lead of grid voltage over plate voltage and so no control of the output current is possible. Hence the maximum current will flow for all values of $\theta$. If $E_{b'a'}$ is in phase with $E_{ba}$, then the current can be controlled from zero to a maximum value by varying $C$ from very large values to very small values.

It is seen from the geometry of the circle diagram that

$$\tan \frac{\theta}{2} = \frac{R}{X_C} = \omega RC$$

For the present case with $R = 10^6$ ohms, $C = 0.001$ $\mu$f, and $\omega = 120\pi$, then tan $(\theta/2) = 0.377$, from which $\theta = 41.2°$. By Eq. (12-38), the d-c current is found to be

$$I_{dc} = \frac{440\sqrt{2}}{400\pi} (1 + \cos 41.2°) = 0.87 \text{ amp}$$

**12-9. D-c Bias Control.** It is possible to control the magnitude of the average rectified current of a thyratron by varying the d-c grid bias voltage on the tube. An a-c plate supply voltage must be used. This is called the "bias-control" method. The operation of this method of control is most easily understood by referring to Fig. 12-33.

The tube will conduct at the point where $E_c$ intersects the critical grid curve, the angle $\varphi$ of the diagram. Clearly, if the negative grid voltage is too large ($E_c > E'$), the grid-voltage line will not intersect the critical grid characteristic and conduction will not be possible. The maximum negative bias is that for which the grid voltage line is tangent to the critical grid curve, and the tube conducts for one-quarter of the cycle. For less negative values of the bias voltage, the angle $\varphi$ at which conduction begins is less than 90 deg. Control is evidently possible over the range from full conduction to half conduction.
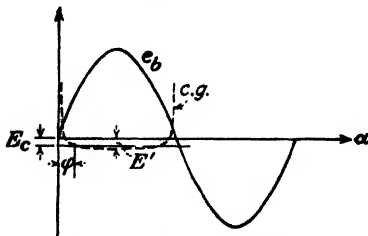


Fig. 12-33. The bias method of control. Conduction starts at the phase $\varphi$ at which the d-c grid voltage $E_c$ equals the critical grid voltage.

A curve showing the variation of the d-c plate current with bias voltage is shown in Fig. 12-34. As predicted, the plate current can be varied from a maximum value to a value about one-half of this maximum. If one

attempts to obtain smaller values of rectified current, the tube ceases to conduct and the current falls to zero, as shown. An unstable critical condition analogous to that which prevails in the phase-shift method of control for $\theta = 180°$ occurs here at the point $\varphi = 90°$.
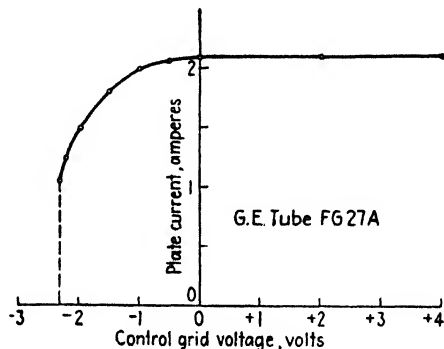


FIG. 12-34. The average plate current in a thyratron as a function of the d-c grid voltage in the bias method of control.

**12-10. Bias Phase Control.** It is a shortcoming of the d-c bias method of control discussed in Sec. 12-9 that it provides control only over the range from maximum current to approximately one-half this value. It is possible to extend the range of control from the maximum value to zero by combining a-c and d-c grid excitation. A circuit for such control is illustrated in Fig. 12-35. The $RC$ network serves to provide a fixed shift between the grid and plate voltages. Control is then obtained by varying the d-c grid bias $E_c$ either positively or negatively.

Suppose, for convenience, that $R = X_C$. The a-c grid voltage will then *lag* the plate voltage by 45 deg and will have a maximum value that is 0.707 times the maximum of the plate voltage, as can easily be seen from a simple sinor diagram. Assume also that the critical grid curve coincides with the zero voltage axis. Then, with zero bias $E_c = 0$, conduction will start at



FIG. 12-35. A circuit for bias phase control of a thyratron. The grid voltage is the sum of the a-c voltage across the capacitor $C$ and the d-c bias voltage $E_c$.

the 45-deg point in each positive half cycle, as shown in Fig. 12-36a. From Eq. (12-38) the d-c current is closely approximated by

$$I_{dc} = \frac{I'}{2}(1 + \cos \varphi) \qquad (12\text{-}42)$$
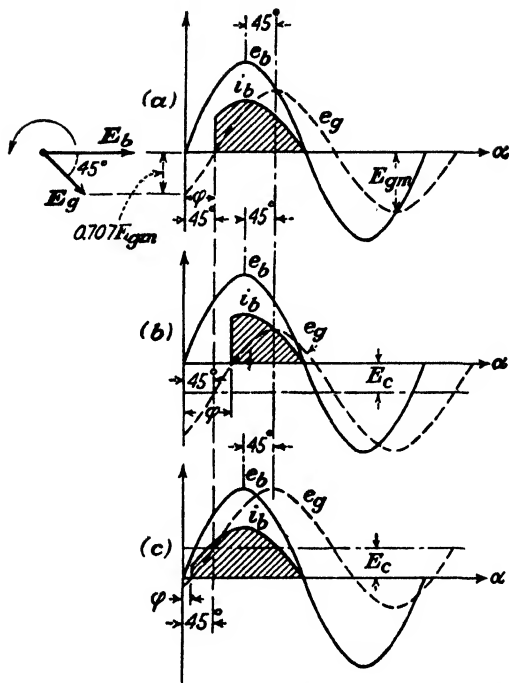
Fig. 12-36. Current variations with bias phase control. The phase angle between plate and grid voltages is kept constant at 45 deg, and the bias voltage is varied: (a) $E_c = 0$; (b) $E_c$ negative; (c) $E_c$ positive. The sinors generating the $E_b$ and $E_c$ curves are shown at the left.

where $I' = E_m/\pi R_l$ represents the maximum possible d-c current. For $\varphi = 45°$,

$$\frac{I_{dc}}{I'} = \frac{1}{2}(1 + \cos 45°) = 0.85$$

This is indicated in Fig. 12-37.

If now a negative d-c bias is applied, this is equivalent to sliding the $e_g$ curve downward with respect to the $e_b$ curve. This is indicated in Fig. 12-36b. The a-c grid voltage $e_g$ intersects the zero voltage axis farther along in the cycle, and the plate current is decreased. It can be seen from the diagram that the minimum rectified current will exist when the negative bias $E_c$ equals the peak a-c grid voltage $E_{gm}$. Conduction will then start at an angle $\varphi = 135°$. Under this condition, $I_{dc}/I' = \frac{1}{2}(1 + \cos 135°) = 0.15$. If the d-c bias is made more negative than this value, no conduction is possible.

If the bias is made positive, this is equivalent to sliding the $e_g$ curve upward with respect to the $e_b$ curve. Conduction will start at some angle that is smaller than 45 deg. This is shown in Fig. 12-36c. An inspection of this diagram shows that when the d-c bias equals 0.707 times the peak grid alternating voltage ($E_c = 0.707E_{gm}$), conduction will take place over the full half cycle and maximum current will be delivered.

A plot of $I_{dc}/I'$ as a function of $E_c/E_{gm}$ is given in Fig. 12-37. Different shapes of control curves of this type are obtained for each phase-shift angle. An angle of lead results in a quite different control characteristic from the same angle of lag. The method of obtaining these control curves
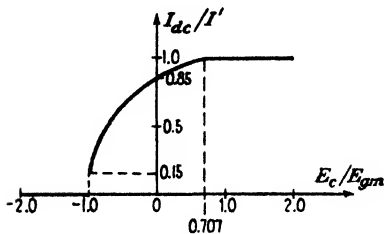


FIG. 12-37. The control characteristic of the network in Fig. 12-35.
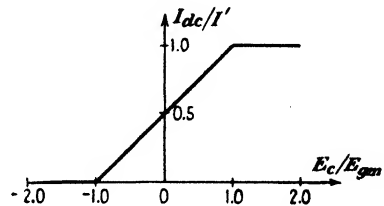
FIG. 12-38. Linear control is obtained with the bias phase method if the grid voltage lags the plate voltage by 90 deg.

is summarized as follows: The sinor $\mathbf{E}_b$ is drawn horizontally, and the sinor $\mathbf{E}_c$ is drawn at the given angle with respect to it. Then the $e_b$ and $e_g$ curves are drawn as the vertical projections of the corresponding sinors, as they are imagined to rotate in a counterclockwise direction. This ensures the proper orientation of the a-c grid voltage with respect to the plate voltage. The entire analysis may be invalid unless this is done carefully. The $e_g$ curve is now moved upward for a positive d-c bias and downward for a negative d-c bias. The angle at which the $e_g$ curve intersects the zero voltage axis is noted for each value of bias, and the corresponding d-c current is calculated from Eq. (12-42).

One special case is of importance. If the phase-shift angle is 90 deg with the grid voltage lagging the plate voltage, then linear control from zero to maximum current is obtained as illustrated in Fig. 12-38 The proof of this statement is left to the reader (Prob. 12-30e).

**12-11. On-off Control.** A variety of circuits exist which permit on-off control of a thyratron. Such circuits would be used when it is desired to use the thyratron as an arcless switch or contactor.

Consider a positive-grid thyratron which operates with a-c current on the plate and with a d-c bias on the grid. The critical grid characteristic is a straight line, as can be seen from Fig. 12-16. If the grid voltage is more

positive than this critical grid voltage, conduction takes place. If the grid voltage is less than this critical value, the tube does not conduct. Thus the tube acts as a power switch, with the grid controlling the "on" or "off" status.

There are other instances when continuous control is not obtained, only an on-off control being exercised. Consider, for example, the phase-shift control circuits of Sec. 12-8. It is seen from Fig. 12-26 that if the circuit parameters are so adjusted that the phase shift between the grid and plate potentials is in the neighborhood of 180 deg either full rectified current or zero current is obtained.
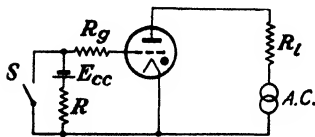


FIG. 12-39. An on-off thyratron control circuit. With the switch $S$ open the tube does not co..duct. With $S$ closed the tube conducts for approximately 180 deg in each cycle.

Another example of on-off control is illustrated in the circuit of Fig. 12-39. The switch $S$ serves as an arcless contactor to control considerable power through the load $R_l$. With $S$ open, no conduction occurs since $E_{cc}$ is so adjusted that it is more negative than the maximum negative critical grid value (the voltage $E'$ of Fig. 13-33). When the switch $S$ is closed, the grid is tied to the cathode and approximately maximum rectified current is delivered. The resistor $R$ serves to prevent short-circuiting the battery $E_{cc}$ when $S$ is closed. The resistor $R_g$ is large enough to keep the grid current after breakdown within safe limits.

**12-12. D-c Operation of Thyratrons.** Since the grid provides a means of initiating the discharge but not of extinguishing it, the plate voltage must be reduced below the extinction voltage before conduction will cease. In those circuits considered above, the application of an a-c plate voltage resulted in arc extinction once each cycle. If the thyratron is operated with a d-c voltage applied to the plate, then for control, it is necessary that some means be provided for extinguishing the arc. This can be accomplished manually by opening a switch in the plate circuit, by means of a relay in the plate circuit, or by suitable electrical means.

A variety of electrical methods exist for extinguishing the arc in a thyratron. In one method, a series $LC$ circuit is connected between the plate and the cathode  With such an oscillatory circuit, the output when the tube fires will be a pulse of current that tends to oscillate from a positive to a negative value. Since the tube cannot pass a reverse current, the arc will be extinguished. This method has been used to permit voltage pulses applied to the grid to be recorded on a mechanical counter in the plate circuit.

A second method, which is used extensively for the generation of rectangular pulses of relatively short duration, is quite like that just described

except that an open-circuited artificial transmission line replaces the $LC$ circuit.[5]

Many other electrical circuits exist which effectively open the plate circuit.[6] In one such circuit one thyratron is used as a means for extinguishing the arc in a second thyratron (see Prob. 12-43).

**12-13. Control of Ignitrons.** The general features of the ignitron are discussed in Sec. 11-8. In this tube, the arc is established once each cycle of the applied power by means of a high-current surge to the ignitor circuit. Since ignition may be established at any point in the cycle, then control of the average anode current is possible.

The fundamental circuit of the ignitron for simple rectification without phase control is illustrated in Fig. 12-40. The wave forms of the anode voltage, the anode current, and the ignitor current are also shown. Conduction occurs if the anode potential is sufficiently positive to maintain the arc that results from the ignition spark. After conduction begins, the anode-cathode voltage across the ignitron is the arc drop (about 10 to 20 volts). If this is less than the maintaining voltage of the diode, the current through the ignitor rod falls to zero. These conditions are illustrated in the figure. If the drop across the ignitron is greater than the diode maintaining voltage, the ignitor-rod current is not reduced to zero but falls to a small value.



Fig. 12-40. A simple ignitron circuit and the wave forms of the anode voltage, anode current, and ignitor current.

The series rectifier serves two purposes, to limit the ignitor current after ignition on the positive half cycle to a small value (or zero) and to prevent inverse currents on the negative half cycle. The instantaneous power required by the ignitor is relatively high, although the total energy involved is small, owing to the short time interval during which the ignitor operates. Hence, the average ignitor power is small.

The instantaneous power required for ignition is considerably higher than that required by the grid circuit of a thyratron, and the general methods of control of thyratrons are not applicable to ignitron control. A common method of control utilizes a thyratron in the circuit. In this method, which is illustrated in Fig. 12-41, the point of ignition in each

cycle is controlled by controlling the breakdown point of the thyratron, which is in series with the ignitor rod. The grid-control circuit of the thyratron may be any of those studied above, *viz.*, a phase shifter, a phase-shifting network, etc. It should be noted that in this circuit, the ignitor current passes through the load. This may not be desirable in some applications and can be avoided by using the circuit of Fig. 12-42. In this circuit the thyratron plate current, which is also the ignitor current, is
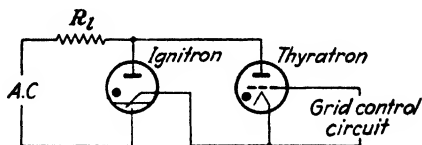


FIG. 12-41. An ignitron controlled by means of a thyratron.

supplied by the energy stored in the capacitor $C$ in the thyratron plate circuit.

To understand the operation of this circuit, suppose that the thyratron is nonconducting. The capacitor $C$ will become charged to the peak value of the transformer voltage through the series gas diode. This auxiliary charging circuit must be so connected as to provide the capacitor polarity as shown. It must also be phased properly with respect to the ignitron supply transformer as shown by the polarity markings on Fig. 12-42 so that the capacitor is charged during the half cycle when the ignitron anode
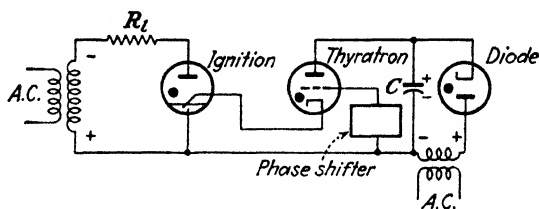


FIG. 12-42. An ignitron control circuit in which the ignitor-rod current does not pass through the load $R_l$.

is negative. If the thyratron grid voltage is now adjusted so that conduction will occur, the capacitor charge will flow through the thyratron and ignitor-rod circuits and breakdown will occur provided that the ignitron anode voltage is sufficiently positive.

The current surge through the thyratron and ignitor-rod circuit will quickly discharge the capacitor. As a result the thyratron plate current will fall below that necessary to maintain the arc, and the current through the thyratron and ignitor-rod circuit will fall to zero. The diode rectifier must now supply enough current to charge the capacitor again before the

cycle is over so that the plate voltage of the thyratron will be at its full value when the grid again assumes control. This action is repeated in each cycle.

## PROBLEMS

**12-1.** A type 1-V vacuum diode whose internal resistance is 200 ohms is to supply power to a 1,000-ohm load from a 300-volt (rms) source of supply. Calculate
   a. The peak load current.
   b. The d-c load current.
   c. The a-c load current.
   d. The d-c tube voltage.
   e. The total input power to the plate circuit.
   f. The efficiency of rectification.
   g. The percentage regulation from no load to the given load.

**12-2.** Show that the maximum d-c output power in a half-wave single-phase circuit occurs when the plate resistance equals the load resistance.

**12-3.** Prove that the regulation of both the half-wave and the full-wave rectifier is given by

$$\% \text{ regulation} = \frac{r_p}{R_l} \times 100\%$$

**12-4.** A full-wave single-phase rectifier consists of a double-diode tube (type 80) the internal resistance of each element of which may be considered to be constant and equal to 500 ohms. These feed into a pure resistance load of 2,000 ohms. The secondary transformer voltage to center tap is 280 volts rms. Calculate
   a. The d-c load current.
   b. The d-c current in each tube.
   c. The a-c voltage across each diode.
   d. The d-c output power.
   e. The efficiency of rectification.
   f. The percentage regulation from no load to the given load.

**12-5.** Verify Eq. (12-34), which gives the ripple factor in a half-wave gas-tube circuit.

**12-6.** Show that the input power to the plate circuit of any rectifier circuit using gas diodes may be expressed in the form

$$P_i = I_{\text{rms}}^2 R_l + E_0 I_{dc}$$

where the symbols have the meanings used in the text.

**12-7.** Gas diodes are used in a full-wave circuit.
   a. Show that the d-c current is

$$I_{dc} = \frac{2E_m}{\pi R_l}\left[\cos\alpha_1 - \left(\frac{\pi}{2} - \alpha_1\right)\sin\alpha_1\right]$$

where $\sin\alpha_1 = E_0/E_m$ and it is assumed that the maintaining and extinction voltages equal $E_0$.
   b. Show that the efficiency of rectification is

$$\eta_r = \frac{4}{\pi}\frac{\left[\cos\alpha_1 - \left(\frac{\pi}{2} - \alpha_1\right)\sin\alpha_1\right]^2}{\frac{\pi}{2} - \alpha_1 - \sin\alpha_1\cos\alpha_1} \times 100\%$$

**12-8.** A gas diode is used in a half-wave circuit to supply a 500-ohm resistor from the 220-volt a-c mains. The breakdown and maintaining voltages are constant at 10 volts. Calculate the readings of the following instruments:

a. A d-c ammeter in series with the load.

b. An a-c ammeter in series with the load.

c. A d-c voltmeter placed across the tube.

d. An a-c voltmeter placed across the tube.

e. A wattmeter whose current coil is in series with the load and whose voltage coil is across the input.

f. The efficiency of rectification.

**12-9.** A mercury-vapor diode is used in a half-wave circuit to supply power to a 15-ohm resistor from a 12-volt rms source of voltage. Assume that the breakdown and maintaining voltages of the tube are constant at 12 volts. Calculate

a. The peak load current.

b. The phase angles in each cycle at which conduction starts and stops.

c. The d-c load current.

d. The a-c load current.

e. The d-c load power.

f. The efficiency of rectification.

**12-10.** A high-pressure gaseous rectifier (see Sec. 11-3) is used to charge a 6-volt battery at a 6-amp charging rate. The tube drop is constant at 10 volts. The filament input is 18 amp at 2.2 volts. If the 120-volt a-c mains are applied directly, calculate

a. The value of the series resistor needed.

b. The over-all efficiency.

If the 120-volt d-c lines are used instead of the a-c lines, calculate the quantities called for in (a) and (b).

Assume that the battery resistance is constant at 0.05 ohm and that the battery voltage is constant at 6 volts throughout the charging process.

**12-11.** In the full-wave single-phase bridge, can the transformer and the load be interchanged? Explain carefully.

**12-12.** A 1-ma d-c meter whose resistance is 10 ohms is calibrated to read rms volts when used in a bridge circuit using copper-oxide elements. The effective resistance of each element may be considered to be zero in the forward direction and infinite in the inverse direction. The sinusoidal input voltage is applied in series with a 5,000-ohm resistor. What is the full-scale reading of this meter?

**12-13.** Find the minimum insulation voltage required of each of the filament transformers in a voltage doubling circuit if

a. The positive side of the output is grounded.

b. The negative is grounded.

The peak alternating voltage is $E_m$.
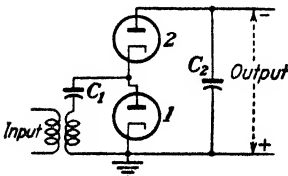


PROB. 12-14.

**12-14.** The circuit shown is a half-wave voltage doubler. Analyze the operation of this circuit. Calculate

a. The maximum possible voltage across each capacitor.

b. The peak inverse voltage of each tube.

c. The cathode-to-ground insulation stress.

Compare this circuit with the bridge voltage doubler of Fig. 12-15. In this circuit the output voltage is negative with respect to ground. Show that if the connections to the cathode and anode of each tube are interchanged, the output voltage will be positive with respect to ground.

**12-15.** The circuit of the preceding problem can be extended from a doubler to a quadrupler by adding two tubes and two capacitors as shown. (*a*) and (*b*) are alternate ways of drawing the same circuit.

*a.* Analyze the operation of this circuit.

*b.* Answer the same questions as those asked in Prob. 12-14.
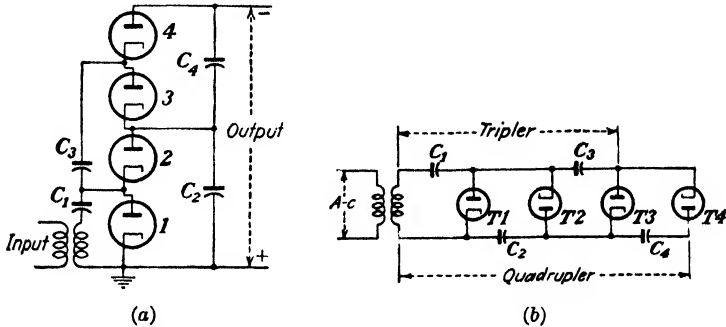


PROB. 12-15.

*c.* Generalize the circuit of this and the preceding problem so as to obtain $n$-fold multiplication when $n$ is any even number. In particular, sketch the circuit for sixfold multiplication.

*d.* Show that $n$-fold multiplication, with $n$ odd, can also be obtained provided that the output is properly chosen.

**12-16.** By connecting two half-wave doublers of the type illustrated in Prob. 12-14 to the same input, show that it is possible to obtain a full-wave quadrupler. Explain the operation of this circuit.

**12-17.** Explain, with the aid of a circuit diagram, exactly what happens in a full-wave single-phase circuit employing two single-anode tank rectifiers when an arc back occurs.

**12-18.** A controlled rectifier is adjusted so as to fire at an angle $\varphi$ in each cycle. The peak of the applied alternating voltage is $E_m$, and the load resistor is $R_l$. Derive expressions for

*a.* The rms tube current.

*b.* The rms voltage drop across the tube.

*c.* The total output power.

Assume that the tube drop is negligible in comparison with $E_m$.

**12-19.** The arc drop in a certain thyratron is 10 volts. The tube is operated from a 440-volt rms sinusoidal source. The load is a resistance of 100 ohms. Calculate the average value of the plate current

*a.* When the grid and plate voltages are in phase.

*b.* When the grid voltage lags the plate voltage by 30 deg.

*c.* When the grid voltage leads the plate voltage by 30 deg.

Assume that the critical grid voltage is zero for all values of plate voltage.

**12-20.** The thyratron circuit of Fig. 12-27 is used to regulate the average current in a 250-ohm plate load resistor of an FG-27A tube. The voltage $E_{ab} = 220$ volts rms at 60 cps, $Z_1$ is a 1,000-ohm resistor, and $Z_2$ is a variable inductor whose resistance may be neglected.

*a.* If the inductance is set at 2.65 henrys, draw a sinor diagram showing the grid and the plate voltages (before conduction starts in each cycle). Sketch these voltages

approximately to scale as a function of $\alpha = \omega t$. Determine the angle at which the tube will fire.

*b.* What is the magnitude of the d-c plate current that will be obtained with the conditions adjusted as in part *a*?

*c.* As the inductance is varied, what are the maximum and minimum values of d-c current that can be obtained?

*d.* If the resistance $R_L$ of the inductor is not negligible, discuss the modifications that must be made in the solution of this problem. In particular, redraw the sinor diagram, taking $R_L$ into account.

*e.* If $Z_1$ is a variable capacitor and $Z_2$ is a 1,000-ohm resistor, repeat parts *a*, *b*, and *c*. For part *a* the capacitance is adjusted to 1.53 $\mu f$.

**12-21.** In the thyratron circuit shown the transformer primary and secondary windings are so arranged that continuous phase-shift control is obtained as $R$ is varied.

*a.* Draw a sinor diagram showing the grid and plate voltages (before conduction starts in each cycle).

*b.* What is the plate current if $R$ is extremely large?

*c.* What is the plate current if $R$ is very small?

*d.* What is the plate current if $R = X_L$?

The transformer reactances may be considered to be infinite.
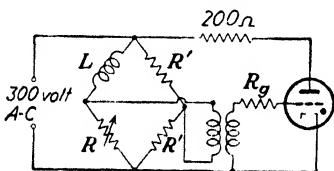


Prob. 12-21.
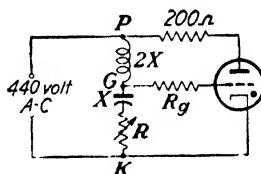


Prob. 12-22.

**12-22.** In the phase-shift-controlled thyratron circuit shown the reactance of the inductor is twice the reactance of the capacitor. The resistance of the choke coil may be neglected. The phase is varied by varying the resistance $R$.

*a.* Draw the sinor circle diagram showing the phase angle $\theta$ between the grid and plate voltage (before conduction starts in each cycle).

*b.* Prove that $\tan(\theta/2) = X/R$ and that the magnitude of the grid voltage equals the magnitude of the plate voltage.

*c.* Between what limits can the d-c load current be varied? To what values of $R$ do these limits correspond?

**12-23.** It is desired to control the current in the 100-ohm resistor in the plate circuit



Prob. 12-23.

of the thyratron shown by means of the phase-shift method.
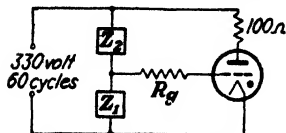
*a.* If a fixed 0.5-$\mu f$ capacitor and an adjustable resistor $R$ are available, which must be $Z_1$ and which must be $Z_2$? Show by a sinor diagram the reason for your choice.

*b.* What value of $R$ will cause the largest d-c plate current? Calculate the value of this current.

*c.* What is the minimum value of the controlled d-c current?

*d.* What must the value of $R$ be in order that 70 per cent of the maximum possible current pass through the load?

*e.* Repeat parts *a*, *b*, *c*, and *d* if a 10-henry inductor (of negligible resistance) is available instead of the 0.5-$\mu f$ capacitor.

**12-24.** *a.* In the thyratron circuit shown, a fixed inductor of 4 henrys and a variable resistor are used. With the aid of a sinor diagram decide whether $Z_1$ or $Z_2$ must be the resistor in order to have control as $R$ is varied.

*b.* What is the maximum reading of a d-c ammeter in series with the load, and for what value of $R$ is this obtained?

*c.* The resistor is set at 1,000 ohms. Calculate the reading of an a-c voltmeter across the tube.
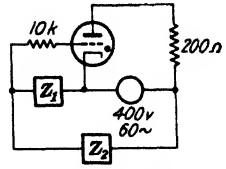
**12-25.** By error, balanced three-phase voltages are applied to the terminals $A$, $K$, and $B$ of Fig. 12-27. If $Z_1 = R$ and $Z_2 = \omega L$, draw sinor diagrams to determine whether or not it is still possible to obtain control. (There are two such diagrams possible, depending upon the phase sequence of the applied voltages.)



PROB. 12-24.

**12-26.** An FG-27A thyratron is operated at a mercury condensation temperature of 35°. The plate supply voltage is 440 volts rms. The load is a 100-ohm resistor. Calculate the d-c plate current under the following conditions:

*a.* The grid bias is $-5$ volts.

*b.* An a-c voltage whose rms value is 5 volts is used instead of a battery in the grid circuit. Assume that the grid lags the plate voltage by 90 deg.

*c.* The grid circuit is excited by the d-c and a-c voltages of parts *a* and *b* in series.

**12-27.** In Fig. 12-35, $C$ is replaced by a 1,000-ohm resistor, and $R$ is replaced by a 1.53-henry inductor. The power line is 115 volts rms at 60 cps. The load is a 250-ohm resistor. Assume that the critical grid curve of the thyratron coincides with the zero voltage axis and that the tube drop is zero.

*a.* If $E_c = 0$, find the angle between the grid and plate voltages before breakdown by means of a sinor diagram.

*b.* If $E_c = 67$ volts (with the grid negative with respect to the cathode), draw a diagram of the instantaneous grid and plate voltages as a function of time before breakdown.

*c.* Under the conditions of part *b*, find the d-c plate current.

**12-28.** In Fig. 12-35 it is desired that the d-c current through the load be varied from maximum current to 2 amp by means of the d-c bias. The magnitude of the reactance $X_C$ is twice the resistance $R$. The power line is 240 volts at 60 cps, and the load resistor is 25 ohms.

*a.* What d-c current is obtained when the d-c bias $E_c$ is zero?

*b.* What is the maximum d-c current, and at what bias is it obtained?

*c.* What bias must be supplied to reduce the d-c current to 2 amp?

**12-29.** A thyratron is controlled by the bias phase method. The plate supply is 400 volts rms, the load resistance is 1,000 ohms, and the a-c grid supply, which lags the plate supply by 30 deg, is 100 volts rms. Plot roughly to scale the d-c load current as a function of the d-c bias from $-200$ to $+200$ volts. In particular, calculate

*a.* The maximum current and the d-c bias at which it is obtained.

*b.* The current at zero d-c bias.

*c.* The bias at which the current just will drop to zero.

**12-30.** Bias phase control is employed in a thyratron circuit. The critical grid starting characteristic of the tube may be assumed to coincide with the zero voltage axis. Sketch roughly to scale the ratio $I_{dc}/I'$ as a function of the ratio of the d-c bias $E_c$ to the peak a-c grid voltage $E_{gm}$ for the following conditions:

*a.* The grid voltage lags the plate voltage by 90 deg.

*b.* The grid voltage leads the plate voltage by 90 deg.

*c.* The grid voltage lags the plate voltage by 150 deg.

*d.* The grid voltage leads the plate voltage by 150 deg.

*e.* For case *a* prove that the result is a straight line, as shown in Fig. 12-38.

**12-31.** An FG-33 positive-grid thyratron is operated at an ambient temperature of 35°. The plate excitation is a sinusoidal voltage whose peak value is 400 volts. The load is a 100-ohm resistor. An a-c grid voltage of adjustable maximum value $E_{gm}$ is applied in phase with the plate voltage. Plot a curve showing the relationship between the average plate current $I_{dc}$ and $E_{gm}$.

**12-32.** The peak sinusoidal voltage applied to the plate of an FG-27A thyratron, operating at 35° ambient temperature, is 400 volts. The load is a 100-ohm resistor.



PROB. 12-33.

The a-c grid voltage lags the plate voltage by a fixed angle of 150 deg. The peak grid voltage $E_{gm}$ is adjustable. Plot the curve showing the relationship between the average plate current $I_{dc}$ and $E_{gm}$.
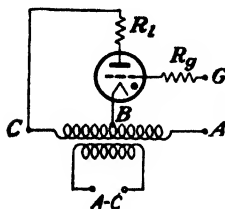
**12-33.** What fraction of the maximum plate current is obtained in the thyratron circuit shown when the grid is connected to points *A*, *B*, and *C*, respectively? Draw diagrams of instantaneous plate and grid voltages in each case.

**12-34.** Two FG-27A tubes operate at a temperature of 50°C in a full-wave controlled rectifier circuit.

*a.* If the firing angle is $\varphi$ deg behind the anode voltage, sketch curves showing the current wave shape in the load.

*b.* Derive expressions for the following:

1. D-c current in load.
2. A-c current in load.
3. Efficiency of rectification.

**12-35.** FG-27A thyratrons are employed in the full-wave controlled rectifier illustrated. The tubes operate at 50°C.

*a.* The lines *a* and *b* of the control circuit are to be connected to the power lines 1
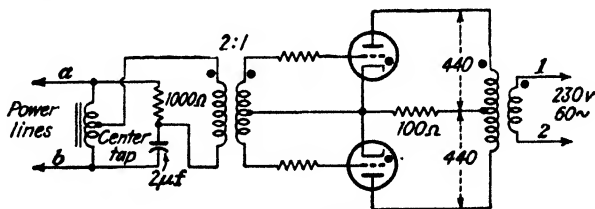


PROB. 12-35.

and 2. Show by a voltage diagram which lines are connected together for conduction to occur for less than a half cycle per tube.

*b.* Calculate the d-c current in the load.

*c.* Calculate the total a-c power to the plate circuit.

**12-36.** The grid and plate circuits of a thyratron are excited by 115-volt sinusoidal voltages. Sketch the wave shape of the output current if the ratio of the frequency of the plate voltage to the frequency of the grid voltage is

*a.* 2.

*b.* ½.

*c.* 3.

*d.* ⅓.

**12-37.** Design a circuit using thyratrons so that the plate current will have the "chopped" sinusoidal wave shape shown. Such a wave shape is employed as the "heat"
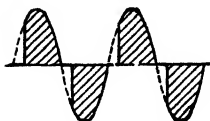


PROB. 12-37.

control of electric resistance welding, although welders require ignitrons in order to meet the current requirements.

**12-38.** An ignitron is used in the circuit of Fig. 12-40. The peak plate voltage is 500 volts, and the load is a 3-ohm resistor. Assume that conduction begins when the ignitor current exceeds 10 amp and that the ignitor rod acts as an ohmic resistance whose magnitude is 7 ohms. Calculate

a. The peak power taken by the rod.
b. The average power taken by the rod.
c. The d-c power supplied to the load.
d. The ratio of (b) to (c).

Neglect the tube drop of the diode in series with the ignitor rod. Neglect the ignitron tube drop. The frequency is 60 cps.

**12-39.** Sketch a single-phase full-wave rectifier circuit using ignitrons and any other auxiliary apparatus needed.

**12-40.** An ignitron is used as a half-wave rectifier. The ignitor is adjusted so that conduction commences 90 deg after the start of each cycle of applied voltage. The tube drop during conduction equals 20 volts. The applied voltage is 300 volts rms, and the load is a 50-ohm resistor. Calculate

a. The d-c load current.
b. The power taken by the tube.
c. The total power delivered to the plate circuit.
d. The rms load current.
e. The total power taken by the load resistor. Obtain the solution to (e) from that in (d), and check this against the result obtained by subtracting the solution to (b) from that to (c).

**12-41.** An ignitron is used in a half-wave circuit to supply power to a 15-ohm resistor from a 320-volt rms source of supply. Breakdown occurs when the potential across the tube in the conduction direction is 225 volts. When the tube conducts, the potential drop across it is 15 volts. Calculate

a. The reading of an a-c ammeter in series with the load.
b. The reading of a d-c ammeter in series with the load.
c. The reading of an a-c voltmeter across the tube.
d. The reading of a d-c voltmeter across the tube.
e. The reading of an a-c voltmeter across the load.
f. The reading of a d-c voltmeter across the load.
g. The power dissipated in the tube.
h. The d-c load power.
i. The average power supplied by the source.
j. The efficiency of rectification.

**12-42.** Two ignitrons are operated in the *inverse-parallel* connection shown. The firing angle is set at 30 deg. Assume ideal transformers and a resistive load.
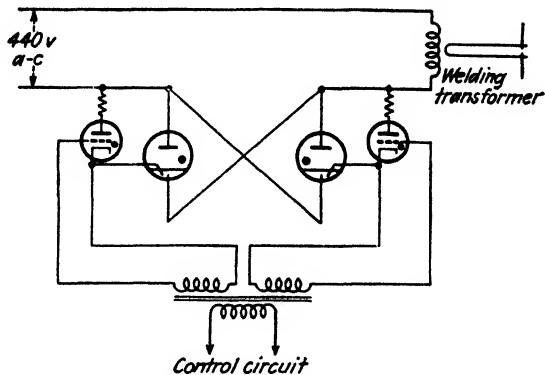


PROB. 12-42.

Draw diagrams showing the following wave shapes in the circuit.

*a.* Voltage across ignitron.

*b.* Voltage across welding transformer.

**12-43.** *a.* In the circuit sketched, neither tube is conducting. A positive pulse is applied to the grid of tube 1 and fires it. Then a positive pulse is applied to the grid of tube 2. This extinguishes tube 1 and fires tube 2. Explain why.

If, now, another positive pulse is applied to tube 1, what will happen?

What is the limitation on the rapidity with which two succeeding pulses may be applied so as to cause the current to shift from one tube to the next?

*b.* The points 1 and 2 are connected together so that a positive pulse is applied simultaneously to the grids of both tubes. Explain clearly what will happen as successive pulses are applied to the grids.

NOTE: It must be remembered that the critical grid starting characteristics vary slightly from tube to tube.
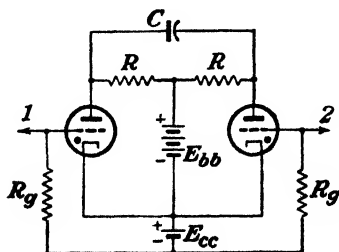


PROB. 12-43.
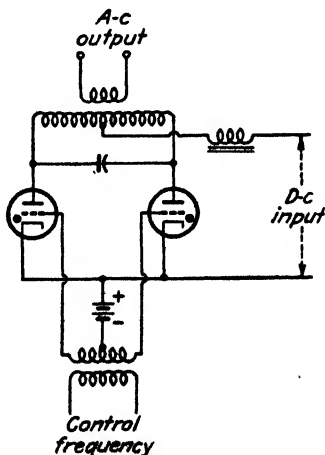


PROB. 12-44.

**12-44.** Consider the separately excited parallel inverter circuit sketched. Derive an expression for the potential across the capacitor $C$, which is also the output potential,

if a perfect output transformer is assumed. If the time constant of the circuit is small compared with the time for one-half cycle, show that the voltage across $C$ is a square wave whose amplitude is equal to the d-c input voltage minus the tube drop and whose frequency equals the control frequency.

**12-45.** A single-phase full-wave rectifier is set to operate with a 30-deg delay angle. The load consists of a resistor in series with an inductor, the inductance of which is so large that the load current may be considered to be constant. The secondary transformer voltage to center tap is 230 volts rms. The tube drop during conduction is constant and is equal to 20 volts.

*a.* Sketch the voltage of each plate and also of the cathodes (with respect to the transformer-secondary center tap) as a function of time.

*b.* Sketch—directly below the curves of part *a*—the anode voltage of each tube as a function of time.

*c.* Calculate the anode voltage of one of the tubes at the instant before it fires.

*d.* Calculate the output voltage across the resistor.

*e.* Sketch—directly below the curves of part *b*—the voltage across the choke as a function of time.

*f.* Repeat parts *a*, *b*, *c*, and *d* under the condition that the load is a pure resistance.

## REFERENCES

1. TONKS, L., and I. LANGMUIR, *Phys. Rev.*, **33**, 195, 1929.
2. GREINACHER, H., *Z. Physik*, **4**, 195, 1921.
   COCKCROFT, J. D., and E. T. S. WALTON, *Proc. Roy. Soc. (London)*, **136**, 619, 1932.
   GARSTAND, W. W., *Electronics*, **4**, 50, February, 1932.
   WAIDELICH, D. L., *ibid.*, **14**, 28, May, 1941; *Proc. IRE*, **29**, 554, 1941.
   WAIDELICH, D. L., and C. L. SHAKELFORD, *Proc. IRE*, **32**, 470, 1944.
   WAIDELICH, D. L., and H. A. K. TAGKIN, *ibid.*, **33**, 449, 1945.
3. SNODDY, L. B., *Physics*, **4**, 366, 1933.
4. HULL, A. W., *Gen. Elec. Rev.*, **32**, 390, 1929.
   HERSKIND, C. C., and E. J. REMSCHEID, *Trans. AIEE*, **65**, 632, 1946.
5. SEELY, S., "Electron-tube Circuits," McGraw-Hill Book Company, Inc., New York, 1950.
6. REICH, H. J., "Theory and Applications of Vacuum Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.
   HENNEY, K., "Electron Tubes in Industry," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937.

# CHAPTER 13

# FILTERS FOR RECTIFIERS

It is usually the requirement of a power supply to provide a substantially ripple-free source of d-c voltage and current from an a-c line. It is demonstrated in the last chapter, however, that the output of a rectifier contains ripple components in addition to the steady term. Because of this, it is customary to include a filter of some type between the rectifier and the load in order to attenuate these ripple components.

The analysis of rectifier circuits containing filters that are composed of simple circuit elements is rather complex, owing to the fact that the rectifier as a driving source is a nonlinear device. Hence the straightforward methods of linear a-c theory cannot, in general, be applied.

No simple complete method of treating nonlinear devices exists.[1] Two methods of analysis are possible. In one, numerical or graphical integration may be used to solve the nonlinear differential equations that control the system performance. The more usual approach is to make reasonable assumptions that reduce the network to an approximate linear equivalent circuit, which may then be analyzed by the usual methods of electric-circuit theory. It is the latter procedure that will be followed throughout. In consequence, the results that are obtained are not exact but do represent good engineering approximations.

**13-1. The Harmonic Components in Rectifier Circuits.** An analytic representation of the output current wave of the single-phase half-wave rectifier is obtained by means of a Fourier series. This series representation has the form

$$i = B_0 + \sum_{k=1}^{\infty} B_k \cos k\alpha + \sum_{k=1}^{\infty} A_k \sin k\alpha \qquad (13\text{-}1)$$

where $\alpha = \omega t$. For convenience, the subscript $b$ is dropped from $i_b$. The coefficients that appear in the series are given by the following integrals

$$
\left.
\begin{aligned}
B_k &= \frac{1}{\pi} \int_0^{2\pi} i \cos k\alpha \, d\alpha \\[2mm]
A_k &= \frac{1}{\pi} \int_0^{2\pi} i \sin k\alpha \, d\alpha \\[2mm]
B_0 &= \frac{1}{2\pi} \int_0^{2\pi} i \, d\alpha
\end{aligned}
\right\}
\qquad (13\text{-}2)
$$

It should be recalled that the constant term $B_0$ in the Fourier series is simply the average or d-c value of the current.

By performing the indicated integrations, using Eq. (12-3) for the explicit expressions for the current over the two specified intervals, there results

$$i = I_m \left[ \frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \sum_{k=2,4,6,\ldots} \frac{\cos k\omega t}{(k+1)(k-1)} \right] \qquad (13\text{-}3)$$

where

$$I_m = \frac{E_m}{r_p + R_l}$$

and where $I_m$ is the peak value of the current and $E_m$ is the maximum value of the transformer voltage. The lowest angular frequency that is present in this expression is that of the primary source of the a-c power. Also, except for this single term of angular frequency $\omega$, all other terms in the final expression are even harmonics of the power frequency.

The corresponding expression for the output of the full-wave rectifier, which is illustrated in Fig. 12-9, may be derived from Eq. (13-3). By recalling that the full-wave circuit consists essentially of two half-wave circuits which are so arranged that one circuit operates during one half cycle and the second operates during the second half cycle, then it is clear that the currents are functionally related by the expression $i_1(\alpha) = i_2(\alpha + \pi)$. The total load current $i = i_1 + i_2$ attains the form

$$i = I_m \left[ \frac{2}{\pi} - \frac{4}{\pi} \sum_{\substack{k \text{ even} \\ k \neq 0}} \frac{\cos k\omega t}{(k+1)(k-1)} \right] \qquad (13\text{-}4)$$

where, as in Eq. (13-3),

$$I_m = \frac{E_m}{r_p + R_l}$$

and where $E_m$ is the maximum value of the transformer voltage to center tap. For convenience, this will be referred to simply as the "transformer maximum voltage."

It will be observed that the fundamental angular frequency $\omega$ has been eliminated from the equation, the lowest frequency in the output being $2\omega$, a second harmonic term. This offers a definite advantage in the effectiveness of filtering of the output. A second desirable feature of the full-wave circuit is the fact that the current pulses in the two halves of the transformer winding are in such directions that the magnetic cycle through which the iron of the core is taken is essentially that of the a-c current. This eliminates any d-c saturation of the transformer core, which would give rise to additional harmonics in the output.

The Fourier series representation of the half-wave and the full-wave circuits using gas diodes can be obtained as above, although the form will be more complex. This greater complexity occurs because the conduction begins at some small angle $\alpha_1$ and ceases at the angle $\pi - \alpha_1$, when it is assumed that the breakdown and the extinction potentials are equal. But since the angle $\alpha_1$ is usually small under normal operating conditions, it will be supposed that Eqs. (13-3) and (13-4) are valid for circuits with either vacuum or gas diodes.

**13-2. Inductor Filters.** The operation of the inductor filter depends on the fundamental property of an inductor to oppose any change of current that may be taking place in the circuit. As a result of this, any sudden changes that might occur in a circuit without an inductor would be smoothed out by the presence of an inductor in the circuit.
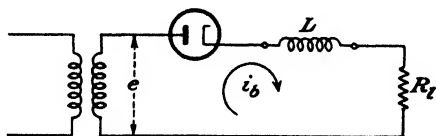


FIG. 13-1.    Half-wave rectifier with choke filter.

Suppose that an inductor or "choke" filter is connected in series with the load in a single-phase half-wave circuit, as illustrated in Fig. 13-1. For simplicity in the analysis, suppose that the tube and choke resistances are negligible. Then the controlling differential equation for the current in the circuit during the time that current flows is

$$e = E_m \sin \omega t = L \frac{di_b}{dt} + R_l i_b \qquad (13\text{-}5)$$

An exact solution of this differential equation may be effected subject to the initial condition that $i_b = 0$ at $t = 0$. The solution is valid only as long as it yields a positive value of current since the diode can conduct only in one direction. The time at which the current falls to zero is called the "cutout point." The solution is given in Prob. 13-1, and the results are illustrated graphically in Fig. 13-2 with $\omega L/R_l$ as a parameter. The effect of changing the inductance on the wave form of the current is clearly seen. An oscillogram of the wave shape in a particular system is given in Fig. 13-3. It should be noted that the simple inductor filter is seldom used with a half-wave circuit.

Suppose that a choke-input filter is applied to the output of a single-phase full-wave rectifier. The circuit is given in Fig. 13-4. An oscillogram showing the wave forms obtained experimentally is given in Fig. 13-5.
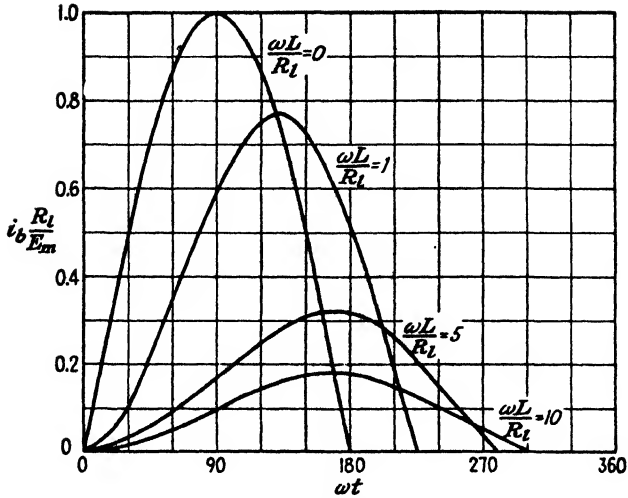
FIG. 13-2. The effect of changing the inductance on the wave form of the output current in a half-wave rectifier with an inductor filter. The load resistance $R$ is assumed constant.
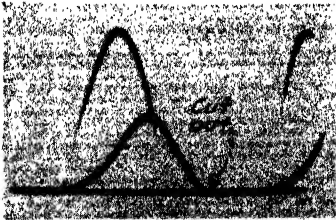


FIG. 13-3. Oscillogram of the load voltage with and without an inductor filter in a single-phase half-wave rectifier.
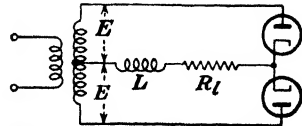


FIG. 13-4. The schematic wiring diagram of a full-wave choke-input-filtered rectifier.
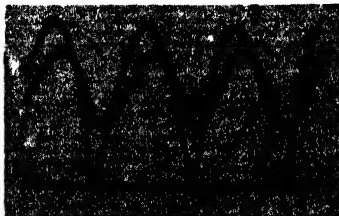


FIG. 13-5. Oscillogram of the load voltage in a single-phase full-wave rectifier circuit with a simple inductor filter. The load voltages with the inductor shorted out are also shown.

An exact solution of the circuit differential equation can be effected (see Prob. 13-2). However, since no cutout occurs in the load current, it is now simpler to proceed by finding an approximate solution. The results will be sufficiently accurate for most applications and will be in a much more useful form than the exact solution.

The voltage that is applied to the circuit comprising the load resistor and the inductor filter is that given in Eq. (13-4), with the current replaced by the voltage ($I_m$ is replaced by $E_m$). It is now noted that the amplitudes of the a-c terms beyond the first are small compared with the amplitude of the first term in the series. Thus, the fourth harmonic frequency term is only 20 per cent of the second harmonic term. Further, since the impedance of the inductor increases with the frequency, better filtering action for the higher harmonic terms results. It is therefore expected that the harmonics in the output will be principally of second harmonic frequency. It is justifiable to neglect all harmonics in the output except the first a-c term. That is, the equivalent circuit of the rectifier under these circumstances is assumed to be that illustrated in Fig. 13-6. For the sake of simplicity and because they introduce little error, the tube drop and the tube resistance will be neglected in the ripple calculations of this chapter.



FIG. 13-6. The equivalent circuit of a full-wave choke-input-filtered rectifier.

In addition, the resistance and leakage inductance of the transformer and the resistance of the inductor will likewise be neglected.

It is noticed that only linear elements exist in the equivalent circuit that is excited by a voltage, which consists of a battery $2E_m/\pi$ in series with an a-c source whose emf is $(-4E_m/3\pi) \cos 2\omega t$. The load current will then be, in accordance with elementary circuit theory,

$$i_l = \frac{2E_m}{\pi R_l} - \frac{4E_m}{3\pi} \frac{\cos(2\omega t - \psi)}{\sqrt{R_l^2 + 4\omega^2 L^2}} \tag{13-6}$$

where

$$\tan \psi = \frac{2\omega L}{R_l} \tag{13-7}$$

That is, the load voltage curve in Fig. 13-5 is expressed by Eq. (13-6) (multiplied by $R_l$).

The ripple factor, defined in Eq. (12-20), becomes

$$r = \frac{\dfrac{4E_m}{3\pi\sqrt{2}} \dfrac{1}{\sqrt{R_l^2 + 4\omega^2 L^2}}}{2E_m/\pi R_l} = \frac{2R_l}{3\sqrt{2}} \frac{1}{\sqrt{R_l^2 + 4\omega^2 L^2}}$$

which may be expressed in the form

$$r = \frac{2}{3\sqrt{2}} \frac{1}{\sqrt{1 + 4\omega^2 L^2/R_l^2}} \qquad (13\text{-}8)$$

This expression shows that filtering improves with decreased circuit resistance or, correspondingly, with increased currents. At no load, $R_l = \infty$, whence the filtering is poorest, and $r = 2/(3\sqrt{2}) = 0.47$. This is the result that applies when no choke is included in the circuit. This result should be compared with Eq. (12-27), which gives 0.482. The difference arises from the higher order terms in the Fourier series that have been neglected in the present calculation.

If the ratio $2\omega L/R_l$ is large, then the ripple factor reduces to

$$r = \frac{1}{3\sqrt{2}} \frac{R_l}{\omega L} \qquad (13\text{-}9)$$

This shows that at any load the ripple varies inversely as the magnitude of the inductance. Also, the ripple is smaller for small values of $R_l$, that is, for high currents. This indicates the desirability of using low resistance chokes to keep the numerator small; also, of using large inductance chokes so as to make the denominator large.

The d-c output voltage is given by

$$E_{dc} = I_{dc}R_l = \frac{2E_m}{\pi} = 0.637E_m = 0.90E_{\text{rms}} \qquad (13\text{-}10)$$

where $E_{\text{rms}}$ is the transformer secondary voltage measured to the center tap. Note that under the assumptions made in the analysis the output voltage is a constant, independent of the load; *i.e.*, perfect regulation exists. Because of the effect of the choke resistance, the resistance of the tube, and the resistance of the transformer winding, the foregoing represents the output only at no load. The output voltage will decrease as the current increases in accordance with the equation

$$E_{dc} = \frac{2E_m}{\pi} - I_{dc}R \qquad (13\text{-}11)$$

where $R$ is the total resistance in the circuit, exclusive of the load.

**13-3. Capacitor Filters.**[2] Filtering is frequently effected by shunting the load with a capacitor. The action of this system depends upon the fact that the capacitor stores energy during the conduction period and delivers this energy to the load during the inverse, or nonconducting, period. In this way, the time during which the current passes through the load is prolonged, and the ripple is considerably decreased.

Consider the half-wave rectifier that is connected according to the diagram of Fig. 13-7. Suppose first that the load resistor $R_l = \infty$. The capacitor will charge to the potential $E_m$, the transformer maximum value. Further, the capacitor will maintain this potential, for no path exists by which this charge is permitted to leak off, since the tube will not pass a negative current. The tube resistance is infinite in the inverse direction, and no charge can flow during this portion of the cycle. Consequently, the filtering action is perfect, and the capacitor voltage remains constant at its peak value, as is seen in Fig. 13-8.



Fig. 13-7. A single-phase half-wave capacitor-filtered rectifier.

Under these conditions,

$$e = e_c + e_b$$

or

$$e_b = e - e_c \tag{13-12}$$

The voltage $e_c$ across the capacitor is, of course, the same as the voltage $e_l$ across the load resistor, since the two elements are in parallel. It is seen from the oscillogram of Fig. 13-8 that the tube voltage is always negative and that the peak inverse voltage is then twice the transformer maximum.



Fig. 13-8. Oscillogram of voltages in a single-phase half-wave condenser-filtered rectifier *at no load*. The output voltage $e_c$ is a constant, indicating perfect filtering. The tube voltage $e_b$ is negative for all values of time, and the peak inverse voltage is $2E_m$.

Hence, the presence of the capacitor causes the peak inverse voltage to increase from a value equal to the transformer maximum when no capacitor filter is used to a value equal to twice the transformer maximum value when the filter is used.

Suppose, now, that the load resistor $R_l$ is finite. Without the capacitor input filter, the load current and the load voltage during the conduction period will be sinusoidal functions of the time, as shown in Fig. 12-4.

The inclusion of a capacitor in the circuit results in the capacitor charging in step with the applied voltage. Also, the capacitor must discharge through the load resistor, since the tube will prevent a current in the negative direction. Clearly, the diode acts as a switch which permits charge to flow into the capacitor when the transformer voltage exceeds the capacitor voltage, and then acts to disconnect the power source when the transformer voltage falls below that of the capacitor.

The analysis will now proceed in two steps. First the conditions during conduction will be considered, and then the situation when the tube is nonconducting will be investigated.

*Diode Conducting.* If the tube drop is neglected, then the transformer voltage is impressed directly across the load. Hence, the output voltage is $e_l = E_m \sin \omega t$. The question immediately arises: Over what interval of time is this equation applicable? In other words, over what portion of each cycle does the diode remain conducting? The point at which the diode starts to conduct is called the "cutin point," and that at which it stops conducting is called the "cutout point." The latter will first be found in the same manner as that indicated for obtaining the cutout point for a half-wave inductor filter. The expression for the tube current is found, and the instant where this current falls to zero is the cutout time.

The expression for the tube current can be written down directly. Since the transformer voltage is sinusoidal and is impressed directly across $R$ and $C$ in parallel, then the sinor current $\mathbf{I}_b$ is found by multiplying the sinor voltage $\mathbf{E}$ by the complex admittance $(1/R) + j\omega C$. Hence

$$
\begin{aligned}
\mathbf{I}_b &= \left( \frac{1}{R} + j\omega C \right) \mathbf{E} \\
&= \left[ \sqrt{\left( \frac{1}{R} \right)^2 + \omega^2 C^2} \; \underline{/\tan^{-1} \omega C R} \right] \mathbf{E}
\end{aligned}
\tag{13-13}
$$

Since $\mathbf{E}$ has a peak value $E_m$, then the instantaneous current is

$$
i_b = E_m \sqrt{\omega^2 C^2 + \frac{1}{R_l{}^2}} \; \sin (\omega t + \psi)
\tag{13-14}
$$

where

$$
\psi \equiv \tan^{-1} \omega C R_l
\tag{13-15}
$$

This expression shows that the use of a large capacitor in order to improve the filtering for a given load $R_l$ is accompanied by a high peak tube current $i_b$. The tube current has the form illustrated in Fig. 13-9a. But it is desirable to use large capacitors for a given load in order to maintain the voltage more nearly constant. As a result, the interval over which the tube conducts may be quite small. Thus, for a certain required average

current demanded by the load, the tube current will become more peaked, and the conduction period will decrease as the capacitor is made larger. The oscillograms of Fig. 13-9b show experimentally determined results.

It is to be emphasized that the use of a capacitor filter may impose serious duty conditions on the rectifying diode, since the average current through the tube may be well within the current rating of the tube, and yet the peak current may be very high. In the case of vacuum diodes the large peak current demands will cause no appreciable damage to the cathode, since the tube current is limited by the temperature saturated value. Of course, if a current in excess of the temperature saturated value is



Theoretical

(a)                                          (b)
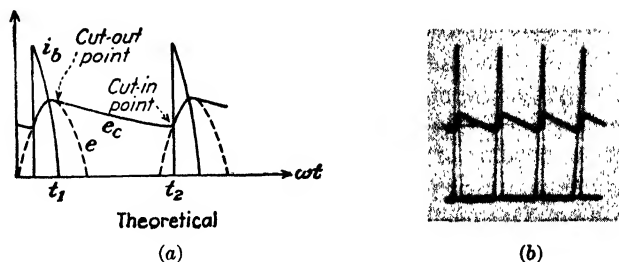
Fig. 13-9. (a) Sketch of tube current and load voltage in a single-phase half-wave capacitor-filtered rectifier. (b) Oscillogram of tube current and load voltage in a full-wave capacitor-filtered rectifier.

specified by the equations, the analysis must be modified, as this saturation value imposes a physical limitation on the output. In this case, the tube current pulse will flatten out, and the time duration will increase. The analysis to follow is not valid under these conditions. In the case of gas tubes, however, if a current higher than the saturation value is required by the circuit, even for very short intervals of time, the tube drop will increase and positive-ion bombardment of the cathode will be the result. If the tube drop exceeds the disintegration voltage of the cathode, the cathode may be permanently injured. It is for this reason that large capacitor input filters should not be used with rectifiers that employ gas diodes, without a careful check of the peak current demands.

The cutout time $t_1$ is found by equating the tube current to zero at this time. Thus, from Eq. (13-14),

$$0 = \sin(\omega t_1 + \psi)$$

or
$$\omega t_1 + \psi = n\pi$$

where $n$ is any positive or negative integer. The value of $t_1$ indicated in Fig. 13-9 in the first half cycle corresponds to $n = 1$ or

$$\omega t_1 = \pi - \psi = \pi - \tan^{-1} \omega C R_l \qquad (13\text{-}16)$$

*Diode Nonconducting.* In the interval between the cutout time $t_1$ and the cutin time $t_2$, the diode is effectively out of the circuit and the capacitor discharges through the load resistor with a time constant $CR_l$. Thus, the capacitor voltage (equal to the load voltage) is

$$e_c = A \epsilon^{-t/CR_l} \tag{13-17}$$

To determine the value of the constant $A$ appearing in this expression, it is noted from Fig. 13-9$a$ that at the time $t = t_1$, the cutout time,

$$e_c = e = E_m \sin \omega t_1$$

whence

$$A = (E_m \sin \omega t_1)\epsilon^{+t_1/CR_l} \tag{13-18}$$

Equation (13-17) thus attains the form

$$e_c = (E_m \sin \omega t_1)\epsilon^{-(t-t_1)/CR_l} \tag{13-19}$$

Since $t_1$ is known from Eq. (13-16), then $e_c$ can be plotted as a function of time. This exponential curve is indicated in Fig. 13-9, and where it intersects the sine curve $E_m \sin \omega t$ (in the following cycle) is the cutin point $t_2$. The validity of this statement follows from the fact that at an instant of time greater than $t_2$, the transformer voltage $e$ (the sine curve) is greater than the capacitor voltage $e_c$ (the exponential curve). Since the tube voltage is $e_b = e - e_c$, then $e_b$ will be positive beyond $t_2$ and the tube will become conducting. Thus, $t_2$ is the cutin point.

The output voltage consists of a section of the input sine curve followed by an exponential section. The cutin time $t_2$ cannot be given by an explicit analytic expression but must be found graphically by the method outlined above.

In principle at least, the foregoing results permit a complete analysis of the capacitor filter to be effected. For given values of $\omega$, $R_l$, $C$, and $E_m$ the tube current is given by Eq. (13-14). If $\alpha \equiv \omega t$, $\alpha_1 \equiv \omega t_1$, and $\alpha_2 \equiv \omega t_2$, the output potential is given by

$$\left. \begin{array}{ll} e_c = E_m \sin \alpha & \text{for } \alpha_2 < \alpha < \alpha_1 \\ \text{and by Eq. (13-19)} & \\ e_c = (E_m \sin \alpha_1)\epsilon^{-(\alpha-\alpha_1)/\omega CR_l} & \text{for } \alpha_1 < \alpha < 2\pi + \alpha_2 \end{array} \right\} \tag{13-20}$$

In these equations $\alpha_1$ and $\alpha_2$ represent the cutout and cutin angles *in the first half cycle*, respectively. The cutout angle is found from Eq. (13-16).

The d-c output voltage, the ripple factor, the peak tube current, etc., may then be calculated. These quantities can be plotted as functions of the parameters $\omega$, $R_l$, $C$, and $E_m$. Such an analysis is quite involved, but it has been carried out,[3] and the results are given in graphical form.

It is possible to make several reasonable approximations which permit an analytic solution to the problem. This approximate solution possesses the advantage that it clearly indicates the dependence of the important factors as a function of the circuit parameters. This approximate analysis is sufficiently accurate for most engineering applications and is carried out in the next section.

**13-4. Approximate Analysis of Capacitor Filters.** Refer to the oscillograms of Fig. 13-10, which show the load voltage for a half-wave and a



(a)                                             (b)
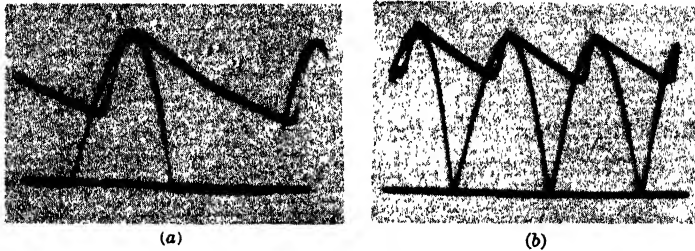
FIG. 13-10. Oscillograms of the load voltage in a single-phase (a) half-wave and (b) full-wave capacitor-filtered rectifier. The load voltages with the capacitor open-circuited are also shown. Since the tube drop, the tube and transformer resistance, and the transformer leakage reactance are not zero, cutin does not occur exactly at the point $e_c = e$, and during conduction the load voltage does not exactly equal the impressed transformer voltage.

full-wave circuit with a capacitor filter. Observe that the voltage curve may be approximated by a broken curve which is made up of portions of straight lines, as shown in Fig. 13-11. The peak value of this wave is $E_m$,
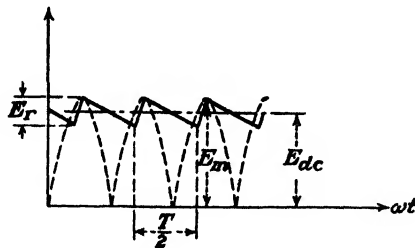


FIG. 13-11. The approximate load-voltage wave form in a full-wave capacitor-filtered rectifier.

the transformer maximum voltage. If the total capacitor discharge voltage is denoted by $E_r$, then from the diagram, the average value of the voltage is

$$E_{dc} = E_m - \frac{E_r}{2} \tag{13-21}$$

The instantaneous ripple voltage is obtained by subtracting $E_{dc}$ from the instantaneous load voltage. The result is shown in Fig. 13-12. The rms value of this "triangular wave" is independent of the slopes or lengths of the straight lines and depends only upon the peak value. Calculation of this rms ripple voltage yields

$$E_{rms}' = \frac{E_r}{2\sqrt{3}} \tag{13-22}$$

The two most important characteristics of the capacitor filter circuit, the regulation and the ripple, are contained in Eqs. (13-21) and (13-22), respectively. It is necessary, however, to express $E_r$ as a function of the load current and the capacitance. If $T_2$ represents the total nonconducting time, then the capacitor, when discharging at the constant rate $I_{dc}$, will

FIG. 13-13. The charging of the output capacitor is assumed to take place along a straight line passing through the zero and the peak values of the input sine curve.

FIG. 13-12. The approximate wave from of the ripple voltage in a full-wave capacitor-filtered rectifier.

lose an amount of charge $I_{dc}T_2$. Hence, the change in capacitor voltage is $I_{dc}T_2/C$ or

$$E_r = \frac{I_{dc}T_2}{C} \tag{13-23}$$

The value of $T_2$ is found from Fig. 13-13. This is the same diagram as in Fig. 13-11 except that the time origin is taken at the peak of the wave for convenience. It is also assumed that the rising straight-line section passes through the zero and the peak values of the sine curve. An inspection of the oscillograms of Fig. 13-10 shows that this is a reasonable approximation. From the geometry of Fig. 13-13, it follows that

$$\frac{T'}{T/4} = \frac{E_m - E_r}{E_m}$$

and

$$T_2 = T' + \frac{T}{4} = \left(1 - \frac{E_r}{E_m}\right)\frac{T}{4} + \frac{T}{4} = \frac{2E_m - E_r}{E_m}\frac{T}{4}$$

Using Eq. (13-21), this reduces to

$$T_2 = \frac{E_{dc}}{E_m}\frac{T}{2} \tag{13-24}$$

Hence, from Eq. (13-23)

$$E_r = \frac{I_{dc}}{C}\left(\frac{E_{dc}}{E_m}\frac{T}{2}\right) = \frac{I_{dc}}{2fC}\frac{E_{dc}}{E_m} \tag{13-25}$$

where $f = 1/T$ is the power supply frequency.

The d-c output is given by Eq. (13-21),

$$E_{dc} = E_m - \frac{E_r}{2} = E_m - \frac{I_{dc}}{4fC}\frac{E_{dc}}{E_m}$$

or

$$E_{dc} = \frac{E_m}{1 + I_{dc}/4fCE_m} \tag{13-26}$$

Figure 13-14 gives graphs of this equation for $f = 60$ cps, $C = 4\ \mu f$, and $E_m = 300\sqrt{2}$ and also $E_m = 450\sqrt{2}$. The experimentally measured
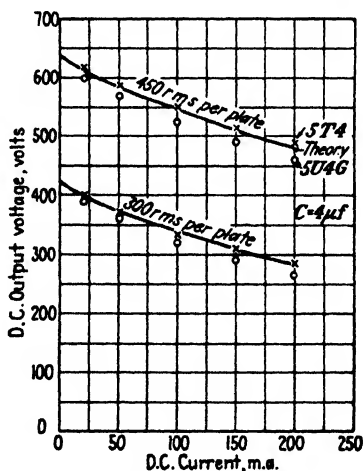


FIG. 13-14. Load voltage vs. load current for a capacitor input filter. $C = 4\ \mu f$. The crosses and circles are the experimentally measured points for a 5T4 and a 5U4G, respectively. These are taken from the RCA manual. The curves are plotted from Eq. (13-26).

values for the 5T4 and the 5U4G tubes are taken from the RCA tube manual and are also indicated on the same graph. The maximum deviation between experiment and theory is less than 10 per cent. This is satisfactory agreement, particularly in view of the simplifying assumptions made in the derivation.

From the graphs of Fig. 13-14 it is seen that the regulation of a capacitor filter is poor. The larger the capacitance, the better will be the regulation. The ripple factor is given by

$$r = \frac{E_{rms}'}{E_{dc}} = \frac{E_r}{2\sqrt{3}E_{dc}} = \frac{I_{dc}}{4\sqrt{3}fCE_m} \tag{13-27}$$

where use has been made of Eqs. (13-22) and (13-25). The ripple is seen to vary directly with the load current and inversely with the capacitance.

In order to keep the ripple low and to ensure good regulation, very large capacitors (of the order of tens of microfarads) must be used. The most common type of capacitor for this rectifier application is the electrolytic capacitor.[4] These capacitors are polarized, and care must be taken to insert them into the circuit with the terminal marked $+$ to the positive side of the output.

If a half-wave circuit is used, then the nonconduction time will be one-half a cycle longer, or

$$T_2 = \frac{T}{2} + \frac{T}{2}\frac{E_{dc}}{E_m} \tag{13-28}$$

Using this expression, the ripple and the output voltages are found to be given by the equations

$$r = \frac{I_{dc}}{4fC\sqrt{3}}\left(\frac{1}{E_m} + \frac{1}{E_{dc}}\right) \tag{13-29}$$

and

$$E_{dc} = \frac{E_m - I_{dc}/4fC}{1 + I_{dc}/4fCE_m} \tag{13-30}$$

A careful inspection of these equations reveals that the ripple of the half-wave circuit is approximately double that of the full-wave circuit. Also, the drop in voltage from no load to a given load for a half-wave circuit is approximately twice that for a full-wave circuit.

The particular features of rectifiers employing capacitor input filters are the small ripple and the high voltage at light load. The no-load voltage is equal, theoretically, to the maximum transformer voltage. The disadvantages of this system are the relatively poor regulation, the high ripple at large loads, and the peaked currents that the tubes must pass.

Refer to Fig. 13-7. It is seen that for a half-wave circuit the voltage across the tube on the inverse cycle is approximately twice the transformer maximum. In the full-wave circuit the inverse voltage across each tube is also twice the maximum transformer voltage measured from the mid-point to either end. However, this is independent of the form of the filter used and is a property of the full-wave circuit itself, and not of the filter system.

**13-5. L-section Filter.** The two types of filtering action considered above may be combined into a single L-section filter. This filter combines the decreasing ripple with increasing load of the series inductor with the increasing ripple with increasing load of the shunt capacitor. Such a filter is illustrated in Fig. 13-15. The inductor offers a high series impedance to the harmonic terms, and the capacitor offers a low shunt impedance to them. The resulting current through the load is smoothed out much more effectively than with either $L$ or $C$ alone in the circuit.
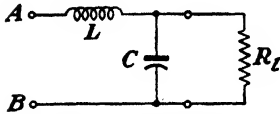


FIG. 13-15. An L-section filter.

The d-c voltage and the ripple factor are readily calculated by taking, for the voltage impressed at the terminals $AB$ of the filter of Fig. 13-15, the first two terms in the Fourier series representation of the output voltage of the rectifier, *viz.*,

$$e = \frac{2E_m}{\pi} - \frac{4E_m}{3\pi} \cos 2\omega t \qquad (13\text{-}31)$$

Thus, the two tubes are replaced by a battery in series with an a-c source having twice the power line frequency. This is the same equivalent circuit that was used in Sec. 13-2 for a full-wave inductor filter. If the resistance in series with the inductance is neglected, then the d-c output voltage equals the d-c input voltage, or

$$E_{dc} = \frac{2E_m}{\pi}$$

If the sum of the tube, transformer, and choke resistances is $R$, then

$$E_{dc} = \frac{2E_m}{\pi} - I_{dc}R \qquad (13\text{-}32)$$

The ripple will now be calculated. Since the object of the filter is to suppress the harmonic components in the system, the reactance of the choke must be large compared with the combined parallel impedance of capacitor and resistor. The latter combination is kept small by making the reactance of the capacitor much smaller than the resistance of the load. Very little error will be introduced, therefore, by assuming that the entire a-c current passes through the capacitor and none through the resistor. Under these conditions the net impedance across $AB$ is approximately $X_L = 2\omega L$, the reactance of the inductor at the second harmonic frequency. The a-c current through the circuit is

$$I_{rms}' = \frac{4E_m}{3\sqrt{2}\pi} \frac{1}{X_L} = \frac{\sqrt{2}}{3} E_{dc} \frac{1}{X_L} \qquad (13\text{-}33)$$

where the resistance $R$ in Eq. (13-32) has been neglected. The a-c voltage across the load (the ripple voltage) is the voltage across the capacitor. This is

$$E_{rms}' = I_{rms}'X_C = \frac{\sqrt{2}}{3} E_{dc} \frac{X_C}{X_L} \qquad (13\text{-}34)$$

where $X_C = 1/(2\omega C)$ is the reactance of the capacitor at the second harmonic frequency. The ripple factor is then given by

$$r = \frac{E_{rms}'}{E_{dc}} = \frac{\sqrt{2}}{3} \frac{X_C}{X_L}$$

$$= \frac{\sqrt{2}}{3} \left(\frac{1}{2\omega C}\right)\left(\frac{1}{2\omega L}\right) \qquad (13\text{-}35)$$

which is, at 60 cps,

$$r = \frac{0.83}{LC} \qquad (13\text{-}36)$$

with $C$ in microfarads and $L$ in henrys.

It is noticed that the effect of combining the decreasing ripple arising with a simple inductor filter and the increasing ripple arising with a simple capacitor filter for increasing loads is a constant ripple circuit independent of load.

The foregoing analysis assumes that a current flows through the circuit at all times. If any cutout points of the type discussed in the previous section exist, this analysis is no longer valid. Consider the conditions that exist when no inductor is used. As already found, current will flow in the tube circuit for a small portion of the cycle and the capacitor will become charged to the peak transformer voltage in each cycle. Suppose that a small inductance is now inserted in the line. Although the time over which tube current



FIG. 13-16. Oscillograms of the tube current in a single-phase full-wave circuit when an L-section filter is used. (a) $L$ less than the critical inductance $L_c$; (b) $L$ equal to $L_c$; (c) $L$ greater than $L_c$. The oscillograms were taken with a fixed value of $L$, and the load resistance was varied.

will exist will be somewhat lengthened, cutout may still occur. As the value of the inductance is increased, a value will be reached for which the tube circuit supplies current to the load continuously, and no cutout occurs. This value of inductance is referred to as the critical inductance $L_c$. Under these circumstances, each tube conducts for its

normal portion of the cycle, and the input voltage to the filter circuit has the form given by Eq. (13-31). It is only under these circumstances that the above-developed L-section filter theory is applicable. The foregoing discussion is illustrated in the oscillograms of Fig. 13-16.

If the rectifier is to pass current throughout the entire cycle, the peak of the a-c component of the current must not exceed the d-c current. But the d-c current is $E_{dc}/R_l$. Also, the peak of the a-c component of the current is $(2E_{dc}/3)(1/X_L)$. The critical condition illustrated in Fig. 13-16b occurs when the d-c current just equals the peak of the a-c component. Therefore, for current to pass during the entire cycle, it is necessary that

$$\frac{E_{dc}}{R_l} \geq \frac{2E_{dc}}{3} \frac{1}{X_L}$$

or

$$X_L \geq \frac{2R_l}{3}$$

The value for the critical inductance is given by

$$L_c = \frac{R_l}{3\omega} \tag{13-37}$$

For a 60-cps input frequency, this becomes

$$L_c = \frac{R_l}{1,130} \tag{13-38}$$

where $R_l$ is expressed in ohms and $L_c$ is in henrys.

It must be remembered that these values of critical inductance have been based not upon the true input voltage but rather upon an approximate voltage that is made up of the d-c term and the first a-c harmonic term in the Fourier series of the true input voltage. It was shown in Sec. 13-2 that this approximation introduced very little error in the calculation of the ripple factor. However, the neglect of the higher harmonic terms introduces an appreciable error in the calculation of the critical inductance.* It is advisable for conservative design to increase the values of $L_c$ calculated from Eq. (13-38) by about 25 per cent.

The effect of the cutout is illustrated in Fig. 13-17, which shows a regulation curve of the system for constant $L$ and a varying load resistor. Clearly,

---

* If $A$ represents the amplitude of the first a-c term in the Fourier series of a wave and $B = 0.1A$ represents the amplitude of the second term, then $A + B = 1.1A$. A 10 per cent error is made if $B$ is neglected in calculating the sum. It is this general process that is involved in the calculation of $L_c$. However, the calculation of the ripple factor requires an evaluation of an expression of the form $\sqrt{A^2 + B^2} = \sqrt{A^2 + (0.1A)^2} = 1.005A$. Hence, if $B$ is neglected, this results in an error of only 0.5 per cent.

when the current is zero ($R_l$ is infinite) the filter is of the simple capacitor type, and the output voltage is $E_m$. With increasing load current, the voltage falls, until at $I = I_c$ (the current at which $L = L_c$) the output potential is that corresponding to the simple L filter with no cutout, or $0.636E_m$. For values of $I$ greater than $I_c$, the change in potential results from the effects of the resistances of the various elements of the circuit.

It is not possible to satisfy the conditions of Eq. (13-37) for all values of load, since at no load this would require an infinite inductance. If good voltage regulation is essential, it is customary to use a "bleeder" resistor in parallel with the load so as to maintain the conditions of Eq. (13-37), even if the useful current is small. This ensures that the inductor exerts its influence in the filter even at zero useful load. That is, $R_l$



Fig. 13-17. The regulation curve of a rectifier with an L-section filter.

in Eq. (13-37) is actually the bleeder resistor and the load resistor in parallel.

A more efficient method than using a bleeder resistor, with its consequent power dissipation, is to make use of the fact that the inductance of an iron-core reactor depends, among other things, upon the magnitude of the d-c current in the winding. Reactors for which the inductance is high at low values of d-c current and which decrease markedly with increased d-c currents are called "swinging chokes." A typical curve for such a reactor is illustrated in Fig. 13-18. The advantage of using such a choke is evident,



Fig. 13-18. The inductance of a typical "swinging choke" as a function of the d-c current through it.

for the condition imposed by Eq. (13-37) is that the inductance should increase with a decrease in load current in order that the filtering action should continue. An inspection of Eq. (13-34) indicates that the ripple will increase with increased load when a swinging choke is used since $X_L$ decreases with increasing loads.

In designing an L-section filter, an inductor must be chosen so as to satisfy Eq. (13-38) for the specified bleeder resistor. Then a capacitor is chosen at least as large as that determined from Eq. (13-36) for the specified tolerable ripple. If a swinging choke is used, the minimum value of its inductance must be used in the calculation of the capacitor value needed.

*Example.* A single-phase full-wave rectifier is to supply 100 ma at 350 volts with a ripple that must be less than 10 volts. Specify the elements of a rectifier using a single L-section filter that will provide the desired results.

*Solution.* The effective resistance of the load is

$$R_l = \frac{350}{0.100} = 3{,}500 \text{ ohms}$$

The ripple factor is

$$r = \tfrac{10}{350} = 0.0286$$

According to Eq. (13-38), the critical inductance for such a filter is

$$L_c = \frac{3{,}500}{1{,}130} = 3.1 \text{ henrys}$$

According to Eq. (13-36), the product $LC$ must be at least as large as

$$LC = \frac{0.830}{0.0286} = 29$$

These calculations specify the minimum values of $L$ and $LC$ that may be used to accomplish the desired filtering. The actual values that will be used are determined by the commercially available inductors and capacitors. The desirability of using standard commercial merchandise is dictated both by availability and by economic considerations.

Since 10-henry chokes having the desired current rating are readily available, such an inductor will be chosen. The capacitor must therefore be about 3 $\mu$f. Since a 4-$\mu$f capacitor is readily available commercially, it will be chosen.

A search through a tube manual reveals several rectifier tubes having the proper ratings. One such tube is the 5Y3G with a maximum d-c output current rating of 125 ma, a maximum plate voltage per plate of 500 volts rms, and a peak inverse voltage of 1,400 volts. The plate characteristic of this tube shows that the tube voltage is 50 volts at 100 ma, corresponding to a resistance of approximately 500 ohms.

The resistance of a 10-henry choke capable of carrying 100 ma is found in a manufacturer's catalogue. A reasonable value is 200 ohms. Similarly, a reasonable value for the transformer resistance is 200 ohms. Hence the total resistance in series with the inductor is $R = 500 + 200 + 200 = 900$ ohms. From Eq. (13-32) the peak transformer voltage is

$$E_m = \frac{\pi}{2}(E_{dc} + I_{dc}R) = \frac{\pi}{2}[350 + (0.1)(900)] = 690 \text{ volts}$$

and

$$E_{rms} = \frac{690}{\sqrt{2}} = 488 \text{ volts}$$

A stock transformer would be purchased whose current rating is at least 100 ma and whose voltage to center tap is close to 488 volts, say 500 volts. If the exact value of choke and transformer resistances were known, a more accurate calculation of the transformer voltage needed could be made.

The peak inverse voltage is $2E_m = (2)(690) = 1,380$ volts. Since the rated peak inverse voltage of the 5Y3G tube is 1,400 volts, it is safe to use this tube in this application.

If the load should be removed accidentally, the circuit will behave as if it had a capacitor input filter and the voltage will rise to the peak transformer voltage to center tap or to 690 volts. Hence the insulation rating of the filament heating transformer should be at least this high.

The inequalities used in the derivation of the expression for the ripple factor will now be checked. $R_l = 3,500; X_L = 4\pi fL = 7,540;$ and $X_C = 1/(4\pi fC) = 332$ ohms. Hence $X_C \ll R_l$, and $X_L \gg X_C$, thus verifying the inequalities assumed.

**13-6. L-section Filter with a Controlled Rectifier.** The above analysis for the critical inductance of the L-section filter applies for a full-wave circuit for which conduction continues for 180 deg in each cycle. Consequently, the results so obtained are not applicable when an L-section filter is used with a controlled rectifier.

The analysis for a full-wave controlled rectifier is considerably more complicated than that above. This greater complexity arises from the fact that the harmonic components contained in the Fourier series of the output voltage from such a unit are larger than the corresponding terms in the output from a rectifier without control and depend on the delay angle. The Fourier series representing the output from a single-phase full-wave controlled rectifier with no cutout may be shown to be

$$e_l = E_{dc} \left[ 1 - \sum_{k=2,4,6,\ldots} \frac{2\sqrt{1 + k^2 \tan^2 \varphi}}{k^2 - 1} \cos (k\alpha - \psi_k) \right] \quad (13\text{-}39)$$

where the d-c output voltage is

$$E_{dc} = \frac{2E_m}{\pi} \cos \varphi \qquad (13\text{-}40)$$

and where $\varphi$ is the delay angle. The phase-shift angle $\psi_k$ is given in Sec. 14-9.

The results of an investigation that was carried out by Overbeck [5] for the critical inductance for controlled full-wave rectifiers for various delay angles are given in Fig. 13-19. These results are in approximate agreement with Eq. (13-38) for the case of zero delay angle.

In the calculation of the ripple, it is found that Eqs. (13-35) and (13-36) are modified only to the extent that the right-hand side of these equations must be multiplied by $\sqrt{1 + 4 \tan^2 \varphi}$. In other words the ripple is increased by this factor.
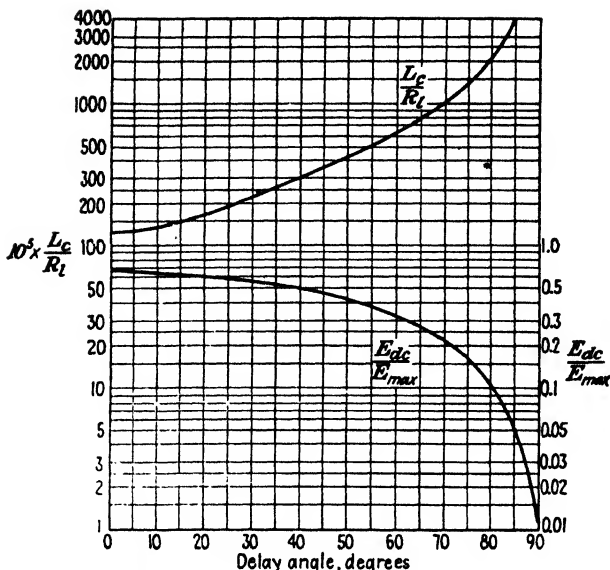
FIG. 13-19. Critical inductance and d-c output voltage as a function of the delay angle in a full-wave rectifier. (*After Overbeck.*)

**13-7. Multiple L-section Filters.** The filtering may be made much more complete through the use of two L-section filters in cascade, as shown in Fig. 13-20. An approximate solution that is sufficiently accurate for practical purposes can be obtained by proceeding according to the development in Sec. 13-5.
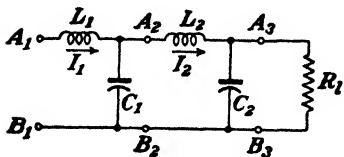


FIG. 13-20. A multiple (two-section) L-section filter.

It is assumed that the reactances of all the chokes are much larger than the reactances of the capacitors. Also, it is assumed that the reactance of the last capacitor is small compared with the resistance of the load. Under these circumstances, the impedance between $A_3$ and $B_3$ is effectively $X_{C_2}$. The impedance between $A_2$ and $B_2$ is effectively $X_{C_1}$, and the impedance between $A_1$ and $B_1$ is effectively $X_{L_1}$. The a-c current $I_1$ is approximately

$$I_1 = \frac{\sqrt{2}E_{dc}}{3}\frac{1}{X_{L_1}}$$

The a-c voltage across $C_1$ is approximately

$$E_{A_2B_2} = I_1 X_{C_1}$$

The a-c current $I_2$ is approximately

$$I_2 = \frac{E_{A_2B_2}}{X_{L_2}}$$

The a-c voltage across the load is approximately

$$I_2 X_{C_2} = I_1 \frac{X_{C_2} X_{C_1}}{X_{L_2}} = \frac{\sqrt{2}E_{dc}}{3}\left(\frac{X_{C_2}}{X_{L_2}}\right)\left(\frac{X_{C_1}}{X_{L_1}}\right)$$

The ripple factor is given by dividing this expression by $E_{dc}$. Hence

$$r = \frac{\sqrt{2}}{3}\frac{X_{C_1}X_{C_2}}{X_{L_1}X_{L_2}} \tag{13-41}$$

A comparison of this equation with Eq. (13-35) indicates the generalization which should be made in order to obtain an expression valid for any number of sections. For example, a multiple L filter of $n$ similar sections has a ripple factor that is given by

$$r = \frac{\sqrt{2}}{3}\left(\frac{X_C}{X_L}\right)^n = \frac{\sqrt{2}}{3}\frac{1}{(16\pi^2 f^2 LC)^n} \tag{13-42}$$

For a multiple L filter of $n$ similar sections, the product $LC$ (henrys $\times$ microfarads) for a specified ripple factor may be evaluated from

$$LC = 1.76\left(\frac{0.471}{r}\right)^{1/n} \tag{13-43}$$

To the approximation that the impedance between $A_1$ and $B_1$ is simply $X_{L_1}$, the critical inductance is the same for the first inductor of a multisection filter as for a single-section unit. The remaining inductors may have any values, since they play no part in determining the cutout condition.

**13-8. Π-section Filters.** A very smooth output may be obtained by using a filter that consists of two capacitors separated by an inductor, as shown in Fig. 13-21. Such filters are characterized by highly peaked tube currents and by poor reg-



FIG. 13-21. A Π-section filter.

ulation as for the simple capacitor input filter. They are used if, for a given transformer, higher voltage than can be obtained from an L-section filter is needed, and if lower ripple than can be obtained from a simple capacitor or an L-section filter is desired.
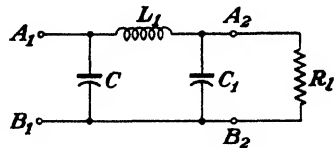
The oscillogram of Fig. 13-22 clearly shows the improved filtering action of a Π-type filter over an L-section filter. Curve $a$ gives the output wave form from the single-phase full-wave rectifier without any filter in the cir-

cuit. Curve *b* shows the output when a capacitor shunts the load. This curve corresponds to that of Fig. 13-10*b*. If an inductor is inserted in the line, resulting in an L-section filter, a fairly smooth output voltage is obtained. This is illustrated in curve *c*. It should be noted that the presence of the inductor causes the output voltage to be equal to the average value of output rectified voltage. The addition of a second capacitor, which results in the Π-type filter, produces the extremely smooth output of curve *d*.
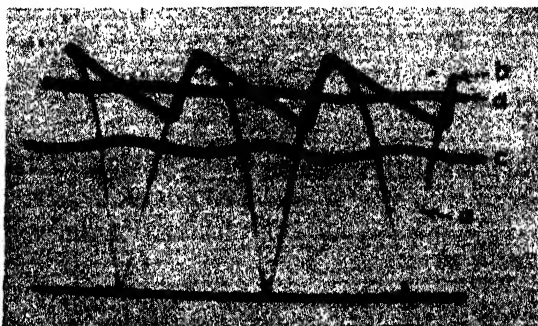


Fɪɢ. 13-22. Oscillograms of the output voltage with various filters for a single-phase full-wave rectifier. (*a*) No filter. (*b*) A capacitor filter. (*c*) An L-section filter. (*d*) A Π-section filter.

The action of this filter can best be understood by considering the inductor and the second capacitor as an L-section filter that acts upon the triangular output-voltage wave from the first capacitor. The output potential is then approximately that from the input capacitor [Eq. (13-26)], decreased by the $IR_L$ drop in the inductor. The ripple contained in this output is reduced by the L-section filter.

The ripple voltage can be calculated by analyzing the triangular wave of Fig. 13-12 into a Fourier series and then multiplying each component by $X_{C_1}/X_{L_1}$ for this harmonic. This leads to rather involved expressions. An upper limit to the ripple can, however, be more easily obtained. If it is assumed that cutout takes place for the entire half cycle (for a full-wave rectifier), then Fig. 13-12 becomes a triangular wave with vertical sides. The Fourier analysis of this wave form is given by

$$e = E_{dc} - \left(\frac{E_r}{\pi}\right)\left(\sin 2\omega t - \frac{\sin 4\omega t}{2} + \frac{\sin 6\omega t}{3} - \cdots\right) \quad (13\text{-}44)$$

From Eq. (13-23), $E_r = I_{dc}T_2/C$, and since it has now been assumed that $T_2 = T/2$, this reduces to

$$E_r = \frac{I_{dc}T}{2C} = \frac{I_{dc}}{2fC}$$

The rms second harmonic voltage is

$$E_2' = \frac{E_r}{\pi\sqrt{2}} = \frac{I_{dc}}{2\pi fC\sqrt{2}} = \sqrt{2}I_{dc}X_C \qquad (13\text{-}45)$$

where $X_C$ is the reactance of $C$ at the second harmonic frequency.

A second method of obtaining the same result, due to Arguimbau,[6] is instructive. If the instantaneous current to the filter is $i$, then the rms second harmonic current $I_2'$ is given by the Fourier component

$$\sqrt{2}I_2' = \frac{1}{\pi}\int_0^{2\pi} i\cos 2\alpha \, d\alpha$$

The current $i$ is in the form of pulses near the peak value of the cosine curve, and hence not too great an error is made by replacing $\cos 2\alpha$ by unity. Since the maximum value of the cosine is unity, this will give the maximum possible value of $I_2'$. Thus,

$$\sqrt{2}I_2' \le \frac{1}{\pi}\int_0^{2\pi} i \, d\alpha = 2I_{dc}$$

because, by definition,

$$I_{dc} \equiv \frac{1}{2\pi}\int_0^{2\pi} i \, d\alpha$$

Hence, the upper limit of the rms second harmonic voltage is

$$E_2' = I_2'X_C = \sqrt{2}I_{dc}X_C$$

which agrees with the first method of analysis in which it was assumed that the cutout took place over the complete half cycle. If this were true, then the charging current could exist only for an infinitesimally small time near the peak of the input voltage or at the points for which $\cos 2\alpha = 1$. This shows the consistency of the two methods of attack.

The voltage $E_2'$ is impressed on an L section, and by using the same logic as in Sec. 13-5 the output ripple is $E_2'X_{C_1}/X_{L_1}$. Hence, the ripple factor is

$$r = \frac{E_{rms}'}{E_{dc}} = \frac{\sqrt{2}I_{dc}X_C}{E_{dc}}\frac{X_{C_1}}{X_{L_1}} = \sqrt{2}\left(\frac{X_C}{R_l}\right)\left(\frac{X_{C_1}}{X_{L_1}}\right) \qquad \cdot (13\text{-}46)$$

where all reactances are calculated at the second harmonic frequency. This is the second harmonic ripple but, just as for the simple inductor filter, very little error is made in neglecting the higher harmonics and considering this as the total ripple.

For 60 cps Eq. (13-46) reduces to

$$r = \frac{3,300}{CC_1L_1R_l} \qquad (13\text{-}47)$$

where the capacitances are in microfarads, the inductances in henrys, and the resistances in ohms.

If the Π section is followed by an L section whose parameters are $L_2$ and $C_2$, then the above reasoning leads to the expression

$$r = \sqrt{2}\left(\frac{X_C}{R_l}\right)\left(\frac{X_{C_1}}{X_{L_1}}\right)\left(\frac{X_{C_2}}{X_{L_2}}\right) \tag{13-48}$$

This can be extended in an obvious fashion to include any number of sections.

If a half-wave circuit is used, then it can be shown that Eqs. (13-46) and (13-48) are still valid provided that all reactances are calculated at the fundamental instead of the second harmonic frequency. Thus, for a single Π section, the half-wave ripple is eight times that for a full-wave circuit. The d-c output voltage is that corresponding to the half-wave simple capacitor filter [Eq. (13-30)], minus the $I_{dc}R_L$ drop in the inductor.

*Example.* Design a power supply using a Π-section filter to give d-c output of 250 volts at 50 ma with a ripple factor not to exceed 0.01 per cent.

*Solution.* The load resistance is $R_l = 250/(50 \times 10^{-3}) = 5,000$ ohms.

From Eq. (13-47), with $C = C_1$, $r = 3,300/C^2LR_l$ or

$$C^2L = \frac{3,300}{10^{-4} \times 5,000} = 6,600$$

There is no unique way of solving this equation for $C$ and $L$. A reasonable commercially available value of $L$ will be chosen, and then $C$ will be calculated. The Thordarson choke type T20C53 has an inductance of 20 henrys at 50 ma. Its d-c resistance is 375 ohms. If this is used, then the corresponding required capacitors have values

$$C = \left(\frac{6,600}{20}\right)^{\frac{1}{2}} = 18.1 \ \mu f$$

Electrolytic capacitors are available in this range. For example, the Cornell-Dubilier type BR 2045 has a rating of 20 $\mu f$ at 450 volts d-c and would be suitable.

The d-c voltage drop in the choke is $(50 \times 10^{-3})(375) = 19$ volts. Hence, the d-c voltage across the first capacitor is $250 + 19 = 269$ volts. The peak transformer voltage to center tap $E_m$ is given by Eq. (13-26),

$$E_{dc} = \frac{E_m}{1 + I_{dc}/4fCE_m}$$

or

$$269 = \frac{E_m^2}{E_m + \dfrac{50 \times 10^{-3}}{4 \times 60 \times 20 \times 10^{-6}}} = \frac{E_m^2}{E_m + 10.4}$$

solving $E_m = 278$ and $E_{rms} = E_m/\sqrt{2} = 196$. Hence a 200-0-200-volt transformer would be used. A suitable tube for this application would be the 6X5. It is rated at 70 ma maximum output current and a peak inverse voltage of 1,250. The peak inverse for this circuit is $2E_m = 556$, which is well within the tube rating.

**13-9. Π-section Filter with a Resistor Replacing the Inductor.** This type of filter is analyzed in the same manner as above. The d-c output is

the value given in Eq. (13-26) for a simple capacitor filter minus the $I_{dc}R$ drop in the resistor. The ripple factor is given by Eq. (13-46) with $X_L$ replaced by $R$. Thus, for a single section

$$r = \sqrt{2} \frac{X_C}{R_l} \frac{X_{C_1}}{R} \tag{13-49}$$

Hence, if the resistor $R$ is chosen equal to the reactance of the choke which it replaces, the ripple remains unchanged. Since this means a saving in the expense, weight, and space of the choke, it is desirable to use the resistor wherever possible. This can be done, however, only for low-current power supplies. Thus, for example, if in a full-wave circuit with an output current of 100 ma, a 20-henry choke is to be replaced by a resistor to give the same ripple, its value must be $R = X_L = 4\pi f L = 15,000$ ohms. The voltage drop in this resistor would be $(15,000)(0.1) = 1,500$ volts! The d-c power dissipated would be $I_{dc}{}^2 R = (0.1)^2(15,000) = 150$ watts! Hence such a substitution would not be a sensible one. However, if the rectifier is to furnish only 10 ma (perhaps for a cathode-ray tube supply) then the drop in the resistor is only 150 volts and the power loss in this resistor is 1.5 watts. The resistor rather than the inductor should be used in such an application.

**13-10. Summary of Filters.** Table 13-1 contains a compilation of the more important information relating to the various types of filters, when used with single-phase full-wave circuits. In all cases tube, transformer, and filter-element resistances are considered negligible, and a 60-cycle power line is assumed. $C$ is expressed in farads, $L$ in henrys, and $R$ in ohms.

TABLE 13-1

SUMMARY OF FILTER INFORMATION

| Filter | None | L | C | L section | Π section |
|---|---|---|---|---|---|
| $E_{dc}$—no load | $0.636E_m$ | $0.636E_m$ | $E_m$ | $E_m$ | $E_m$ |
| $E_{dc}$—load $I_{dc}$ | $0.636E_m$ | $0.636E_m$ | $\dfrac{E_m}{1 + \dfrac{I_{dc}}{240CE_m}}$ | $0.636E_m$ | $\dfrac{E_m}{1 + \dfrac{I_{dc}}{240CE_m}}$ |
| Ripple factor $r$ | 0.48 | $\dfrac{R_l}{1,600L}$ | $\dfrac{I_{dc}}{416CE_m}$ | $\dfrac{0.83 \times 10^{-6}}{LC}$ | $\dfrac{3,330 \times 10^{-12}}{CC_1L_1R_l}$ |
| Peak inverse | $2E_m$ | $2E_m$ | $2E_m$ | $2E_m$ | $2E_m$ |

## PROBLEMS

**13-1.** *a.* Prove that the general solution of the differential equation in Eq. (13-5) is

$$i_b = \frac{E_m}{\sqrt{R_l^2 + \omega^2 L^2}} \left[ \sin(\omega t - \Psi) + e^{-R_l t/L} \sin \Psi \right]$$

where $\tan \Psi = \omega L/R_l$.

*b.* The angle of cutout $\omega t_2$ is that angle at which the current becomes zero. Show that at cutout

$$\sin(\omega t_2 - \Psi) + e^{-(R_l/\omega L)\omega t_2} \sin \Psi = 0$$

Plot a semilog curve of $\omega t_2$ vs. $\omega L/R_l$, with $\omega L/R_l$ in the range from 0.1 to 1,000.

*c.* Verify the curves of Fig. 13-2. In particular, check the value for $\omega L/R_l = 5$.

**13-2.** A single-phase full-wave rectifier uses gas diodes. The tube drop and internal resistance of the tubes may be neglected. Assume an ideal transformer.

*a.* Prove that one tube conducts for one half cycle and that the other tube conducts for the remaining half cycle of the input line voltage, if the load consists of a resistor $R$ in series with an inductor $L$.

*b.* Find the analytic expression for the load current in the interval

$$0 \leq \alpha = 2\pi ft \leq \pi$$

HINT: Set up the differential equation for the load current $i$ in this interval. The solution of this equation will consist of a steady-state a-c term added to a "transient" term. Evaluate the arbitrary constant in the "transient" term by noting that the current repeats itself at intervals of $\pi$ in $\alpha$, so that $i(0) = i(\pi)$.

*c.* Evaluate the d-c current $I_{dc}$ by averaging the instantaneous current.

*d.* Evaluate the first term in the Fourier series for the current, and compare with Eq. (13-6). Compare this method of attack with that used in Sec. 13-2.

**13-3.** Prove that the rms value of the triangular voltage depicted in Fig. 13-12 is given by Eq. (13-22).

**13-4.** A single-phase full-wave rectifier uses an 83 gas tube. The transformer voltage is 350 volts rms to center tap. The load consists of a 4-$\mu f$ capacitor in parallel with a 2,500-ohm resistor. The tube drop and the transformer resistance and leakage reactance may be neglected.

*a.* Calculate the cutout angle.

*b.* Plot to scale the output voltage and the tube current (see Fig. 13-9). Determine the cutin point graphically from this plot, and find the peak tube current corresponding to this point.

*c.* Repeat parts *a* and *b*, using a 16- instead of a 4-$\mu f$ capacitor.

**13-5.** *a.* Show that Eq. (13-26) reduces to

$$E_{dc} = E_m - \frac{I_{dc}}{4fC}$$

provided that

$$\frac{I_{dc}}{4fCE_m} \ll 1$$

*b.* Show that this result is obtained if it is assumed that the capacitor discharges for the complete half cycle $T/2$.

**13-6.** The circuit of Fig. 13-15 can be analyzed by the methods of elementary a-c theory without making the approximations used in Sec. 13-5. Assuming that the input

voltage to the filter is given by Eq. (13-31), prove that the ripple factor is

$$r = \frac{\sqrt{2}/3}{\sqrt{\left(\dfrac{X_L}{R_l}\right)^2 + \left(\dfrac{X_L}{X_C} - 1\right)^2}}$$

Under what condition does this reduce to the simpler equation (13-35)?

**13-7.** By error the capacitor of an L-section filter is connected to the input side of the inductor. Examine this filter analytically, and derive an expression for

a. The regulation of the system.

b. The ripple factor.

Compare these results with those in Sec. 13-5.

**13-8.** The output of a full-wave rectifier is fed from a 400-0-400-volt transformer. The load current is 0.1 amp. Two 4-μf capacitors are available. The tube drop of each gas tube is 15 volts. The circuit resistance exclusive of the load is 500 ohms.

a. Calculate the value of inductance for a two-stage L-section filter. The inductances are to be equal. The ripple factor is to be 0.0001.

b. Calculate the d-c output voltage.

**13-9.** Given two equal capacitors $C$ and two equal inductors $L$. Under what circumstances will it be better to use a double L-section filter than to use a single section with the inductors in series and the capacitors in parallel?

**13-10.** An L-section filter is used in the output of a full-wave rectifier that is fed from a 375-0-375-volt transformer. The load current is 0.2 amp. Two 4-μf capacitors and two 20-henry chokes are available. The drop in the gas rectifier tubes is 15 volts.

a. Calculate the 120-cycle ripple voltage, if a single-section filter is used, with the two chokes in series and two capacitors in parallel.

b. Repeat part a for a two-section filter.

c. Calculate the 240-cycle ripple voltage if a single-section filter is used.

**13-11.** Design a power supply for a load varying from zero to 150 ma at a full-load output voltage of approximately 250 volts. Good regulation is desired. The ripple voltage is not to exceed 0.1 volt.

Specify the type of circuit, the type of filter and the type of tube. Give nominal ratings of all the circuit elements used.

Assume

a. Vacuum-tube rectifiers have a plate resistance of 350 ohms.

b. Gas-tube rectifiers have a tube drop of 10 volts.

c. Power transformers are rated in steps of 50 volts (*i.e.*, you can buy a transformer whose total secondary voltage output is 50 volts, or 100 volts, or 150 volts, etc. rms).

d. The transformer secondary resistance totals 200 ohms.

e. Chokes are rated in steps of 5 henrys.

f. Each choke resistance is 200 ohms.

g. Capacitors are rated in 4-μf. steps.

**13-12.** Design a power supply for a load that varies between the limits of 20 and 100 ma at a nominal voltage of 300 volts. The ripple factor is not to exceed 0.1 per cent, and good regulation is desired. Specify the type of circuit, the type of tube, and the type of filter to use. Give nominal ratings of all the circuit elements used. Make the same assumptions as in Prob. 13-11.

**13-13.** Given a full-wave rectifier circuit, a 375-0-375 volt transformer, $R_l = 2,000$ ohms, gas diodes with a 20-volt drop, two 20-henry chokes and two 16-μf capacitors. The transformer resistance to center tap and each choke resistance is 200 ohms. Calculate the approximate output voltage and ripple factor under the following filter arrangements:

a. The two chokes are connected in series with the load.

*b.* The two capacitors are connected in parallel across the load.

*c.* A single-section L filter, consisting of the two inductors in series and the two capacitors in parallel.

*d.* A two-section L filter.

*e.* A Π-section filter, using both inductors.

**13-14.** Derive an expression for the ripple in a Π-section filter when used with a half-wave rectifier, subject to the same approximations as those in Sec. 13-8 for the full-wave case.

**13-15.** A full-wave single-phase rectifier employs a Π-section filter consisting of two 4-μf capacitors and a 20-henry choke. The transformer voltage to center tap is 300 volts rms. The load current is 50 ma. Calculate the d-c output voltage and the ripple voltage. The resistance of the choke is 200 ohms.

**13-16.** A single center-tapped transformer (350-0-350 volts) is to supply power at two different voltages for certain service. The negative is to be grounded on each system. The low voltage is full wave and is filtered with a two-section L filter. The high voltage is half wave and has a capacitor input filter. Show the schematic diagram for such a system. What is the nominal output voltage of each unit?

**13-17.** What voltages are available from the rectifier circuit shown? A 425-0-425-volt transformer is used. Label the polarities of the output voltages.
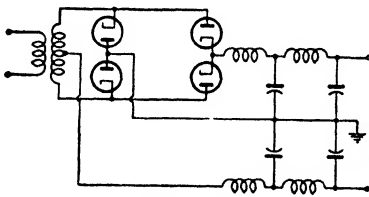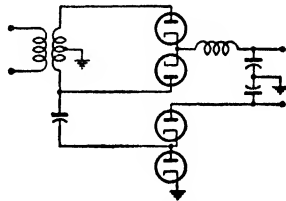


PROB. 13-17.  PROB. 13-18.

**13-18.** The circuit shown is to be used to supply power for an amplifier and also for the accelerating voltage of an associated cathode-ray tube. What output voltages are obtained if a 350-0-350-volt center-tapped transformer is used? (HINT: See Prob. 12-15.)
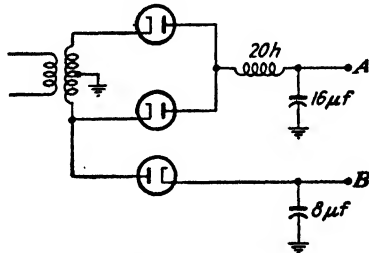


PROB. 13-19.

**13-19.** The circuit shown operates from a 300-0-300-volt transformer.

*a.* What are the magnitude and polarity of the d-c voltage at *A*; at *B*; under no load?

*b.* What is the peak inverse on each tube?

*c.* If the load current at *A* is 100 ma, what is the voltage at *A*?

*d.* If the load current at *B* is 20 ma, what is the voltage at *B*?

## REFERENCES

1. Electrical Engineering Staff, Massachusetts Institute of Technology, "Electric Circuits," Chap. XIII, John Wiley & Sons, Inc., New York, 1940.
   PREISMAN, A., "Graphical Constructions for Vacuum Tube Circuits," McGraw-Hill Book Company, Inc., New York, 1943.
2. STOUT, M. B., *Elec. Eng.*, **54,** 977, 1935.
3. WAIDELICH, D. L., *Trans. AIEE*, **60,** 1161, 1941.
   SCHADE, O. H., *Proc. IRE*, **31,** 341, 1943.
4. DEELEY, P. M., "Electrolytic Capacitors," The Cornell-Dubilier Electric Corp., South Plainfield, N.J., 1938.
   BROTHERTON, M., "Capacitors," D. Van Nostrand Company, Inc., New York, 1946.
5. OVERBECK, W. P., *Proc. IRE*, **27,** 655, 1939.
6. ARGUIMBAU, L., "Vacuum Tube Circuits," John Wiley & Sons, Inc., New York, 1948.

### General References

Dow, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.
DUNHAM, C. R.: *IEEJ*, **75,** 278, 1934.
EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," 3d ed., McGraw-Hill Book Company, Inc., New York, 1949.
LEE, R.: "Electronic Transformers and Circuits," John Wiley & Sons, Inc., New York, 1947.
REICH, H. J.: "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.

# CHAPTER 14

# POLYPHASE RECTIFIERS

THE single-phase rectifiers which are discussed in some detail in Chap. 12 are used extensively to provide sources of rectified power at voltages to several kilovolts, and for currents to several amperes. If larger powers are required, it is customary and desirable to employ polyphase rectifiers for the purpose. Such polyphase systems are used for supplying power for railway systems, for supplying the large d-c currents required for electroplating, for supplying the moderate-current high voltage required for the plate circuits of radio transmitters, or, in fact, for any application that requires large d-c currents.

A number of reasons exist for preferring polyphase rectifiers for high power service. One reason is that most a-c power is generated and distributed as three-phase power, and a rectifying system which operates directly from the three-phase lines is clearly desirable. Second, it will be seen in what follows that the ripple in the output of a polyphase rectifier which operates without a filter decreases with an increase in the effective number of phases. Moreover the lowest harmonic that exists in the output of a rectifier is directly dependent on the number of effective phases. The larger the number of phases, the higher is the first harmonic term, with the consequent ease of filtering with a simple and generally inexpensive filter. Third, the transformers and associated equipment are utilized to better advantage in certain polyphase circuits than with single-phase circuits, so that for a given rectifier output, the rating of the auxiliary equipment is smaller than with the single-phase system.

Although the primary source of power is usually a three-phase system, considerable advantage may be realized by using more than three-phases. The desired phase transformation is effected by means of transformers to yield 6-phase, 12-phase, or other polyphase power. The analysis to follow will be based on a 3-phase primary power source, but the results will apply to a $p$-phase rectified output.

Because of the greater current capacity of gas tubes and the fact that the tube drop remains substantially constant during the conducting portion of the cycle (thereby providing a high and substantially constant efficiency), polyphase rectifiers seldom, if ever, employ vacuum diodes. If moderate values of current are required at high voltage, mercury-vapor

diodes are usually employed.  If high currents are required at low voltage, Se or CuO rectifier stacks may be used.  For high-current sources at moderate values of voltage, polyphase pool-cathode tanks or banks of ignitrons or excitrons are employed.  The general operating theory is substantially the same in either case.

The use of rotating machinery, either the synchronous converter or the synchronous motor-generator set, for converting from a-c power into d-c
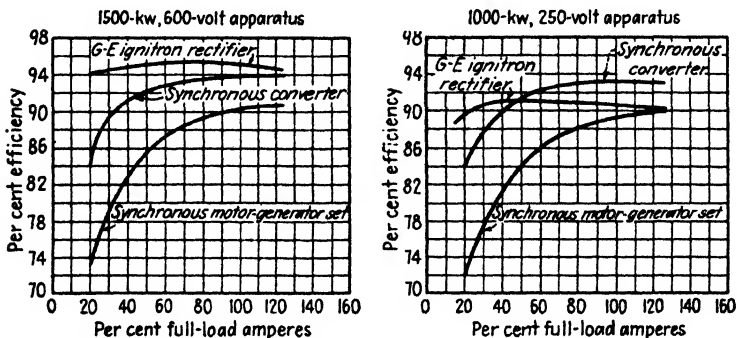


FIG. 14-1.　Curves showing the relative efficiencies of comparable capacity conversion systems for two different output voltages.  (*Courtesy of General Electric Co.*)

power is very widespread.  The use of metallic rectifier stacks for low-voltage high-current service has become increasingly important during the past decade.  Also, the use of mercury pool rectifier systems is gradually displacing the use of rotating machinery for the larger capacities.  The efficiency curves given in Fig. 14-1 will indicate one of the reasons for this.
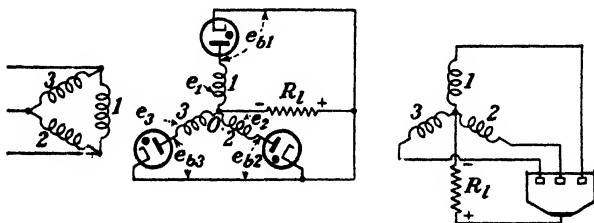


FIG. 14-2.　Three-phase half-wave rectifier connections using either gas-filled tubes or a pool-cathode tank.

**14-1. Elementary Theory.**  The circuit of the three-phase half-wave rectifying system for either diode or tank operation is given in Fig. 14-2. *The corresponding primary and secondary transformer windings are shown parallel to each other*, a method of representation that will be adhered to throughout the text.  The symbol $e_s$ represents the voltage at anode $S$ with respect to the neutral $O$ of the transformer secondary.

The operation of such a system is made clear by reference to Figs. 14-2 and 14-3. At a time such as $t_1$ (indicated on Fig. 14-3), only $e_1$ is positive, and so only anode 1 is passing current. Since the volt drop across the tube is small, then most of the voltage appears across the load $R_l$. At an instant later than $t_2$, the transformer voltage $e_2$ also becomes positive. How-



ever, the second tube remains nonconducting because the voltage of the plate with respect to the cathode, $e_{b2}$, is still negative. That this is so can be seen from Fig. 14-2. It is noticed that

$$e_{b2} = e_2 - e_l$$

where $e_l$ is the voltage across the load $R_l$. However, when tube 1 conducts, then $e_l = e_1 - E_0$, where $E_0$ is the constant tube drop. Hence

$$e_{b2} = e_2 - e_1 + E_0 \qquad (14\text{-}1)$$
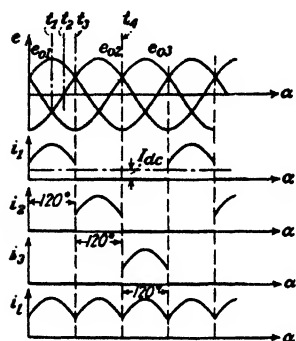
FIG. 14-3. Transformer voltages, tube currents, and load current in a three-phase half-wave rectifier.

Thus, so long as $e_1$ is greater than $e_2$, $e_{b2}$ will be less than $E_0$ and tube 2 will not conduct.

At the time $t_3$, however, $e_2$ equals $e_1$. At an instant later, $e_2$ exceeds $e_1$, and tube 2 will fire because $e_{b2} > E_0$. Further, since the plate voltage on the first rectifier is

$$e_{b1} = e_1 - e_2 + E_0$$

as is evident from the way Eq. (14-1) has been derived, then $e_{b1}$ will fall below $E_0$ and conduction of tube 1 will cease.

By following the same sequence of arguments, it may be shown that the current will transfer to tube 3 at the time $t = t_4$. Thus each tube conducts for 120 deg, or one-third of the total cycle. The output current is the sum of the currents through each of the anodes and is given by $i_l$ of Fig. 14-3. Actual oscillograms showing the output from such a polyphase system are given in Figs. 14-4 and 14-5.*

Also indicated in Fig. 14-3 is the average or d-c component of current that exists in anode 1. The magnitude of this d-c current is obtained from an analysis of the output-current wave form and is given in the next section. Since this is the value of the d-c current through each transformer secondary winding, it will cause some saturation of the transformer core. This is undesirable, for a distortion of the output wave form may result.

* All oscillograms showing characteristics of polyphase rectifiers were taken using type 83 tubes at a load current of approximately 100 ma. At these low currents, the transformer leakage reactance has little effect.

The effect of the d-c current in the transformer winding may be overcome by using transformers that are provided with two secondary windings, connected in "zigzag," as shown in Fig. 14-6. The two windings are connected into separate circuits in such a way that the instantaneous current in each winding, though of the same magnitude, will be in the opposite



FIG. 14-4. Oscillograms of transformer voltages and load current in a three-phase half-wave rectifier.

FIG. 14-5. Oscillograms showing the load current and the tube currents in a three-phase half-wave rectifier.

direction, and the mmf's due to each current will cancel each other. Specifically, the current $I_1$ is equal to the current $I_1'$ but is in the direction opposite to $I_1$. It must be remembered that 1 and 1' are the two secondary windings on the same core, both of which are fed by the same primary winding 1". 

Although this connection eliminates the effects of core saturation, it does so at the expense of economy, since it does not utilize the voltage of each winding to yield the highest possible output voltage. This condition is indicated in Fig. 14-6, where it is seen that the two secondary windings in series give but $\sqrt{3}E$, instead of $2E$ volts. This results from the addition of two sinor voltages that are 60 deg apart.

For the same reasons that the single-phase full-wave circuit is more de-sirable than the single-phase half-wave circuit, the three-phase full-wave circuit is more desirable than the three-phase half-wave system. The full-wave circuit supplies higher power at a higher efficiency than the half-wave circuit.
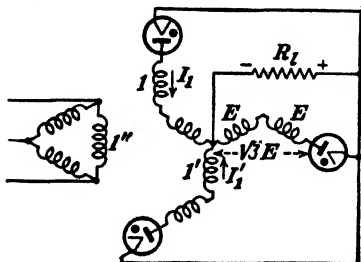


Fig. 14-6. "Zigzag" secondary wind-ings used to eliminate saturation of the transformer cores resulting from the d-c current.
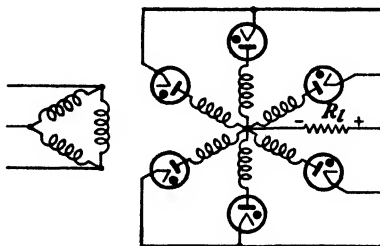
Fig. 14-7. Three-phase full-wave or six-phase half-wave rectifier connections.

If three transformers with center taps or with double secondary wind-ings are available, the six rectifier units may be connected as in Fig. 14-7. The physical behavior of this circuit is identical with that of the three-phase half-wave rectifier, except that each tube conducts for 60 instead of
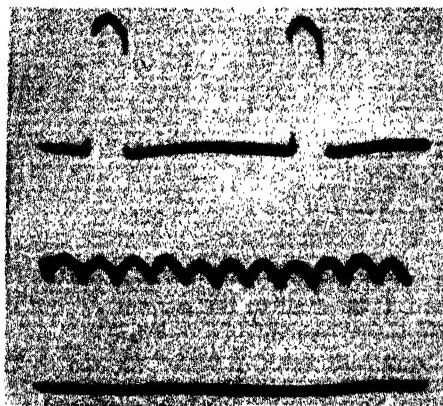


Fig. 14-8. Oscillogram showing one tube current and the load current in a six-phase half-wave rectifier.

120 deg. The oscillogram of Fig. 14-8 shows the current through one tube and also the load current.

If three transformers without center taps are available, then full-wave rectification is possible through the use of the three-phase bridge circuit,

shown schematically in Fig. 14-9. The output wave shape is the same as that shown in Fig. 14-8 for the six-phase half-wave circuit. However, the individual tube currents differ from those in the six-phase half-wave circuit, since each pair of tubes conducts for two successive 60-deg intervals
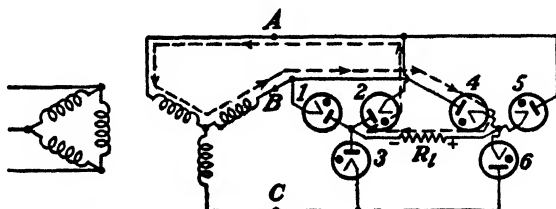


FIG. 14-9. Three-phase full-wave bridge circuit. At the instant shown the potential of point $B$ is positive with respect to point $A$, and tubes 4 and 2 are conducting. The corresponding current path is indicated by the broken line.

during each cycle. As was also true in the case of the single-phase bridge circuit, two tubes always act in series to pass current to the load. Though such a three-phase bridge circuit cannot be used with a single multianode pool-cathode tank because of the manner of connection, it is readily adaptable for use either with a bank of ignitrons or excitrons or with externally heated gas diodes. If ignitrons or excitrons are used, these must be insulated from ground because the cathodes in Fig. 14-9 are not all at the same potential.

From the way in which the circuit diagram has been drawn, it is noted that tubes 1, 2, and 3 form a Y-connected system which is connected in parallel with a second group of tubes 4, 5, and 6. This latter group has the plates connected to the lines feeding the cathodes of the first group.



FIG. 14-10. In the three-phase bridge circuit two tubes pass current simultaneously. Each tube conducts for 120 deg of each cycle as shown.

To understand the operation of this rectifier, refer to Fig. 14-10, in which the three-phase line voltages $e_{BA}$, $e_{AC}$, and $e_{CB}$ are indicated by solid lines. The negative of these voltages $e_{AB}$, $e_{CA}$, and $e_{BC}$, respectively, are indicated by the broken lines. The diodes that conduct over each 60-deg interval of the output voltage are also indicated. Thus, at the instant when the voltage $e_{BA}$ is a maximum ($B$ being positive with respect to $A$),*

* In accordance with the voltage-notation convention explained in Appendix VIII, $e_{BA}$ represents the voltage drop from point $B$ to point $A$. If $e_{BA}$ is positive at a given instant of time, then $B$ is at a higher potential than $A$ at this instant.

the plates of tubes 4 and 2 are positive, so that these tubes are conducting current. The current path through these tubes, the transformer secondaries, and the load is given in Fig. 14-9. When the voltage $e_{BC}$ exceeds $e_{BA}$, then the current will switch from tubes 4 and 2 to tubes 4 and 3. The explanation for this commutation is very similar to that given for Fig. 14-3 for the three-phase half-wave circuit. The reader should check the commutation sequence indicated in Fig. 14-10. In particular, note that the tubes conduct in pairs and that each tube carries current for 120 deg in each cycle, as predicted.

**14-2. Harmonics in $p$-phase Rectifiers.** The general case of a $p$-phase gaseous rectifier will be considered. The results so obtained will apply to the single-phase full-wave circuit ($p = 2$), a three-phase half-wave circuit ($p = 3$), a six-phase half-wave or a three-phase full-wave circuit ($p = 6$),
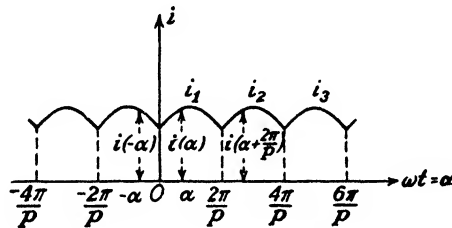


FIG. 14-11. The current wave shape of a $p$-phase rectifier.

and circuits represented by higher values of $p$. This analysis will not apply to the simple-phase half-wave circuit, which has been considered in some detail separately in Chap. 12.

By neglecting the effect of the small tube drop, the general form of the output current from a $p$-phase system is that illustrated in Fig. 14-11. This is merely an extension of the form of the current that has been shown to exist in Fig. 12-9 for the single-phase full-wave circuit, and in Figs. 14-5, and 14-8. This wave can be expressed analytically, for the regions specified, by the relations

$$\left.\begin{aligned}
i_1 &= I_m \cos\left(\alpha - \frac{\pi}{p}\right) & 0 \leq \alpha \leq \frac{2\pi}{p} \\[2mm]
i_2 &= I_m \cos\left(\alpha - \frac{3\pi}{p}\right) & \frac{2\pi}{p} \leq \alpha \leq \frac{4\pi}{p} \\[2mm]
i_3 &= I_m \cos\left(\alpha - \frac{5\pi}{p}\right) & \frac{4\pi}{p} \leq \alpha \leq \frac{6\pi}{p}
\end{aligned}\right\} \quad (14\text{-}2)$$

etc.

where $\alpha = \omega t$.

The validity of these equations is verified by applying them to the individual cases already considered in detail, namely, $p = 2, 3, 6$. It is to be noted that each expression differs from the previous one only through the change of phase of $2\pi/p$. Hence, with

$p = 2$, the phase between currents is $2\pi/2 = 180°$; 1-phase full-wave
$p = 3$, the phase between currents is $2\pi/3 = 120°$; 3-phase half-wave
$p = 6$, the phase between currents is $2\pi/6 = 60°$; 6-phase half-wave

In order to analyze the wave form of the output current in a Fourier series, the existence of certain conditions of symmetry, which simplifies the analysis rather markedly, is noted. The existing symmetries are

$$i(\alpha) = i(-\alpha)$$

and

$$i(\alpha) = i\left(\alpha + \frac{2\pi}{p}\right)$$

These equations state that the value of the current at any phase $\alpha$ is the same as that at the phase $-\alpha$, and also at the phase $\alpha + 2\pi/p$. That these are valid can be seen from an inspection of Fig. 14-11.

The first symmetry which is known as "zero-axis symmetry," requires that the general Fourier series representation

$$i = B_0 + \sum_{k=1}^{\infty} B_k \cos k\alpha + \sum_{k=1}^{\infty} A_k \sin k\alpha \qquad (14\text{-}3)$$

must satisfy the condition that

$$B_0 + \Sigma B_k \cos k\alpha + \Sigma A_k \sin k\alpha = B_0 + \Sigma B_k \cos(-k\alpha) + \Sigma A_k \sin(-k\alpha)$$

Noting that $\cos k\alpha = \cos(-k\alpha)$, whereas $\sin k\alpha = -\sin(-k\alpha)$, then, for equality, $A_k = -A_k$. This can be true only if $A_k = 0$. Hence, zero-axis symmetry requires that the Fourier series contain no sine terms. Because of this, the Fourier series becomes

$$i = B_0 + \sum_k B_k \cos k\alpha \qquad (14\text{-}4)$$

This shows that only the cosine harmonic terms are present in the resulting series.

Further, by virtue of the second symmetry, the series must satisfy the relation

$$B_0 + \Sigma B_k \cos k\alpha = B_0 + \Sigma B_k \cos k\left(\alpha + \frac{2\pi}{p}\right)$$

from which

$$\cos k\alpha = \cos k\left(\alpha + \frac{2\pi}{p}\right) = \cos\left(k\alpha + \frac{2\pi k}{p}\right)$$

Since the cosine of one angle can equal the cosine of another angle only when the two angles differ by an integral multiple of $2\pi$, it then follows that

$$\frac{2\pi k}{p} = 2\pi n$$

where $n$ is an integer. It follows from this that

$$k = np \tag{14-5}$$

This is a rather remarkable condition, for it shows that a large number of terms in the series (14-4) will be absent. Specifically, it shows that for $p = 2$, the case of the single-phase full-wave system, only the second, fourth, sixth, etc., harmonics are present [see Eq. (13-4)]. For a three-phase half-wave system, $k = 3n$, and so only the third, sixth, ninth, etc., harmonics are present. This shows that, the greater the number of phases of the supply system, the higher will be the order of the first a-c component that is present in the output wave. This condition is of extreme importance, for the filtering becomes increasingly simplified as the order of the harmonic increases.

In order to determine the explicit form of the series represented by Eq. (14-4), it is noted that the term

$$B_0 = \frac{1}{2\pi} \int_0^{2\pi} i \, d\alpha$$

simply represents the average current between the limits 0 and $2\pi$. Since the current repeats itself in the interval $2\pi/p$, then $B_0$ can be obtained by averaging over this interval. Thus

$$B_0 = \frac{p}{2\pi} \int_0^{2\pi/p} I_m \cos\left(\alpha - \frac{\pi}{p}\right) d\alpha$$

This becomes, upon carrying out the indicated integration,

$$B_0 = I_m \frac{p}{\pi} \sin \frac{\pi}{p} \tag{14-6}$$

The general coefficient in the Fourier series, $B_k$, is

$$B_k = \frac{1}{\pi} \int_0^{2\pi} i \cos k\alpha \, d\alpha$$

The value of this integral is $p$ times the same integral between the limits 0 and $2\pi/p$. This step is valid because both $i$ and $\cos k\alpha = \cos np\alpha$ repeat

at intervals of $2\pi/p$. Thus,

$$B_k = I_m \frac{p}{\pi} \int_0^{2\pi/p} \cos\left(\alpha - \frac{\pi}{p}\right) \cos k\alpha \, d\alpha$$

Performing the indicated integration yields

$$B_k = -\left(\frac{2p}{\pi} I_m \sin \frac{\pi}{p}\right)\left[\frac{1}{(k+1)} \frac{1}{(k-1)}\right]$$

Finally, the general expression for the output-current wave form of the $p$-phase system becomes

$$i = \left(I_m \frac{p}{\pi} \sin \frac{\pi}{p}\right)\left[1 - \sum_{k=p,2p,3p,\ldots} \frac{2 \cos k\omega t}{(k+1)(k-1)}\right] \qquad (14\text{-}7)$$

The expressions for several specific cases are written out explicitly.

$p = 2$, single-phase full-wave

$$i = \frac{2I_m}{\pi}\left(1 - \frac{2}{1 \times 3} \cos 2\omega t - \frac{2}{3 \times 5} \cos 4\omega t \right.$$
$$\left. - \frac{2}{5 \times 7} \cos 6\omega t - \cdots\right)$$

$p = 3$, three-phase half-wave

$$i = \frac{3\sqrt{3}}{2\pi} I_m \left(1 - \frac{2}{2 \times 4} \cos 3\omega t - \frac{2}{5 \times 7} \cos 6\omega t \right. \qquad (14\text{-}8)$$
$$\left. - \frac{2}{8 \times 10} \cos 9\omega t - \cdots\right)$$

$p = 6$, three-phase full-wave; six-phase half-wave

$$i = \frac{6I_m}{2\pi}\left(1 - \frac{2}{5 \times 7} \cos 6\omega t - \frac{2}{11 \times 13} \cos 12\omega t \right.$$
$$\left. - \frac{2}{17 \times 19} \cos 18\omega t - \cdots\right)$$

An inspection of these expressions indicates that not only does the order of the lowest harmonic term in the output increase with an increased number of phases but the magnitude of this term is smaller for large values of $p$.

The ripple factor can be evaluated from the Fourier series of the output wave. It follows from the first of Eqs. (14-8) for the single-phase full-wave system that

$$I_{dc} = \frac{2I_m}{\pi} = 0.636 I_m$$

Furthermore, if a wave is of the form

$$i = B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \cdots \qquad (14\text{-}9)$$

then the rms value of this wave is given by the well-known expression

$$I_{\text{rms}} = \sqrt{B_0{}^2 + \frac{B_1{}^2 + B_2{}^2 + B_3{}^2 + \cdots}{2}} \qquad (14\text{-}10)$$

Hence

$$I_{\text{rms}}' = \frac{2I_m}{\pi} \sqrt{\frac{1}{2}\left[\left(\frac{2}{1 \times 3}\right)^2 + \left(\frac{2}{3 \times 5}\right)^2 + \cdots\right]} = 0.307 I_m$$

This is, of course, simply the rms value of the a-c components, since the d-c term has been omitted from the expression. The ripple factor defined in Eq. (12-20) as

$$r \equiv \frac{I_{\text{rms}}'}{I_{dc}}$$

is then

$$r = \frac{0.307}{0.636} = 0.482$$

This is the value that has already been found [see Eq. (12-27)].

Suppose now that all terms in the Fourier expansion except the first harmonic term are neglected. The ripple factor becomes under these conditions, since

$$I_{\text{rms}}' = \frac{2I_m}{\pi} \sqrt{\frac{1}{2}\left(\frac{2}{1 \times 3}\right)^2} = 0.300 I_m$$

simply

$$r = \frac{0.300}{0.636} = 0.472$$

It is noted that this value is only about 2 per cent different from the value which results when all the terms in the expansion are used. In view of this, only the first harmonic term in the Fourier series for the case $p = 2$ or higher will be retained in calculations of the ripple factor, since the results will be within 2 per cent of the correct value. Under this approximation the output current for the $p$-phase system may then be written as

$$i = \left(I_m \frac{p}{\pi} \sin \frac{\pi}{p}\right)\left(1 - 2 \frac{\cos p\omega t}{p^2 - 1}\right) \qquad (14\text{-}11)$$

It follows from this expression that

$$I_{dc} = I_m \frac{p}{\pi} \sin \frac{\pi}{p} \qquad I_{rms}' = \left(I_m \frac{p}{\pi} \sin \frac{\pi}{p}\right)\left(\frac{\sqrt{2}}{p^2 - 1}\right) = \frac{\sqrt{2}I_{dc}}{p^2 - 1} \quad (14\text{-}12)$$

The ripple factor attains the form

$$r = \frac{\sqrt{2}}{p^2 - 1} \qquad (14\text{-}13)$$

This expression shows that the ripple factor decreases rapidly as the number of phases increases. The ripple factors are 0.472, 0.177, and 0.0404 for $p = 2, 3$, and 6, respectively. Thus, a six-phase system without a filter has a ripple of only 4 per cent.

**14-3. D-c Output Voltage, and Efficiency of Rectification in a $p$-phase Rectifier.** The effect of the tube drop $E_0$ was neglected in the last section. However, it will be taken into account in the calculation of the d-c output voltage and the rectifier efficiency.

By analogy with Eq. (14-12) for the output current, the d-c output voltage will be $E_m(p/\pi) \sin (\pi/p)$ if the tube drop is neglected. Since the output voltage is reduced by the amount of the tube drop, then clearly

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p} - E_0 \qquad (14\text{-}14)$$

This expression is valid, of course, only for those circuits in which only one tube conducts at a time. For the bridge circuits (see Figs. 12-13 and 14-9) in which two tubes conduct simultaneously, $E_0$ in this expression must be replaced by $2E_0$. Since this expression is independent of the load current, it appears that the regulation of a gaseous rectifier system is perfect. This is true only when the tubes are fed from ideal transformers (those with zero leakage reactance and zero resistance). The effect of the transformer leakage reactance will be considered in Sec. 14-5.

A plot of this expression (the small tube drop being neglected) is given in Fig. 14-12. It will be noted that as the



Fig. 14-12. Ratio of d-c output voltage to peak transformer voltage as a function of the number of phases in a polyphase rectifier. (Tube drop is neglected.)

number of phases is increased the ratio of the average or d-c voltage to the peak voltage of the transformer $E_m$ increases and approaches the value unity. That is, $E_{dc}$ approaches $E_m$ as the number of phases is increased.

The efficiency of rectification, defined as the ratio of the output d-c power to the input a-c power, may readily be calculated. The secondary transformer voltage $e$ is given by

$$e = E_m \cos\left(\alpha - \frac{\pi}{p}\right)$$

and the corresponding current $i$ is given by

$$i = \frac{E_m \cos\left(\alpha - \frac{\pi}{p}\right) - E_0}{R_l}$$

(14-15)

for the period from 0 to $2\pi/p$ during conduction. The a-c input power to the system is

$$P_i = \frac{p}{2\pi} \int_0^{2\pi/p} ei \, d\alpha$$

$$= \frac{p}{2\pi} \int_0^{2\pi/p} \left[ E_m \cos\left(\alpha - \frac{\pi}{p}\right) \right] \times \left[ \frac{E_m \cos\left(\alpha - \frac{\pi}{p}\right) - E_0}{R_l} \right] d\alpha$$

which integrates to

$$P_i = \left(\frac{p}{2\pi}\right) \frac{E_m{}^2}{R_l} \left[ \frac{\pi}{p} + \sin\frac{\pi}{p}\left(\cos\frac{\pi}{p} - \frac{2E_0}{E_m}\right) \right]$$

(14-16)

The corresponding output d-c power is

$$P_{dc} = \left(\frac{p}{\pi}\right)^2 \frac{E_m{}^2}{R_l} \left(\sin\frac{\pi}{p} - \frac{\pi}{p}\frac{E_0}{E_m}\right)^2$$

(14-17)

since $P_{dc} = E_{dc}{}^2/R_l$, where $E_{dc}$ is given by Eq. (14-14). The efficiency of rectification is given by the ratio of Eq. (14-17) to Eq. (14-16) and may be written in the form

$$\eta_r = \frac{200p}{\pi} \left[ \frac{\left(\sin\frac{\pi}{p} - \frac{\pi}{p}\frac{E_0}{E_m}\right)^2}{\frac{\pi}{p} + \sin\frac{\pi}{p}\left(\cos\frac{\pi}{p} - 2\frac{E_0}{E_m}\right)} \right] \quad \% $$

(14-18)

This expression is seen to remain substantially constant and independent of the load current. It depends only upon the number of phases $p$, the transformer maximum voltage $E_m$, and the tube drop $E_0$. The general form of the variation predicted by this expression is shown in Fig. 14-13.

It must be emphasized that the value of $\eta_r$ given by this expression is valid only for those circuits in which a single tube conducts at any time.

For the bridge circuits where two tubes conduct simultaneously, $E_0$ must be replaced by $2E_0$. Furthermore, the symbol $E_m$ must be correctly interpreted. In most cases, it represents the maximum transformer secondary alternating voltage to the common junction of the interconnected transformer bank. In the case of the three-phase zigzag connection, it still represents the phase voltage to the common junction, although this value will be $\sqrt{3}$ times the maximum winding voltage. In the case of the three-phase bridge circuit, it represents the line-to-line voltage and is $\sqrt{3}$ times the peak a-c value of each winding of the secondary. In other cases, a careful examination of the wiring diagram of the system must be made to determine the appropriate value.



FIG. 14-13. Efficiency of rectification as a function of the number of phases in a polyphase rectifier. (Tube drop is neglected.)

The advantages of the polyphase systems are evident: higher values of d-c current; higher efficiency of rectification; smoother output; and the existence of higher harmonics in the output, which permits effective filtering by means of simple filters.

*Example.* Compare the efficiency of rectification of two 600-volt 1,000-kw three-phase rectifiers, one of which consists of a polyphase tank (cathode-anode fall = 25 volts), the other of a bank of ignitrons (cathode-anode fall = 14 volts).

*Solution.* For the polyphase tank: By Eq. (14-14),

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p} - E_0$$

from which

$$E_m = \frac{E_{dc} + E_0}{\frac{p}{\pi} \sin \frac{\pi}{p}} = \frac{600 + 25}{\frac{3}{\pi} \sin \frac{\pi}{3}} = 755 \text{ volts}$$

By Eq. (14-18),

$$\eta_r = \frac{200 \times 3}{\pi} \left[ \frac{\left( \sin \frac{\pi}{3} - \frac{\pi}{3} \times \frac{25}{755} \right)^2}{\frac{\pi}{3} + \sin \frac{\pi}{3} \left( \cos \frac{\pi}{3} - \frac{2 \times 25}{755} \right)} \right] = 92.4\%$$

For the bank of ignitrons:

$$E_m = \frac{614}{\frac{3}{\pi} \sin \frac{\pi}{3}} = 742 \text{ volts}$$

whence, from Eq. (14-18),

$$\eta_r = \frac{200 \times 3}{\pi} \left[ \frac{\left( \sin \frac{\pi}{3} - \frac{\pi}{3} \times \frac{14}{742} \right)^2}{\frac{\pi}{3} + \sin \frac{\pi}{3} \left( \cos \frac{\pi}{3} - \frac{2 \times 14}{742} \right)} \right] = 94.3\%$$

The percentage improvement resulting from the use of the ignitrons is

$$\frac{94.3 - 92.4}{94.3} \times 100 = 2\%$$

**14-4. The Transformer Utilization Factor.** If the curves of Fig. 14-3 are examined critically, it will be observed that the current flows through each anode for times corresponding to 120 deg of each cycle for the case $p = 3$. That is, the transformer that supplies the power to an anode is being utilized for one-third of the cycle and remains idle for the remaining two-thirds of the cycle, or the time required for each of the other anodes to pass through their conducting periods. Evidently, the larger is the number of phases, the shorter is the period during which any particular transformer is supplying power to the output circuit. The transformer "utilization factor" gives a measure of the effectiveness with which the transformers in a polyphase rectifier system are used.

The utilization factor may be specified for either the primary or secondary winding of the transformer and is defined as *the ratio of the d-c output power supplied to the load to the volt-ampere capacity of the transformer bank.* Thus

$$\text{Utilization factor (UF)} = \frac{P_{dc}}{\text{volt-amp}} \tag{14-19}$$

This definition applies for both the primary and the secondary transformer windings. In each case, $P_{dc}$ refers to the d-c load power, whereas volt-amperes refers to either the primary or the secondary volt-amperes, respectively. The utilization factor is evidently less in magnitude than unity.

The utilization factor gives a measure of the output power that may be obtained from a transformer bank having a given volt-ampere rating when used in any particular rectifier circuit. Rewrite Eq. (14-19) in the following form, which specifies the rating of transformers for a given output power:

$$\text{Transformer rating (volt-amp)} = \frac{P_{dc}}{\text{UF}} \tag{14-20}$$

An approximate analytic expression for UF is readily derived. It will be supposed that the output current of each transformer during its conduct-
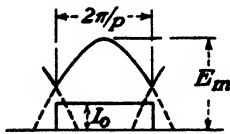


Fig. 14-14. In an ideally filtered rectifier the transformer current is constant while it is delivering power to the load.

ing period is a pure d-c value, as shown in Fig. 14-14, since the ultimate purpose of the rectifier system is to convert a-c power into d-c power. This involves an assumption that is not seriously in error for polyphase rectifiers. The error resulting from this assumption is very small when an inductor is placed in series with the load. The oscillograms of Figs. 14-15 and 14-16 show the "square" or "block" waves that exist under these conditions.
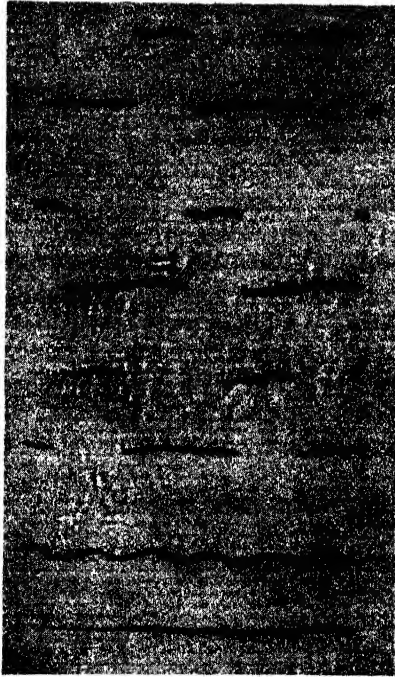
Fig. 14-15.  Oscillogram of the tube currents and the load current in a three-phase half-wave rectifier with a small inductance in series with the load.  (Compare with Fig. 14-5, which was taken under identical conditions except that the inductance was shorted out.)
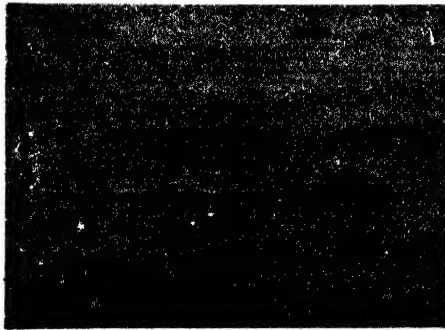


Fig. 14-16.  A tube current and the load currents in a six-phase half-wave circuit with a small inductance in series with the load.  The same inductance was used in this circuit as in the circuit of Fig. 14-15.  Note how much more effective the filtering is for the six-phase than for the three-phase circuit.  (Compare with Fig. 14-8, in which all conditions were the same except that the inductance was shorted out.)

Under these conditions, the average or d-c value of the voltage during the conducting part of the cycle is, from Eq. (14-14) (the tube drop being neglected),

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p}$$

The rms value of the current in the ideally filtered output system of Fig. 14-14 is readily seen to be

$$I_{rms} = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} i^2 \, d\alpha} = \sqrt{\frac{1}{2\pi} \int_0^{2\pi/p} I_0^2 \, d\alpha} = \frac{I_0}{\sqrt{p}} \qquad (14\text{-}21)$$

The total volt-amperes of the transformer secondaries is

$$p \left( \frac{E_m}{\sqrt{2}} \right) \left( \frac{I_0}{\sqrt{p}} \right) = \sqrt{\frac{p}{2}} \, E_m I_0$$

The d-c power supplied to the load is

$$E_{dc} I_{dc} = I_0 E_m \frac{p}{\pi} \sin \frac{\pi}{p} \qquad (14\text{-}22)$$

The ratio of these two terms is the transformer secondary utilization factor defined in Eq. (14-19),

$$\text{Secondary UF} = \frac{\sqrt{2p}}{\pi} \sin \frac{\pi}{p} \qquad (14\text{-}23)$$

A plot of this expression as a function of the number of phases $p$ is given in Fig. 14-17. It is noted that rectifiers with $p = 3$ utilize the transformer secondaries to the best advantage, although even for this system the utilization factor is only 0.675. The mathematical maximum value of (14-23), obtained by setting the derivative equal to zero, occurs at $p = 2.69$. This has no physical meaning, of course, since $p$ must be integral. Actually, Figs. 14-12, 14-13, and 14-17 should possess ordinates only at the discrete points $p = 2, 3, 4$, etc.
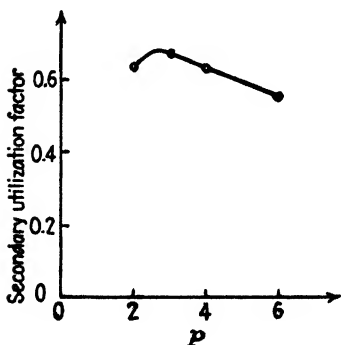


FIG. 14-17. Transformer secondary utilization factor as a function of the number of phases.

The problem of determining the transformer primary utilization factor is somewhat more complicated than that of finding the transformer secondary utilization factor. This complication arises from the necessity of determining the wave shape of the current in the primary winding when the current in the secondary winding is the rectangular wave shape illustrated in Fig. 14-14. The de-

sired result may be obtained in the following way: The current wave illustrated in Fig. 14-14 .may be expressed analytically over the complete cycle by means of a Fourier series of the form given by Eq. (14-3), *viz.*,

$$i = B_0 + \sum_{k=1}^{\infty} B_k \cos k\alpha + \sum_{k=1}^{\infty} A_k \sin k\alpha \qquad (14\text{-}24)$$

where, of course, $B_0$, the average value for the block wave, may readily be found. Now, if the transformer is considered to be perfect, all terms contained in the Fourier expansion (14-24) will be reflected into the transformer primary with the exception of the d-c term, since d-c current can-
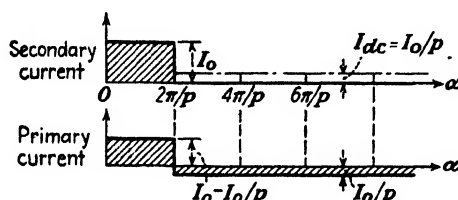


Fig. 14-18. The wave shape of the secondary current and the corresponding primary current in an ideally filtered $p$-phase rectifier. These are designated as *block* currents.

not be transformed. Consequently, the wave shape of the current in the primary winding will be simply that of the secondary current with the d-c term absent. This requires that the axis of the current wave in the secondary must be shifted by an appropriate amount in order to obtain the current in the primary winding. The amount of shift is such that the net area of the curve is zero. That is, the resultant curve has as much area above as below the axis. These results are illustrated in Fig. 14-18.

The rms value of the primary current illustrated above is

$$I_{rms} = \sqrt{\frac{\left(\dfrac{2\pi}{p}\right)\left(I_0 - \dfrac{I_0}{p}\right)^2 + \left(2\pi - \dfrac{2\pi}{p}\right)\left(\dfrac{I_0}{p}\right)^2}{2\pi}}$$

This reduces to

$$I_{rms} = \frac{I_0}{p} \sqrt{p - 1} \qquad (14\text{-}25)$$

Hence, the primary volt-amperes is

$$p\left(\frac{E_m}{\sqrt{2}}\right)\left(\frac{I_0}{p}\sqrt{p-1}\right) = \sqrt{\frac{p-1}{2}}\, E_m I_0 \qquad (14\text{-}26)$$

The ratio of Eq. (14-22) to Eq. (14-26) is the transformer primary utiliza-

tion factor.   This ratio becomes

$$\text{Primary UF} = \sqrt{\frac{2}{p-1}\frac{p}{\pi}\sin\frac{\pi}{p}} \qquad (14\text{-}27)$$

This expression is valid only for those systems in which each transformer supplies power to the rectifier system for $2\pi/p$ degrees in each cycle.   Con-



sequently, this expression is not valid either for the bridge circuits or for the six-phase half-wave circuit.

In the case of the six-phase half-wave circuit, an inspection of Fig. 14-7 shows that the primary current results from the action of two secondary windings, since there are two secondary windings per transformer required in this system.   The transformer primary utilization factor may readily be calculated for this system.   Let the two secondary windings be denoted by the letters $A$ and $B$.   Then Fig. 14-19 shows the secondary currents $i_A$ and $i_B$; the corresponding primary current components $i_A'$ and $i_B'$ (obtained from

Fig. 14-19.   Block-current diagram of one transformer of a six-phase half-wave rectifier.

$i_A$ and $i_B$ by subtracting the d-c value); and the resultant primary current $i_A' - i_B'$ (the negative sign arises from the fact that the currents are in opposite directions in the two secondary windings).

Except for the exciting current in the primary of a transformer, a balance is required between the mmfs produced by the secondary and the primary currents.   This means that in order to find the primary current in a rectifier transformer which is provided with multiple secondary windings, due account must be taken of the relative number of turns of each winding, as well as the direction of the currents in the several windings.   In the above illustration all secondary windings have the same number of turns.   However, this is not always true (see Fig. 14-29).

From Fig. 14-19 the rms value of the current in each transformer primary is found to be

$$I_{\text{rms}} = \sqrt{\frac{2I_0{}^2}{6}} = \frac{I_0}{\sqrt{3}}$$

The volt-ampere rating of the $\Delta$-connected transformer bank is

$$3\left(\frac{I_0}{\sqrt{3}}\right)\left(\frac{E_m}{\sqrt{2}}\right) = \sqrt{\frac{3}{2}}\,I_0 E_m$$

Also, the d-c output power of the system is

$$I_0 E_m \frac{6}{\pi} \sin \frac{\pi}{6} = \frac{3}{\pi} I_0 E_m$$

Therefore,

$$\text{Primary UF} = \frac{3/\pi}{\sqrt{3/2}} = \frac{\sqrt{6}}{\pi} = 0.777 \qquad (14\text{-}28)$$

From the foregoing analyses, it is evident that the single-phase full-wave circuit, the three-phase half-wave circuit, and the six-phase half-wave circuit possess very low utilization factors. This means that the kva capacity of the rectifier transformers must be much greater than the d-c kilowatt output. It should also be noted that the transformer primary utilization factor is higher than the corresponding secondary utilization factor. A little thought will show that this will always be true if the secondary current carries a d-c component. This means, of course, that unlike transformers used for ordinary a-c power applications the kva rating of a rectifier transformer secondary is usually different from and higher than that of the primary winding.

By virtue of its operation, the three-phase full-wave bridge circuit makes much better use of its rectifier transformers than any of the above-mentioned circuits. A careful study of the operation of this system (see Figs. 14-9 and 14-10) shows that the transformer currents in the three secondary windings $A$, $B$, and $C$ are as pictured in Fig. 14-20.



FIG. 14-20. The secondary block currents in the three transformers of Fig. 14-9 (the three-phase bridge circuit). Current toward the neutral of the transformer bank is taken as positive.

It should be noted that there is no resultant d-c component of current in the transformer windings. This eliminates any effect of core saturation. Furthermore, the primary currents will have exactly the same wave shape as the secondary currents. As a result, the transformer primary and secondary utilization factors will be the same.

To find these factors, it is noted that

$$E_{dc} = \sqrt{3} E_m \frac{p}{\pi} \sin \frac{\pi}{p}$$

where $E_m$ is the transformer maximum voltage. The factor $\sqrt{3}$ arises from the fact that two transformers supply power simultaneously (the voltages are 120 deg apart in time phase). Since the output is that of a

three-phase full-wave circuit, $p = 6$, whence this expression reduces to

$$E_{dc} = \frac{3\sqrt{3}}{\pi} E_m$$

The d-c output power is given by

$$E_{dc}I_0 = \frac{3\sqrt{3}}{\pi} E_m I_0$$

The rms current is

$$\sqrt{\tfrac{2}{3}I_0{}^2} = \sqrt{\tfrac{2}{3}}I_0$$

and the rms transformer voltage is $E_m/\sqrt{2}$. The total volt-ampere rating is

$$3\left(\sqrt{\frac{2}{3}}\,I_0\right)\left(\frac{E_m}{\sqrt{2}}\right) = \sqrt{3}I_0 E_m$$

Hence

$$\mathrm{UF} = \frac{(3\sqrt{3}/\pi)E_m I_0}{\sqrt{3}E_m I_0} = \frac{3}{\pi} = 0.955 \qquad (14\text{-}29)$$

for either the primary or the secondary windings. The reciprocal of this is 1.05, which means that the transformer rating (both primary and secondary kva) need be only 5 per cent greater than the d-c kilowatt output power.

**14-5. The Effect of Leakage Reactance on Rectifier Operation.**[1] The foregoing analyses have explicitly supposed that the transformers were perfect. That is, the losses were considered to be negligible, and the leakage inductance was considered to be zero. This idealization is approximately true. However, the presence of the leakage reactance does have marked effects upon the operation of the rectifier system. It causes two anodes to fire simultaneously, resulting in an electrical connection between the windings of the transformers of the "overlapping" phases (see Fig. 14-21). Thus, instead of the arc transferring immediately
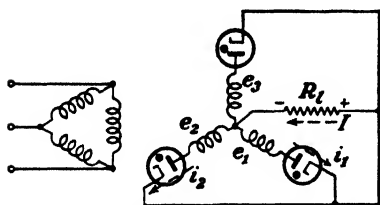


FIG. 14-21. Two tubes conduct simultaneously during overlap. The load current is assumed constant so that $I = I_0$.

from one anode to the next in firing sequence, as depicted in Fig. 14-3, a short region exists during overlap in which both anodes conduct. The overlap results from the presence of the leakage inductance $L$, which consists of the leakage inductance of the secondary winding plus the effects of the line and primary leakage inductances reflected into the secondary.

The application of Kirchhoff's law around the path of the overlapping phases yields

$$\left(e_1 - L\frac{di_1}{dt}\right) - \left(e_2 - L\frac{di_2}{dt}\right) = 0 \tag{14-30}$$

where, for simplicity, the tube drops have been neglected. Furthermore, since two anodes are passing current, then

$$I_0 = i_1 + i_2 \tag{14-31}$$

It will be assumed that the load circuit is highly inductive, so that the load current $I_0$ is constant.
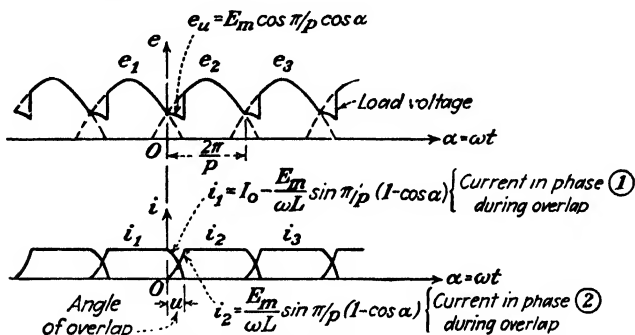


Fig. 14-22. The load voltage and individual tube currents taking into account overlap in a $p$-phase rectifier.

If the origin $\alpha = 0$ is chosen as shown in Fig. 14-22, the transformer voltages $e_1$ and $e_2$ are given by the expressions

$$e_1 = E_m \cos\left(\alpha + \frac{\pi}{p}\right) \qquad e_2 = E_m \cos\left(\alpha - \frac{\pi}{p}\right) \tag{14-32}$$

By combining Eqs. (14-30), (14-31), and (14-32), there results

$$2L\frac{di_1}{dt} = E_m \cos\left(\alpha + \frac{\pi}{p}\right) - E_m \cos\left(\alpha - \frac{\pi}{p}\right) = -2E_m \sin\frac{\pi}{p}\sin\alpha$$

When $\alpha = \omega t$ equals zero, anode 1 is carrying all the current, so that $i_1 = I_0$ at $\alpha = 0$. Integrating the above equation subject to this initial condition yields

$$i_1 = I_0 - \left(\frac{E_m}{\omega L}\sin\frac{\pi}{p}\right)(1 - \cos\alpha) \tag{14-33}$$

during the overlap period. The current $i_2$ becomes

$$i_2 = I_0 - i_1 = \left(\frac{E_m}{\omega L}\sin\frac{\pi}{p}\right)(1 - \cos\alpha) \tag{14-34}$$

during the overlap period. These currents are shown in Fig. 14-22.

During that portion of the cycle when one tube only is carrying current, the load voltage equals the transformer voltage feeding this anode (the tube drop being neglected). However, during the overlap period the output voltage is $e_1 - L\,di_1/dt$ or $e_2 - L\,di_2/dt$, as is evident from an inspection of Fig. 14-21. From the known values of $e$ and $i$, the load voltage during the overlap period reduces to

$$e_u = E_m \cos \frac{\pi}{p} \cos \alpha \tag{14-35}$$

which is also

$$e_u = \frac{e_1 + e_2}{2} \tag{14-36}$$

This shows that the load voltage during overlap is the average of the transformer induced voltages of the two phases.

In order to calculate the angle of overlap $u$, it is to be noted from the diagram that $i_1 = 0$ when $\alpha = u$. Thus, from Eq. (14-33)

$$0 = I_0 - \left(\frac{E_m}{\omega L} \sin \frac{\pi}{p}\right)(1 - \cos u)$$

from which it follows that

$$\cos u = 1 - \frac{\omega L I_0}{E_m \sin (\pi/p)} \tag{14-37}$$

An inspection of Fig. 14-22 shows that the most serious consequence of the presence of appreciable leakage reactance is an output d-c voltage $E_{dc}$, which will vary with the load current. That is, an appreciable voltage regulation will result from the existence of the leakage reactance. The exact dependence of the output voltage upon the load current is readily determined. Thus,

$$E_{dc} = \frac{p}{2\pi} \int_0^{2\pi/p} e_l\,d\alpha = \frac{p}{2\pi}\left(\int_0^u e_u\,d\alpha + \int_u^{2\pi/p} e_2\,d\alpha\right)$$

The result of the integration is

$$E_{dc} = \left(\frac{p}{2\pi} E_m \sin \frac{\pi}{p}\right)(1 + \cos u) \tag{14-38}$$

Substituting the value of $\cos u$ from Eq. (14-37) into this expression gives

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p} - p \frac{\omega L I_0}{2\pi} \tag{14-39}$$

This shows that the voltage decreases linearly with the load current. If the tube drop $E_0$ is to be taken into account, it is only necessary to reduce this expression by this value $E_0$.

**14–6. The Double-Y Rectifier Circuit.**  It was shown in Sec. 14–4 that the three-phase full-wave bridge circuit is characterized by high transformer primary and secondary utilization factors.  For this reason, plus others to be considered below, this circuit recommends itself for certain applications where single-anode tubes are to be used.  However (see Fig. 14-9), this circuit cannot be used with multianode pool-cathode tanks.

A very important rectifier circuit, which is known as a *double-Y circuit*, is illustrated in Fig. 14-23.  The double-Y circuit possesses a transformer primary utilization factor equal to that of the bridge circuit, but the transformer secondary utilization factor is somewhat lower than that of the
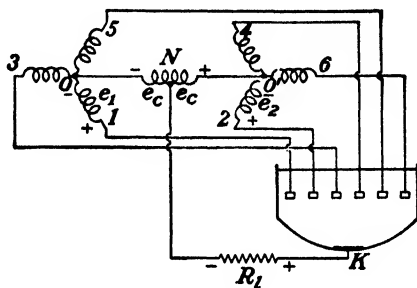


Fig. 14-23.  The double-Y rectifier connections.  Winding 1 is 180 deg out of phase with 4, winding 2 is 180 deg out of phase with 5, and winding 3 is 180 deg out of phase with 6.

bridge circuit.  There are, however, several features that more than offset this, the most important being that pool-cathode multianode or single-anode tanks may be used.  This permits, of course, the very high currents that are available from the pool cathode.  Also, the peak anode current in such a system is one-half the total load current, whereas in the bridge circuit the peak anode and the load currents are equal.  Consequently, the double-Y connection recommends itself for systems that are to supply high currents and is used on most large six-phase or higher phase rectifiers.

It will be noted that this system consists essentially of two separate three-phase Y-connected secondary systems whose corresponding windings are displaced by 180 electrical degrees with respect to each other.  The two three-phase systems have their neutral points connected together by a center-tapped reactor, generally referred to as the "interphase transformer," which is marked $ONO'$ in the diagram.

The three-phase system of voltages of group I, consisting of anodes 1, 3, and 5, are shown in Fig. 14-24.  Directly below are indicated the three-phase voltages of group II, consisting of anodes 2, 4, and 6.  It should be noted that the voltage of anode 4 is the negative of that of anode 1, 5 is the negative of 2, and 6 is the negative of 3.  The action of the circuit can

best be understood if the interphase transformer is considered simply as a "commutating reactor"[2] that causes two anodes to fire simultaneously.

Consider, for example, the instant of time $t_1$ shown on the curves of Fig. 14-24. At this instant, the voltages $e_3$, $e_4$, and $e_5$ are negative, so that anodes 3, 4, and 5 are not conducting. The voltages $e_1$ and $e_2$ are positive, and anodes 1 and 2 conduct *simultaneously*. Although the voltage $e_6$ is positive, the anode-cathode voltage of anode 6 is $+ e_6 - e_2$, which is negative, so that anode 6 is not conducting. (This result is obtained by apply-
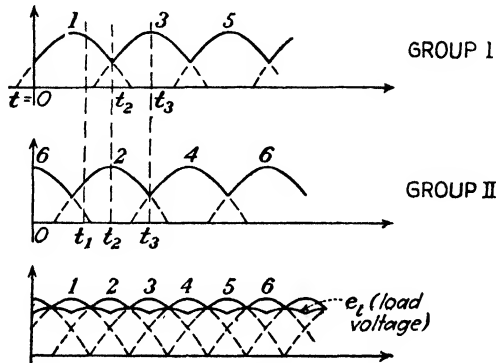


Fig. 14-24. The internal and external behavior of a double-Y-connected rectifier. Anodes 1, 3, and 5 act as a simple three-phase half-wave rectifier. The same is true of anodes 2, 4, and 6. Each anode conducts for 120 deg in each cycle. The two groups act effectively in parallel so that at every instant of time one anode of each group is conducting. The load voltage is the average of the output voltages of the two groups.

ing Kirchoff's law to the path $6O'2K6$ of Fig. 14-23, the tube drop being neglected for simplicity.) Consider the path $1ONO'2K1$, which connects anodes 1 and 2 together through the interphase transformer. If $e_c$ is the voltage to center tap of this reactor (the "commutating voltage"), then Kirchhoff's law around this path gives

$$e_1 - 2e_c - e_2 = 0$$

from which

$$e_c = \tfrac{1}{2}(e_1 - e_2) \qquad (14\text{-}40)$$

This means, of course, that anode 2 is also passing current, and it is the difference in voltage between the two conducting anodes that appears across the reactor.

The load voltage is found by choosing a path through the load, through one-half of the interphase reactor, and through either anode 1 or anode 2. The result is either

$$e_l = e_1 - e_c \qquad \text{or} \qquad e_l = e_2 + e_c \qquad (14\text{-}41)$$

By adding these expressions,

$$e_l = \tfrac{1}{2}(e_1 + e_2) \qquad (14\text{-}42)$$

This shows that *the load voltage is the average value of the transformer voltages* $e_1$ *and* $e_2$. This is clearly shown in Fig. 14-24.

At the time $t_2$ when $e_1$ equals $e_3$, the current shifts from anode 1 to anode 3, exactly as in a simple three-phase system. However, anode 2 continues to carry current. The commutating voltage shortly after this instant becomes

$$e_c = \tfrac{1}{2}(e_3 - e_2)$$

and the load voltage becomes

$$e_l = \tfrac{1}{2}(e_2 + e_3)$$

(The leakage reactance of the rectifier transformers has been neglected.)

At the time $t_3$ the current transfers from anode 2 to anode 4, which then conducts simultaneously with tube 3.

From the foregoing reasoning, it can be concluded that internally group I behaves independently of group II, each functioning as an ordinary three-phase half-wave rectifier, and each tube conducts for 120 deg/cycle. Since two anodes always conduct simultaneously, these two groups may be considered to be acting essentially in parallel. The difference in voltage between the transformer potentials to the anodes that are conducting appears across the commutating reactor. The load voltage is the average value of the output voltage of group I and group II. This load voltage appears to have, from Fig. 14-24, twice the frequency of either group I or group II. It will be shown that this is true and that



FIG. 14-25. Oscillogram verifying the behavior of the double-Y-connected rectifier as depicted in the sketches of Fig. 14-24.

the lowest harmonic frequency in the output voltage is a sixth harmonic of the supply frequency. An oscillogram showing these features is given in Fig. 14-25.

The Fourier series of the output voltage of group I is that given by Eq. (14-8), *viz.*,

$$e_I = \frac{3\sqrt{3}}{2\pi} E_m \left( 1 - \frac{\cos 3\alpha}{4} - \frac{2\cos 6\alpha}{35} - \cdots \right)$$

It is evident from Fig. 14-24 that $e_{II}$ can be obtained from $e_I$ merely by shifting the horizontal axis by 60 deg, *i.e.*, by changing $\alpha$ to $\alpha - 60$ deg.

Thus,

$$e_{II} = \frac{3\sqrt{3}}{2\pi} E_m \left[ 1 - \frac{1}{4} \cos 3(\alpha - 60) - \frac{2}{35} \cos 6(\alpha - 60) - \cdots \right]$$

$$= \frac{3\sqrt{3}}{2\pi} E_m \left( 1 + \frac{1}{4} \cos 3\alpha - \frac{2}{35} \cos 6\alpha + \cdots \right)$$

Hence, the load voltage becomes

$$e_l = \frac{1}{2}(e_I + e_{II}) = \frac{3\sqrt{3}}{2\pi} E_m \left( 1 - \frac{2}{35} \cos 6\alpha - \cdots \right) \qquad (14\text{-}43)$$

It should be noted that the third harmonic term cancels out of the expression and that the lowest frequency in the output voltage is the sixth harmonic, as predicted. The d-c load voltage is

$$E_{dc} = \frac{3\sqrt{3}}{2\pi} E_m$$

which is the d-c value of each three-phase group, despite the fact that the output voltage contains no third harmonic term.

The voltage that appears across the commutating reactor is a third harmonic of the supply frequency, and not a sixth harmonic as appears in the load voltage. This follows from Eq. (14-40), from which it is found that

$$e_c = \frac{1}{2}(e_I - e_{II}) = -\frac{3\sqrt{3}}{8\pi} E_m \left( \cos 3\alpha + \frac{\cos 9\alpha}{10} + \cdots \right) \qquad (14\text{-}44)$$

This expression when plotted indicates that the voltage across the commutating reactor has a triangular wave form (compare with Prob. 14-9).



This result is illustrated in the oscillogram of Fig. 14-26. It is noticed, therefore, that there is no d-c voltage across the interphase reactor. This means, of course, that no saturation results from the presence of a d-c current component in this reactor. Also, the lowest frequency voltage that appears across this reactor is a third harmonic, and only odd multiples of $3\omega$ will appear across this reactor. That is, all even multiples of $3\omega$ cancel out.

Fig. 14-26. Oscillogram showing the approximately triangular wave form of the voltage across the commutating reactor of a double-Y-connected rectifier.

The subtraction indicated in Eq. (14-44) may also be carried out graphically from the plotted voltages $e_I$ and $e_{II}$ of Fig. 14-24. This construction confirms the conclusion of the preceding paragraph that the voltage across the commutating reactor is approximately triangular in wave shape and of a frequency $3\omega$.

The entire foregoing analysis is based upon the fact that Kirchhoff's voltage equation is satisfied around every closed loop. However, one additional condition must be fulfilled in order to have a physically realizable system, *viz.:* the current through any anode can never be negative. Consider, therefore, the conditions that exist when anodes 1 and 2 are conducting simultaneously, as at the instant $t_1$ of Fig. 14-24. The conducting circuit of Fig. 14-23 is redrawn in Fig. 14-27, the nonconducting anodes having been omitted from the diagram. It is assumed that the load circuit is highly inductive so that the *load* current is constant and equal to $I_0$. The tube currents, however, are not equal. $i_1$ is decreasing and $i_2$ is increasing at this instant. The d-c component of the current delivered by each is $I_0/2$. The instantaneous current in anode 1 is then
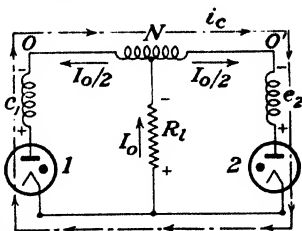


Fig. 14-27. In the double-Y-connected rectifier the current in the conducting anodes may be considered to be made up of two components. These are: one-half the constant direct current, $I_0/2$, and a commutating or circulating current $i_c$.

and that in anode 2 is

$$i_1 = \frac{I_0}{2} - i_c$$
$$i_2 = \frac{I_0}{2} + i_c$$

(14-45)

where $i_c$ is the commutating or circulating current indicated in Fig. 14-27. This commutating current is determined by the voltage $e_c$ [Eq. (14-44)] and the reactance $X$ of the interphase reactor.

Since $i_1$ cannot be negative, the minimum value of $I_0$ is twice the maximum value of $i_c$. This minimum load current value is called the "critical load current." The interphase transformer is so designed that its reactance $X$ limits the commutating current to a low value. The critical load current is usually less than 2 per cent of the full load current, and the circulating current must be kept smaller than one-half of this value.

Since the foregoing theory is not valid for currents less than the critical value, what is the true situation at these low currents? Consider the extreme case of zero load current. Figure 14-27 shows that if $I_0 = 0$ then the current through anode 1 is $-i_c$ and that through anode 2 is $+i_c$. Since, however, neither anode can pass negative current (the electrons can flow only from cathode to anode within each tube), then $i_c$ must equal zero. There can therefore be no commutating current and the commutating voltage must also be zero. Under these circumstances the interphase reactor may be considered to be missing from the circuit. This means that the

double-Y circuit must reduce to the simple six-phase half-wave circuit (compare with Fig. 14-7). As a result, each tube conducts for 60 deg in each cycle, and the output voltage is $1.35E_{rms}$ [obtained from Eq. (14-14) with $p = 6$ and $E_0 = 0$]. However, for currents $I_0$ that are greater than the critical value, the load voltage is that of a three-phase half-wave system, or $1.17E_{rms}$. Thus, the voltage at no load is $1.35/1.17$ times the load voltage. This results in a 15 per cent voltage regulation in the range from no load to the critical load.

In the range from no load to the critical load, the situation is precisely that discussed in the last section in which the matter of overlap was in-
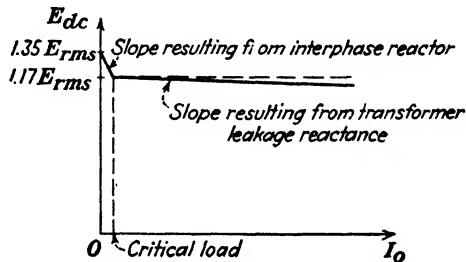


Fig. 14-28.   Regulation curve of the double-Y circuit.

vestigated. Because of the overlap, which results in this case from the presence of the reactance of the interphase transformer (the leakage reactance of the transformer being negligible in comparison with this value), conduction proceeds for more than 60 deg as the load current is increased. The angle of overlap increases with load current until the critical load is reached. At this point, each anode is conducting for 120 instead of 60 deg, and the rectifier has attained the double-Y features. As the load is increased beyond the critical value, only a small increase in overlap occurs, resulting now from the effect of the transformer leakage reactances.

A typical voltage regulation curve is shown in Fig. 14-28. The large slope due to the interphase reactor is indicated; and the gradually decreasing voltage, given by Eq. (14-39), resulting from the transformer leakage reactance is also shown.

A group of double-Y circuits may be connected together by commutating reactors to provide higher than effective 6-phase operation. This requires, of course, that the appropriate phase shift between groups be introduced in order to yield the desired results. One such circuit is the quadruple zigzag rectifier, which is illustrated in Fig. 14-29. It is a 12-phase rectifier which behaves like two double-Y circuits in parallel. These circuits retain the desirable features of the 3-phase circuits (*viz.*, conduction for 120 deg of each cycle), although the load voltage contains only harmonics of higher order.[3]
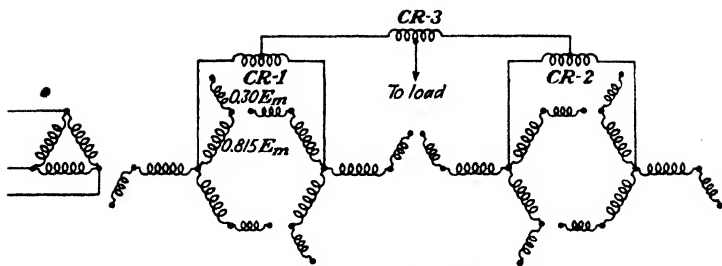
FIG. 14-29. The transformer connections of a quadruple zigzag rectifier system.

**14-7. Rectifier-design Data.** The main results of the analysis of the rectifier circuits outlined above are summarized in Table 14-1. All the items contained therein can be obtained from the formulas and the methods of analysis developed above.

TABLE 14-1 *

RECTIFIER-DESIGN DATA

| Circuit | One-phase full-wave | One-phase bridge | Three-phase half-wave | Three-phase bridge | Double-Y | Six-phase half-wave |
|---|---|---|---|---|---|---|
| Circuit diagram........ | 12–8 | 12–13 | 14–2 | 14–9 | 14–23 | 14–7 |
| Transformer secondary rms volts/deg........ | $1.11E_{dc}$ | $1.11E_{dc}$ | $0.855E_{dc}$ | $0.427E_{dc}$ | $0.855E_{dc}$ | $0.741E_{dc}$ |
| Peak inverse voltage.... | $3.14E_{dc}$ | $1.57E_{dc}$ | $2.09E_{dc}$ | $1.045E_{dc}$ | $2.09E_{dc}$ | $2.09E_{dc}$ |
| Average tube current... | $0.5I_{dc}$ | $0.5I_{dc}$ | $0.33I_{dc}$ | $0.33I_{dc}$ | $0.167I_{dc}$ | $0.167I_{dc}$ |
| Peak tube current.. ... | $I_{dc}$ | $I_{dc}$ | $I_{dc}$ | $I_{dc}$ | $0.5I_{dc}$ | $I_{dc}$ |
| Transformer secondary kva rating........... | $1.57P_{dc}$ | $1.11P_{dc}$ | $1.481P_{dc}$ | $1.047P_{dc}$ | $1.481P_{dc}$ | $1.814P_{dc}$ |
| Transformer primary kva rating........... | $1.11P_{dc}$ | $1.11P_{dc}$ | $1.21P_{dc}$ | $1.047P_{dc}$ | $1.047P_{dc}$ | $1.283P_{dc}$ |
| Principal ripple frequency.............. | $2f$ | $2f$ | $3f$ | $6f$ | $6f$ † | $6f$ |
| Ripple factor.......... | 0.472 | 0.472 | 0.177 | 0.0404 | 0.0404 | 0.0404 |

* The values in this table are based upon the following assumptions: zero transformer resistance; zero transformer reactance; zero tube drop; very large reactance in series with the load so that the tube current is constant (except for the ripple frequency and ripple factor).

† The load ripple frequency is $6f$, but the interphase reactor ripple frequency is $3f$.

Suppose that the three circuits that yield effective six-phase output are compared. The six-phase half-wave circuit possesses no features that

compensate for the low transformer utilization factors, and so this circuit is seldom used.

The three-phase bridge circuit is characterized by (1) high transformer primary and secondary utilization factors, (2) two tubes in series during the conduction period, (3) low inverse peak voltage per tube. Though (1) and (3) are desirable features, the drop in the two conducting tubes results in a reduced rectifier efficiency. Also, by virtue of its connection, a multi-anode tank cannot be used with this circuit. This circuit is used extensively for high-voltage moderate-current service.

The double-Y circuit is characterized by (1) poor transformer secondary utilization factor, but with a high primary utilization factor, (2) high inverse peak voltage per anode, (3) low average tube current. By virtue of (3) and the fact that this connection is possible with multianode tanks, although modern practice favors the use of single-anode tanks for rectifier service, this circuit is recommended for moderate-voltage high-current service. Other advantages arise from the fact that the arc drop is lowered as a result of the low peak current. This results in an increased efficiency. Furthermore, since a lower current is commutated, the regulation is improved.

*Example.* The plates of a certain radio-frequency power amplifier require 4.5 amp at 16,000 volts. Design a polyphase rectifier that will supply this power from a 220-volt three-phase supply.

*Solution.* The three-phase bridge circuit is used for the reasons given in the foregoing discussion. From the data in column 5 of Table 14-1, it is found that

$$\text{Peak-inverse voltage} = 1.045 \times 16{,}000 = 16{,}700 \text{ volts}$$
$$\text{Average tube current} = 0.33 I_{dc} = 1.5 \text{ amp}$$
$$\text{Peak-tube current} = 4.5 \text{ amp}$$

Western Electric type 255-A mercury-vapor diodes, for which

$$\text{Peak-inverse-voltage rating} = 20{,}000 \text{ volts}$$
$$\text{Peak-tube-current rating} = 5.0 \text{ amp}$$

would be satisfactory tubes to use in this circuit.

*Transformer characteristics*:

$$\text{Transformer-secondary rating} = \frac{1.047 \times 16{,}000 \times 4.5}{3} = 25.1 \text{ kva}$$
$$\text{Transformer-primary rating} = 25.1 \text{ kva}$$
$$\text{Transformer-secondary voltage} = 0.427 \times 16{,}000 = 6{,}830 \text{ volts rms}$$
$$\text{Transformer-primary voltage} = 220 \text{ volts rms}$$

**14-8. Controlled Polyphase Rectifiers.** Many of the circuits considered above for polyphase rectifiers, using either the mercury pool-cathode tanks or single-anode units, have been adopted for use with grid pool tanks, ignitrons, and excitrons. Figure 11-17 is a photograph of a six-tank excitron assembly.

Figure 14-30 shows a typical circuit of a three-phase double-Y controlled ignitron circuit.[4] Each ignitron is controlled by an individual thyratron. For simplicity, only one ignitron is shown connected to its control thyratron. The Δ-connected primaries of the grid transformers are supplied from the three-phase supply through a phase shifter. The output voltage
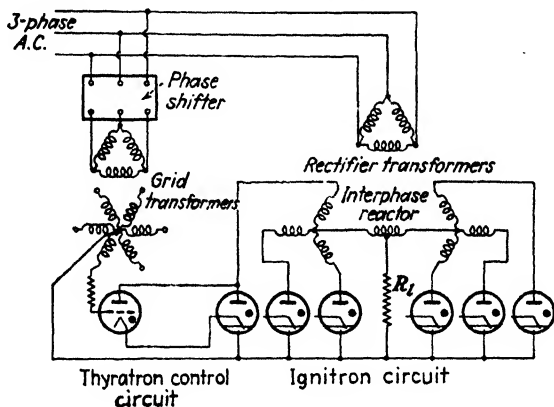


FIG. 14-30. A double-Y ignitron circuit with thyratron control. The complete ignitor-rod control circuit is shown for only one ignitron.

of this system for a pure resistance load with an angle of delay $\varphi$ is shown in Fig. 14-31. This output wave shape will be considered in some detail in Sec. 14-9.

The phase-shifting circuit above can also be employed with grid pool tanks. Another common circuit for grid pool tanks is the bias phase con-
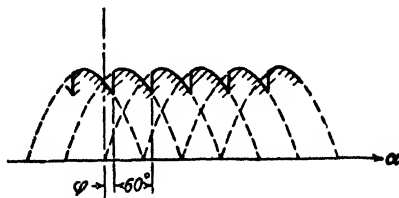


FIG. 14-31. The output-voltage wave shape of a controlled six-phase rectifier.

trol discussed in Sec. 12-10,[5] which has been extended for polyphase operation. The circuit connections for a three-phase tank employing this method of control are shown in Fig. 14-32.

The grid transformers are energized from the same a-c source as the rectifier transformers. The use of the zigzag connection for the power transformers provides a phase displacement of 30 deg between the grid

voltage and the corresponding anode voltage.  In addition, the use of an auxiliary d-c grid bias, which is obtained from a single-phase rectifier system, together with the alternating grid voltage, provides a grid potential that may be smoothly varied over an appreciable range.  This is evident
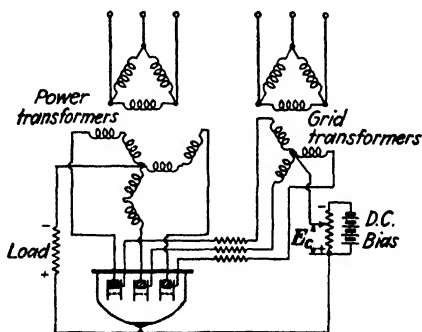


F ɪ ɢ. 14–32.  Three-phase grid pool tank (see Sec. 11-9).  The bias-phase method of control is used.

from an inspection of Fig. 14-32, from which it is seen that the instantaneous potential $e_c$ is the corresponding value of the grid phase voltage less the d-c bias potential.  Obviously, the grid potential is zero when the grid alternating voltage is positive and just equal to the d-c grid bias.
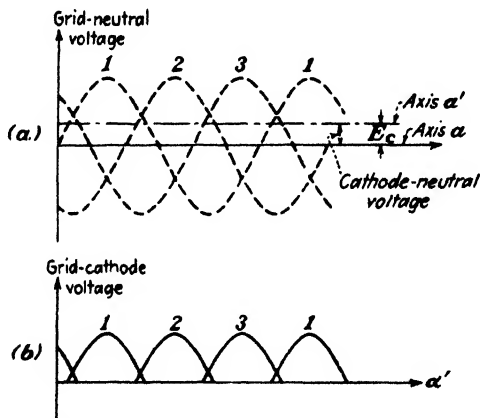


Fɪɢ. 14–33.  The grid circuit potentials in a grid-controlled three-phase tank rectifier.

The curves of Fig. 14-33a represent the potentials of the grid with respect to the neutral of the Y-connected grid transformers, the constant d-c bias $E_c$ giving the cathode voltage with respect to the neutral.  The

difference of these two potentials, which gives the grid potential with respect to the cathode, is the polyphase system of voltages taken with respect to the displaced axis $\alpha'$. This has been separately redrawn in Fig. 14-33b, with only the positive lobes of the grid potential with respect to the cathode being indicated.

The positive portions of the three-phase voltages applied to the anodes are shown in Fig. 14-34. Also included in this figure are the cathode-grid
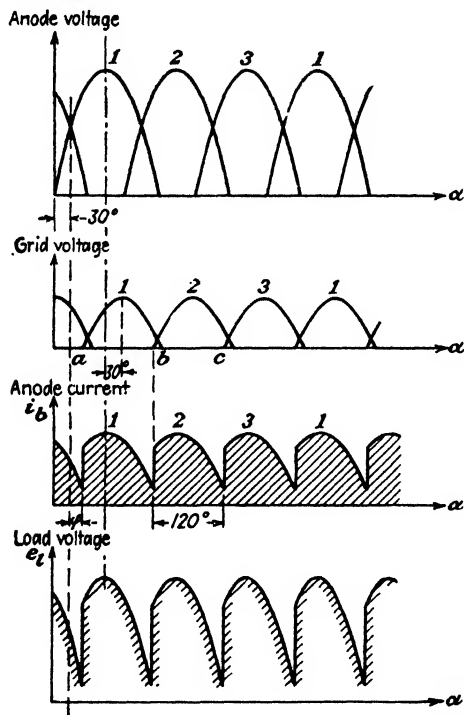


FIG. 14-34. Current and voltage wave shapes in a three-phase grid-controlled tank with the grid voltage adjusted as in Fig. 14-33.

voltages, which have been redrawn from Fig. 14-33b. The proper phase relationship between these two sets of voltages has been maintained. That is, the peak of the grid voltage occurs 30 deg later than the peak of the corresponding anode voltage.

If it is assumed that the critical grid starting characteristic is the zero voltage axis, then conduction will begin at the point $a$ of phase 1, point $b$ of phase 2, and point $c$ of phase 3. That is, conduction will begin at the points where the grid-cathode voltage crosses the zero axis. The output

wave shape is, of course, exactly the same as that which would arise if three ignitrons or three excitrons, adjusted for the same delay angle $\varphi$, were used.

Careful consideration reveals that the delay angle $\varphi$ can be varied between the limits 0 and 90 deg. If the d-c bias $E_c$ is zero, the grid-cathode voltage crosses the zero axis at a point 30 deg beyond the start of the anode-voltage cycle. This corresponds to zero delay angle, and the output is exactly that from a three-phase mercury-arc rectifier without control. The other extreme condition occurs when the d-c bias is equal to the peak value of the grid-transformer phase voltage. The grid-cathode voltage will just graze the zero axis at the 90-deg point of its cycle, or at the 120-deg point of the anode-voltage cycle. This corresponds to a 90-deg delay angle. If the d-c bias were increased beyond this value, the grid-cathode voltage would always be negative and conduction would never occur.

**14-9. $p$-phase Rectifiers with Delay.**[6] From the discussion in the foregoing section, it should be clear that the d-c output current, and so the d-c output voltage of a controlled polyphase rectifier, depends upon the angle of delay. In order to obtain explicit expressions for this dependence, it will be assumed, for simplicity, that the tube drop, the transformer resistance, and the transformer leakage reactance are negligible.
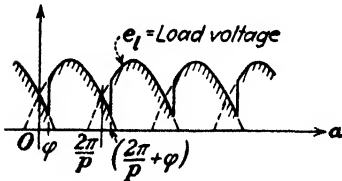


Fig. 14-35. The load voltage in a $p$-phase rectifier with delay.

An inspection of Fig. 14-35 reveals that conduction starts at the angle $\varphi$ and continues over the range $2\pi/p$, or until $\alpha = \varphi + (2\pi/p)$. The instantaneous load voltage in this range is

$$e_l = E_m \cos\left(\alpha - \frac{\pi}{p}\right)$$

Hence the d-c voltage is given by

$$E_{dc} = \frac{p}{2\pi} \int_{\varphi}^{\varphi+(2\pi/p)} E_m \cos\left(\alpha - \frac{\pi}{p}\right) d\alpha$$

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p} \cos \varphi \tag{14-46}$$

By comparing this result with Eq. (14-14) for the output d-c voltage from a system without delay, it is seen that

$$E_{dc} = E_{dc}{}^0 \cos \varphi \tag{14-47}$$

where $E_{dc}{}^0$ denotes the zero-delay d-c voltage given by Eq. (14-14). This simple result shows that the d-c output voltage at a delay angle $\varphi$ is equal to $\cos \varphi$ times the output with zero retardation.

Equation (14-47) actually represents the $B_0$ term in the Fourier series of the output wave. The harmonic terms can be found by the usual methods of analysis, just as was done in Sec. 14-2 for the output wave of the $p$-phase system without control. The analysis leads to the terms

and

$$
\left.\begin{array}{l}
B_k = -\dfrac{2E_{dc}{}^0}{k^2 - 1}\left(\cos k\varphi \cos \varphi + k \sin k\varphi \sin \varphi\right) \\[4mm]
A_k = -\dfrac{2E_{dc}{}^0}{k^2 - 1}\left(\sin k\varphi \cos \varphi - k \cos k\varphi \sin \varphi\right)
\end{array}\right\} \qquad (14\text{-}48)
$$

where

$$
k = np \qquad n = 1, 2, 3, \ldots
$$

It should be noted that the Fourier series contains sine components as well as cosine components since the wave no longer possesses zero axis symmetry. However, the wave still repeats at intervals of $2\pi/p$ rad. The latter condition again requires that only those frequency components which are integral multiples of $p$ are present. The magnitude of the $k$th harmonic term is given by $\sqrt{A_k{}^2 + B_k{}^2}$, which reduces to

$$
-\frac{2E_{dc}{}^0 \cos \varphi}{k^2 - 1} \sqrt{1 + k^2 \tan^2 \varphi}
$$

The complete Fourier series of the rectifier with delay is then given by

$$
e_l = E_{dc}\left[1 - \sum_{k = p, 2p, 3p, \ldots} \frac{2\sqrt{1 + k^2 \tan^2 \varphi}}{k^2 - 1} \cos (k\alpha - \psi_k)\right] \quad (14\text{-}49)
$$

where

$$
\tan \psi_k = \frac{A_k}{B_k}
$$

which is a function of $\varphi$.

If the retardation angle $\varphi$ is zero, this expression reduces to the Fourier series of the $p$-phase system without retardation, as it should [see Eq. (14-7)]. It is seen that the magnitude of each harmonic term of the system with delay is $\sqrt{1 + k^2 \tan^2 \varphi}$ times the corresponding term without delay. Since this term is greater than unity, it means that the ripple factor will be increased.

The foregoing analysis is not valid for all values of $\varphi$ but applies only for retardations within a limited range. For example, suppose that the angle of delay is greater than 30 deg in a three-phase system. The output from such a system is illustrated in Fig. 14-36. It is noticed that the current falls to zero for a certain portion of each cycle. If $\beta$ denotes the angle at which the current through anode 1 falls to zero, and with $\varphi$ the pickup angle of anode 2 the rectifier is nonconducting over the angular range $\varphi - \beta$. The load current thus consists of portions of sine waves separated by zero-current regions.
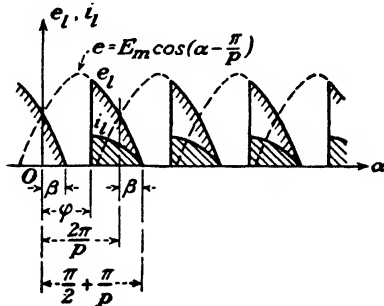


Fig. 14-36. A $p$-phase system with delay such that the load current is zero for a portion of the cycle. This condition prevails for all values of the delay angle $\varphi$ greater than $\beta$. For a three-phase rectifier $\beta = 30$ deg.

The maximum delay angle for which the current will pass continuously through the load is equal to the angle $\beta$. To find this angle, it is noted from the figure that

$$\frac{2\pi}{p} + \beta = \frac{\pi}{2} + \frac{\pi}{p}$$

from which

$$\beta = \frac{\pi}{2} - \frac{\pi}{p} \tag{14-50}$$

Clearly, the equations that have been derived above for $E_{dc}$ and $e_l$ are valid only for values of $\varphi$ that are less than $\beta$.

In order to obtain the d-c output voltage for angles of delay that are greater than $\beta$, it must be remembered that conduction will begin at the angle $\varphi$ and will stop when the anode voltage falls to zero. With respect to the origin $O$, as indicated in the figure, the anode voltage becomes zero at the angle $(\pi/2) + (\pi/p)$. Then

$$E_{dc} = \frac{p}{2\pi} \int_{\varphi}^{(\pi/2)+(\pi/p)} E_m \cos\left(\alpha - \frac{\pi}{p}\right) d\alpha$$

This integral becomes

$$E_{dc} = \frac{pE_m}{2\pi}\left[1 - \sin\left(\varphi - \frac{\pi}{p}\right)\right]$$

$$= \frac{E_{dc}{}^0}{2\sin(\pi/p)}\left[1 - \sin\left(\varphi - \frac{\pi}{p}\right)\right] \qquad \varphi \geq \beta \qquad (14\text{-}51)$$

The Fourier components for the present case can be found by straightforward analysis. The results are rather complicated in form, however. They can be found in the references cited (in graphical form for purposes of calculation).

It is seldom desirable in practice to permit the current to fall to zero for even a very short interval of time. It is customary, therefore, to make the load highly inductive so that such a condition cannot occur. *With suffi-*
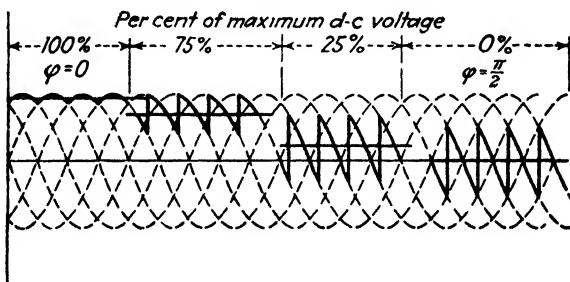


Fig. 14-37. The broken curves are the six-phase plate transformer voltages in a grid-controlled rectifier. The load consists of a resistor $R_l$ in series with an inductor $L$. The solid curves are the rectifier output voltages (across $R_l$ and $L$ in series) for various delay angles. The inductance is assumed so large that the voltage across $R_l$ is constant. This d-c voltage is indicated by the solid horizontal lines.

*cient inductance in the circuit* the instantaneous load current never falls to zero, and Eq. (14-47) will then be valid for all values of $\varphi$ from 0 to 90 deg. At this extreme condition $\cos\varphi = \cos 90° = 0$, and the output voltage is zero. The d-c output voltages for various values of delay angle are shown as heavy horizontal lines in Fig. 14-37 for a six-phase rectifier. Note that since the instantaneous output voltage (across $R_l$ and $L$ in series) equals the transformer voltage as long as the tube conducts, it may be negative. It is only required that the load current never become negative. This means that the average value of a solid curve must never be negative. The difference between a solid curve and the solid horizontal line through it is the instantaneous voltage which appears across the inductor $L$.

**14-10. The Effect of Leakage Reactance in Controlled Rectifiers.** The presence of leakage reactance in the rectifier transformers causes overlap

in a controlled rectifier, just as it does in rectifiers with zero delay angle. Figure 14-38 shows the output voltage and the anode currents under these conditions. The load voltage during the overlapping periods is the average value of the anode voltages of the overlapping anodes. This is the condition found in Eq. (14-36), and the reasons here are precisely the same
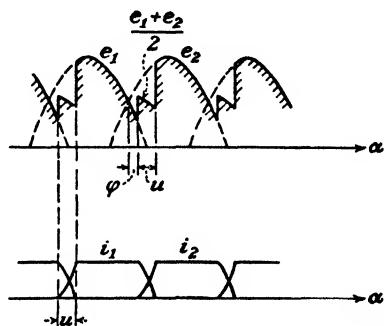


FIG. 14-38. The effect of leakage reactance on the operation of a controlled rectifier. The load current is assumed to be constant.

as for the case considered in Sec. 14-5. The remainder of the diagram should be clear from the discussion in the preceding sections.

The d-c output voltage under the present conditions of delay plus overlap is found to be

$$E_{dc} = E'_{dc}{}^0 \left[ \frac{\cos \varphi + \cos (\varphi + u)}{2} \right] \tag{14-52}$$

which reduces, of course, to the forms deduced previously for the cases where either $\varphi$ or $u$ or both are zero.

The generation of a large number of harmonics on the output, or d-c, side of a rectifier will obviously result in the presence of harmonics in the input. The presence of these harmonics in either the a-c or the d-c side may cause considerable disturbance in neighboring telephone lines.[7] It is usually possible to avoid this difficulty by careful disposition of the rectifier and the communication feeders. In a few cases, it has been found necessary to use resonant shunts, one for each disturbing harmonic, in order to reduce this telephone interference to a tolerable value.

**14-11. Controlled Rectifier Applications.** Because the output voltage can be varied from a maximum voltage to zero, controlled rectifiers are adaptable to varied applications. One such application is a variable-speed drive for d-c motors.[8] The armature of a separately excited motor is fed from the rectifier, and any desired speed may be maintained by adjusting the grid to give any desired voltage. Furthermore, the motor can be started or stopped by means of the grid controlling unit.

The voltage regulation of a mercury-arc rectifier is about 5 per cent.[9] This decrease in voltage with load results from the transformer leakage reactance, as discussed in Sec. 14-5. However, by employing controlled rectifiers, it is possible to obtain characteristics resembling those of a d-c compound-wound generator. This is accomplished by employing a small delay at zero load and then decreasing this delay as the load increases. In this way, either effective flat compounding or effective overcompounding can be accomplished. The variation of delay angle with load can be manually controlled or may be made automatic by employing proper controllers.[10] In a similar manner, it is possible to obtain constant d-c output voltage despite variations of the alternating input voltage. In this case, it is necessary to employ circuits that regulate the delay angle with variations in the input voltage.

Any attempt to obtain a large range of voltages with grid control results in the introduction of a large number of harmonic terms. This results in a poor power factor and an increased ripple. In order to avoid these disadvantages of grid control, it is sometimes advisable to use tapped rectifier transformers. This makes it possible to provide large changes in voltage by varying the transformer tap, the smooth variation between such changes being provided by the grid-control feature.

Controlled rectifiers are desirable for use on systems that may be damaged by voltage surges. Such controlled units permit the voltage to be brought up to its rated value slowly. For example, the use of a controlled rectifier for the power supply of a radio transmitter considerably decreases the initial surge charging rate of the filter capacitors.[11] In addition, the plate potential applied to the transmitting tubes can be applied slowly.

### PROBLEMS

**14-1.** Prove that the ripple factor of a $p$-phase gas-rectifier system is approximately given by

$$r = \frac{\sqrt{2}}{p^2 - 1} \frac{1}{1 - E_0/E_1}$$

where $E_1 = E_m (p/\pi) \sin (\pi/p)$ and $E_0$ is the constant tube drop.

**14-2.** Prove that for a $p$-phase vacuum rectifier

$$E_{dc} = E_m \frac{p}{\pi} \sin \frac{\pi}{p} - I_{dc} r_p$$

and

$$\eta_r = \frac{2v}{\pi} \frac{\sin^2 (\pi/p)}{(\pi/p) + \sin (\pi/p) \cos (\pi/p)} \frac{100\%}{1 + r_p/R_l}$$

where the symbols have the meanings used in the text. It is assumed that $p > 1$, that bridge-type circuits are *not* under consideration, and that each tube conducts for $2\pi/p$ rad in each cycle.

**14-3.** Consider the three-phase zigzag rectifier circuit of Fig. 14-5. Draw complete wiring and sinor diagrams for the six secondary windings. Show that, if thé windings are connected properly, balanced three-phase voltages are obtained, each of magnitude $\sqrt{3}\,E$ volts, where $E$ is the magnitude of one of the six secondary windings.

**14-4.** Show that the transformer currents in the three-phase bridge circuit of Fig. 14-9 are as depicted in Fig. 14-20.

**14-5.** It is found that an appreciable amount of telephone interference occurs because of the proximity of telephone lines to the feeders from a polyphase rectifier. In order
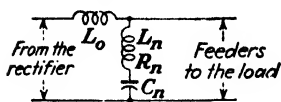


PROB. 14-5.

to reduce this interference, resonant shunts, of the type shown, are used across the lines. $L_n$ and $C_n$ are adjusted for resonance at the $n$th harmonic, which is to be suppressed. The resistance of the series choke $L_0$ can be neglected. The resistance $R_n$ of $L_n$ is small compared with the load resistance and with the reactance of $L_0$ at the $n$th harmonic angular frequency $np\omega$.
Show that under these conditions the $n$th harmonic voltage from the rectifier is multiplied by the factor $R_n/np\omega L_0$ at the feeders.

**14-6.** If there is a large enough inductance in series with the load so that the load current may be considered to be constant, show that the efficiency of rectification is given by $E_{dc}/(E_{dc} + E_0)$ for those circuits in which a single tube conducts at a time.

**14-7.** Sketch the "block" waves for the secondary and primary transformer currents in a double-Y circuit, assuming constant output current. Find the primary and secondary utilization factors.

**14-8.** Sketch the "block" waves for the *line* currents in the following polyphase gas rectifier systems:

a. Three-phase half-wave.
b. Six-phase half-wave.
c. Three-phase bridge.
d. Double-Y.

The primary-transformer bank is assumed to be Δ-connected.

**14-9.** Show that the Fourier series of the triangular wave shown is given by

$$e = -\frac{8A}{\pi^2}\left(\cos\alpha + \frac{1}{3^2}\cos 3\alpha + \frac{1}{5^2}\cos 5\alpha + \cdots\right)$$

Compare the first few terms in this series with those of Eq. (14-44) for the voltage across the interphase reactor in a double-Y polyphase rectifier.
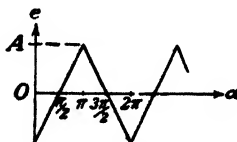


PROB. 14-9.

**14-10.** a. Calculate the inductance of the commutating reactor in a double-Y 625-volt 1,000-kw tank rectifier in order that the "critical" point occur at 2 per cent of the full load current.

*b.* Calculate the maximum permissible transformer leakage inductance if a 4 per cent regulation from critical load to full load is desired.

**14-11.** Given a 200-kw 600-volt double-Y rectifier system. Calculate

*a.* The transformer secondary voltage.

*b.* The efficiency of rectification.

*c.* The direct current in each anode.

*d.* The rms current in each anode.

*e.* The kva rating of the transformer primaries and secondaries.

Assume a constant tube drop of 20 volts. The load is so inductive that the output current may be considered to be constant. Neglect overlap.

**14-12.** In the preceding problem the net leakage reactance of each transformer is 0.4 millihenry. Calculate

*a.* The angle of overlap.

*b.* The percentage regulation from critical load to full load.

*c.* The percentage regulation from no load to full load.

**14-13.** Verify all the items in one of the columns of Table 14-1. (The column is to be designated by the instructor.)

**14-14.** In the double-Y rectifier does the circulating current $i_c$ depend upon the load current

*a.* For loads above the critical load?

*b.* For loads below the critical load?

Explain carefully.

**14-15.** The tube drop in a 3,000-kw 600-volt three-phase double-Y rectifier is 15 volts.

*a.* Calculate the size of inductance needed to reduce the ripple factor to one-tenth of 1 per cent at full load.

*b.* Calculate the ripple factor at half rated load for the value of inductance used in part *a.*

**14-16.** A 4-henry choke having a 5-amp d-c rating and a 0.25-$\mu$f capacitor (20,000-volt rating) are available. Calculate the percentage improvement possible when an L-section filter is used instead of a simple inductance choke in the output of the three-phase bridge rectifier of the illustrative example of page 440. Calculate the ripple voltage in each case.

**14-17.** The circuit of a six-phase forked Y rectifier is illustrated. Determine the anode currents, the primary winding currents, and the line currents. Assume that the primary windings are connected first in $\Delta$ and then in Y.
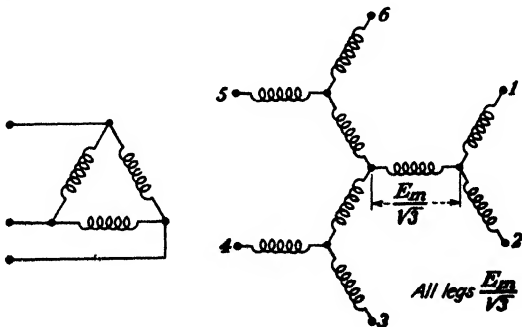


PROB. 14-17.

**14-18.** Repeat Prob. 14-17 for the twelve-phase forked star shown.



Inner legs $0.815E_m$
Outer legs $0.30E_m$
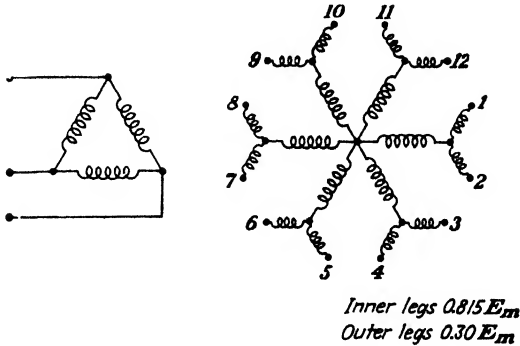
PROB. 14-18.

**14-19.** Each transformer of a group of three contains the primary winding, four windings of relative voltage $0.815E_m$, and four windings of relative voltage $0.30E_m$. These transformers are to feed a rectifier system in quadruple zigzag (two interconnected zigzag double-Y systems as shown in Fig. 14-29) to yield a 12-phase output.

*a.* Discuss the reasons for including such windings.

*b.* What frequencies appear across each commutating reactor?

*c.* Determine the anode currents, the primary winding currents, and the line currents of this system. Assume that the primary windings are connected first in Δ and then in Y.

**14-20.** Repeat Prob. 14-19 for the zigzag double six-phase circuit shown.



Commutating
reactor

Inner legs $0.815E_m$
Outer legs $0.30E_m$
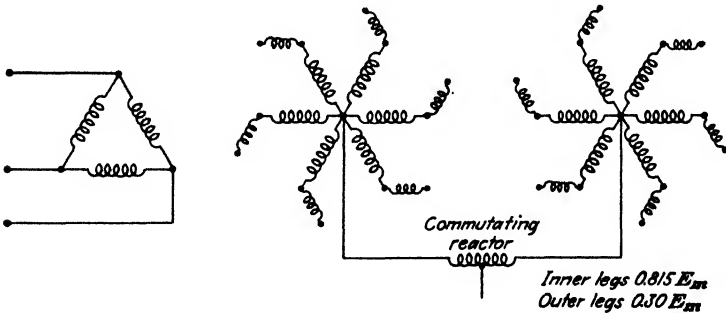
PROB. 14-20.

**14-21.** Obtain the Fourier components of the current wave in a $p$-phase rectifier with delay; *i.e.*, verify Eqs. (14-47) and (14-48).

**14-22.** The leakage reactance of the transformers that supply a three-phase 625-volt 1,000-kw grid-controlled tank rectifier causes a full-load regulation of 5 per cent. At what delay angle must the system be set at no load in order to have flat compounding? The delay angle is, of course, reduced to zero as the output increases from zero to full load.

**14-23.** Verify Eq. (14-52), giving the magnitude of the d-c output voltage, taking into account both delay and overlap.

## REFERENCES

1. MARTI, O. K., and H. WINOGRAD, "Mercury Arc Rectifiers—Theory and Practice,' McGraw-Hill Book Company, Inc., New York, 1930.
2. DORTORT, I. K., *Allis-Chalmers Elec. Rev.*, **4**, 9, 1939.
3. MARTI and WINOGRAD, *op. cit.*
   READ, J. C., *Engineer*, **163**, 418, 1947.
4. COX, J. H., and G. F. JONES, *Trans. AIEE*, **58**, 618, 1939.
5. JOURNEAUX, D., *Elec. Eng.*, **53**, 976, 1934.
6. DEMONTVIGNER, M., *Rev. gén. élec.*, **32**, 625, 659, 1932.
   BROWN, H. D., and J. J. SMITH, *Trans. AIEE*, **52**, 973, 1933.
   HERSKIND, C. C., *Elec. Eng.*, **53**, 926, 1934.
   STEBBINS, F. O., and C. W. FRICK, *ibid.*, **53**, 1259, 1934.
   RISSIK, H., *Electrician*, **123**, 146, 1939.
7. Edison Electric Institute, *Pub.* E-1, April, 1937.
   EVANS, R. D., and H. N. MULLER, *Trans. AIEE*, **58**, 861, 1939.
   RISSIK, H., *Electrician*, **124**, 37, 1940.
8. Allis-Chalmers Manufacturing Co., *Bulletin* 1176, Milwaukee, Wis.
9. KILGORE, L. A., and J. H. Fox, *Elec. Eng.*, **56**, 1134, 1937.
10. GOODHUE, W. M., and R. B. POWER, *ibid.*, **55**, 1200, 1936.
11. DURAND, S. R., and O. KELLER, *Proc. IRE*, **25**, 570, 1937.

### General References

ARMSTRONG, R. W.: *Proc. IRE*, **19**, 78, 1931.
EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," 3d ed., McGraw-Hill Book Company, Inc., New York, 1949.
GLASER, A., and MÜLLER-LUBECK: "Einführung in die Theorie der Stromrichter," Verlag Julius Springer, Berlin, 1935.
GUNTHERSCHULZE, A., and N. A. DeBRUYNE: "Electric Rectifiers and Valves," John Wiley & Sons, Inc., New York, 1928.
JOLLEY, L. B. W.: "Alternating Current Rectification and Allied Problems," John Wiley & Sons, Inc., New York, 1927.
MARTI, O. K., and H. WINOGRAD: "Mercury Arc Rectifiers—Theory and Practice," McGraw-Hill Book Company, Inc., New York, 1930.
MIT Staff: "Applied Electronics," Chap. VII, John Wiley & Sons, Inc., New York, 1943.
PRINCE, D. C., and F. B. VOGDES: "Principles of Mercury Arc Rectifiers and Their Circuits," McGraw-Hill Book Company, Inc., New York, 1927.
REICH, H. J.: "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.
RISSIK, H.: "The Rectification of Alternating Current," English Universities Press, London, 1938.

# CHAPTER 15

# PHOTOELECTRICITY AND PHOTOELECTRIC CELLS

$\mathrm{T}$HE liberation of electrons from matter under the influence of light is known as the *photoelectric effect*. This effect was discovered by Hertz [1] in 1887. This subject has received the attention of physicists since that time although photoelectricity was not put to much practical use until about 1928, when the phototube was developed for the sound unit of the "talking" motion pictures. As the quality of the commercial cells has improved, the range of application has been extended to a very varied field.

Three classes of phenomena are classified as photoelectric phenomena: (1) the liberation of electrons under the influence of light; (2) the generation of an emf by chemical or physical reactions produced by the light; (3) the changes of electrical conductivity with changes in light intensity. The reactions of the first type are termed "photosensitive," the usual photoelectric cell being the most common example. The second class of reaction is termed "photovoltaic," the Photronic and Photox cells being of this type. The third class of reaction is termed "photoconductive," an example being furnished by the changes in electrical conductivity of selenium under the action of light.

**15-1. Photoemissivity.** The experimental features that are characteristic of the photoemissive effect are the following: [2]

1. The photoelectrons liberated from the photosensitive surface possess a range of initial velocities. However, a definite negative potential when applied between the collector and the emitting surface will retard the fastest moving electrons. This indicates that the emitted electrons are liberated with all velocities from zero to a definite maximum value. The maximum velocity of the emitted electrons is given by the relation

$$\tfrac{1}{2}mv_{\mathrm{max}}^2 = eE_r \qquad (15\text{-}1)$$

where $E_r$ is the retarding potential, in volts, necessary to reduce the photocurrent to zero. As the potential is increased, the number of electrons to the collector increases until saturation occurs. It is to be noted from Fig. 15-1, in which are plotted curves showing the variation of photocurrent $I_{p\lambda}$ with collector potential (the light intensity $j$ is a parameter), that $E_r$ and hence $v_{\mathrm{max}}$ are independent of the light intensity.

454

2. If the photoelectric current is measured as a function of the collector potential for different light frequencies $f$ and equal intensities of the incident light, the results obtained are essentially those illustrated in Fig. 15-2.[3] It is observed that the greater the frequency of the incident light, the greater must be the retarding potential to reduce the photocurrent to zero. This means, of course, that the maximum velocity of emission of the photoelectrons increases with the frequency of the incident light in a
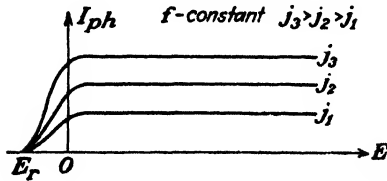


FIG. 15-1. Photocurrent vs. plate voltage with light intensity as a parameter. The frequency of the incident light is a constant.
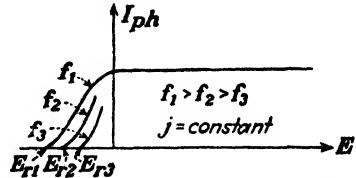
FIG. 15-2. Photocurrent vs. plate potential with the frequency of incident light as a parameter. The light intensity is a constant.

measurable way. Experimentally, it is found that a linear relationship exists between $E_r$ and $f$.

The experimental facts listed under 1 and 2 may be summarized in the statement that the *maximum energy of the electrons liberated photoelectrically is independent of the light intensity but varies linearly with the frequency of the incident light.*

3. If the saturation current is plotted as a function of the light intensity, the results show that the photoelectric current is directly proportional to the intensity of the light.

4. The foregoing photoelectric characteristics are practically independent of temperature, within wide ranges of temperature.

5. The electrons are emitted immediately upon the exposure of the surface to light. The time lag has been experimentally determined to be less than $3 \times 10^{-9}$ sec.[4]

6. Photoelectric cells are selective devices. This means that a given intensity of light of one wave length, say, red light, will not liberate the same number of electrons as an equal intensity of light of another wave length, say, blue light. That is, the *photoelectric yield*, defined as the photocurrent (in amperes) per watt of incident light, depends upon the frequency of the light. Alternative designations of the term "photoelectric yield" to be found in the literature are relative response, quantum yield, spectral sensitivity, specific photosensitivity, and current-wave-length characteristic. The relative response curves for the alkali metals are shown in Fig. 15-3.

Curves of these types are obtained experimentally in the following way: Light from an incandescent source is passed through the prism of a mono-chromator for dispersion, a narrow band of wave lengths being selected by means of an appropriately placed slit system. The current given by the photoelectric surface when exposed to the light passing through the system of slits is noted. The current given by a blackened thermopile when exposed to the same light is also noted. The ratio of these two readings is plotted vs. the wave length of the incident light. Blackened thermopiles are used because they absorb all radiation incident upon them equally, regardless of the wave length. This permits a measure of the energy contained in any part of the spectrum to be made. An automatic spectral sensitivity curve tracer has been designed for obtaining these curves quickly with the aid of a cathode-ray tube.[5]



FIG. 15-3. Spectral sensitivity as a function of wave length for the alkali metals. (*E. F. Seiler, Astrophys. J.,* **52**, 129, 1920.)

**15-2. Photoelectric Theory.** The foregoing experimental facts find their explanation in the electronic theory of metals and in the light-quantum hypothesis of Planck. As already noted, the light-quantum hypothesis regards light as consisting of photons (discrete units or bundles of light) that possess some characteristics similar to those of material particles. Associated with light of frequency $f$ are a number of photons, each of which has an associated energy, $hf$ joules, where $h$ is Planck's constant. The greater the intensity of the light, the larger is the number of photons present, but the energy of each photon remains unchanged. Of course, if the light beam is heterogeneous rather than monochromatic, then the energy of the photons therewith associated will vary and will depend upon the frequency.

If monochromatic light of frequency $f$ falls upon a metal whose work function is $E_W$, the velocity of the emitted electron is, according to Einstein's equation,[6]

$$\tfrac{1}{2}mv^2 \leqq hf - eE_W \qquad (15\text{-}2)$$

The significance of this equation becomes apparent when considered in the light of the electronic theory of matter. Since photoelectric devices are operated at normal temperatures, the completely degenerate distribution function must be employed. This is given in Fig. 15-4, which shows the energy distribution function at low temperatures, and also the potential-energy barrier at the surface of the metal (see Chap. 4).

This figure indicates that the electrons within the metal exist in energy levels ranging from zero to a maximum energy, $E_M$ electron volts, but none have energies greater than this value. If an electron possessing the energy $E_M$ receives the photon of light energy $hf$ and travels normal to the surface of the metal, the kinetic energy that it will have, upon escaping from the metal, will be $hf - eE_W$ joules. This follows directly from the significance of the work function $E_W$, which is the minimum energy that must be supplied in order to permit the fastest moving surface-directed
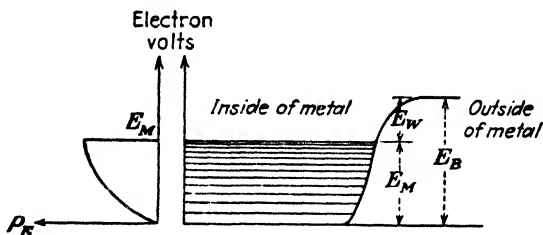


FIG. 15-4. Energy-level diagram for the free electrons within a metal. The potential-energy barrier at the surface of the metal is also shown.

electrons just to surmount the potential barrier at the surface of the metal and to escape.

Since some of the electrons in levels other than the highest energy states may absorb the incident photons, an energy greater in magnitude than $E_W$ will be expended when they escape. This fact explains the inequality of Eq. (15-2)

According to this equation, the retarding potential $E_r$ that will just repel the fastest moving electron is given by

$$eE_r = \tfrac{1}{2}mv_{\max}^2 = hf - eE_W \tag{15-3}$$

which is in agreement with the experimental facts 1 and 2. This result shows that the maximum energy of the escaping electrons varies linearly with the frequency and is independent of the light intensity. This latter condition follows from the fact that the intensity of the incident light does not enter into this expression. This equation was experimentally verified by Millikan.[7] He plotted retarding voltage vs. frequency and obtained a straight line. The slope of this line gives the value of the ratio $h/e$. The value of this ratio by this method agrees very well with that found from other experiments. The intercept of the Einstein line with the axis of abscissa gives $E_W$ (provided that corrections are made for contact difference of potential). The value of the work function obtained photoelectrically agrees well with that measured thermionically for the same emitter.

The minimum frequency of light, known as the "threshold frequency," that can be used to cause photoelectric emission can be found from the equation above by setting the velocity equal to zero. The result is

$$f_c = \frac{eE_W}{h} \tag{15-4}$$

The corresponding wave length, known as the "long-wave-length limit" or the "threshold wave length," beyond which photoelectric emission cannot take place is

$$\lambda_c = \frac{c}{f_c} = \frac{ch}{eE_W} = \frac{12,400}{E_W} \quad \text{A} \tag{15-5}$$

where A denotes the angstrom unit ($10^{-10}$ m) and where $E_W$ is expressed in electron volts. For response over the entire visible region, 4,000 to 8,000 A, the work function of the photosensitive surface must be less than 1.55 volts. This follows directly from Eq. (15-5).

The similarity between this equation and Eq. (9-5) is evident. Yet the two expressions represent physical processes that are almost direct complements of each other. In the present case the energy is transferred to an electron by a light quantum. In Chap. 9, the light quantum was emitted when electrons jumped from one atomic energy level to another.

*Example.* A tungsten surface having a work function of 4.52 ev is irradiated with the mercury line, 2,537 A. What is the maximum speed of the emitted electrons?

*Solution.* The electron-volt equivalent of the energy of the incident photons is $12,400/2,537 = 4.88$ ev. According to the Einstein equation the maximum energy of the emitted electrons is

$$4.88 - 4.52 = 0.36 \text{ ev}$$

From Eq. (2-13) the corresponding velocity is

$$v_{\max} = 5.93 \times 10^5 \sqrt{0.36} = 3.56 \times 10^5 \text{ m/sec}$$

The fact that the photoelectric current is strictly proportional to the light intensity is readily explained. A greater light intensity merely denotes the presence of a larger number of photons. Further, since each photon is equally effective in ejecting electrons, the number of electrons per second ejected must be proportional to the light intensity.

If it is remembered that the distribution function of electrons in metals varies very little with temperature, then fact 4 is evident. Strictly speaking, however, the totally degenerate distribution function applies only at the temperature 0°K. At room temperature, therefore, a few electrons will have emission velocities greater than those predicted by Eq. (15-3). These considerations are illustrated in Fig. 15-5. A frequency slightly less than the threshold value has been chosen. The distribution function has been shifted up in energy by the amount $hf/e$ electron volts on the assump-

tion that each electron is capable of absorbing the incident photons $hf$ and so having its energy increased by this amount.

As is evident from an inspection of this curve, and as predicted by the Einstein equation, no photoelectric emission should exist under the conditions of absolute zero of temperature and for photons of frequency $f$. At room temperature, some electrons will possess sufficient energy to overcome the surface barrier, and these will be emitted. These are represented by the shading under the broken distribution curve. This shows that no absolutely sharp long-wave-length limit exists for any substance, since the
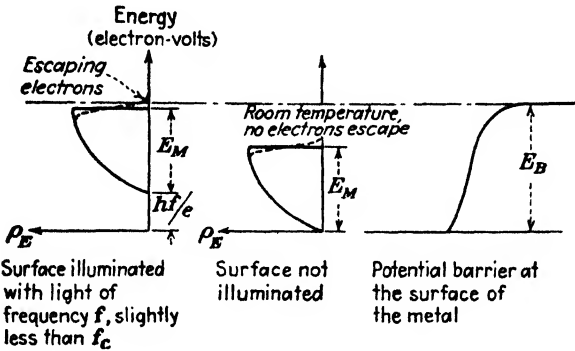


FIG. 15-5. The energy distribution of the free electrons within a metal with and without the surface illuminated (not drawn to scale).

curves, such as those of Fig. 15-1, theoretically approach the axis asymptotically. Fowler [8] investigated this matter theoretically, and this theory provides a method of determining photoelectric work function independent of the temperature of the surface.

According to the Fowler theory the relation between photoelectric current $I_{ph}$, surface temperature $T$, and frequency $f$ for constant intensity of the incident light can be represented by an analytic expression of the form

$$\log \frac{I_{ph}}{T^2} = B + \ln f(x) \tag{15-6}$$

where $f(x)$ is expressed as a series expansion in $x$. In this expression

$$x = \frac{hf - eE_W}{kT} \quad \text{and} \quad B = \text{a constant} \tag{15-7}$$

A plot of $\ln (I_{ph}/T^2)$ vs. $hf/kT$ as determined by Eq. (15-6) is shown in Fig. 15-6. This figure [2] shows the theoretical curve and also the form of typical experimental curves which are obtained at two different temperatures.

Indicated in Fig. 15-6 is the method for determining the work function, and so the photoelectric threshold, of the surface. This requires a displacement of the experimentally observed curve, both horizontally and vertically, in order to superpose it on the theoretical curve. The horizontal displacement is a measure of $eE_W/kT$, and thereby permits a calculation of $E_W$, since the temperature $T$ is known. The vertical displacement is a measure of $B$, which has no particular significance. Very accurate measurements verify this theory.[9] This permits the accurate determination of $E_W$.

For most practical purposes the use of the completely degenerate distribution function even at room temperatures is quite reasonable. Under these circumstances, it is justifiable to consider cutoff to occur sharply for frequencies below the critical value $f_c$.
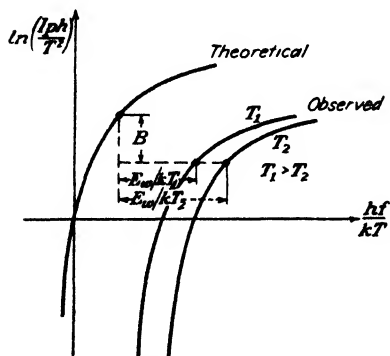


FIG. 15-6. The Fowler method of determining the true photoelectric work function.

A qualitative explanation for the shapes of the spectral response curves of Fig. 15-3 is readily possible. There can be no response for frequencies below $f_c$; hence cutoff occurs at the point $f = f_c$. As $f$ increases above $f_c$, the energy of the incident photon $hf$ increases and some electrons in levels below the maximum energy state are permitted to escape. As a result, the response increases as the frequency increases or, correspondingly, as the wave length decreases. However, a point of maximum response must exist. This follows from the fact that if the energy of the light is $W$ joules, then the number of photons in the light beam is $W/hf$. But since this number decreases with increasing frequency, the photocurrent must decrease as $f$ increases because of the decreased number of photons present. The decrease is accentuated by the rapid decrease in the efficiency of interaction between the radiation and the conduction electrons at the surface with increased frequency. A second peak is sometimes found to occur at the short wave lengths. This is ascribed to the interaction of the radiation with the more tightly bound conduction electrons of the matter. This *volume* photoeffect becomes significant for light in the violet or near ultraviolet. A complete quantitative explanation for the shapes of these curves has not yet been given.

**15-3. Phototubes.** The essential elements of a phototube are a sensitive cathode surface of large area, and a collecting electrode, contained in an evacuated glass bulb. These electrodes may be arranged in numerous ways. In the older type of tube, the cathode was made by distilling the

photosensitive material, generally an alkali metal (usually cesium), on the inner surface of the bulb, which first was silvered in order to ensure good conduction. The anode was made in the form of a straight wire or ring, in any case being small, so that it did not obstruct the light that was



| Nondirectional type | Conventional type | "Low-leakage" type | End type | Cartridge type |
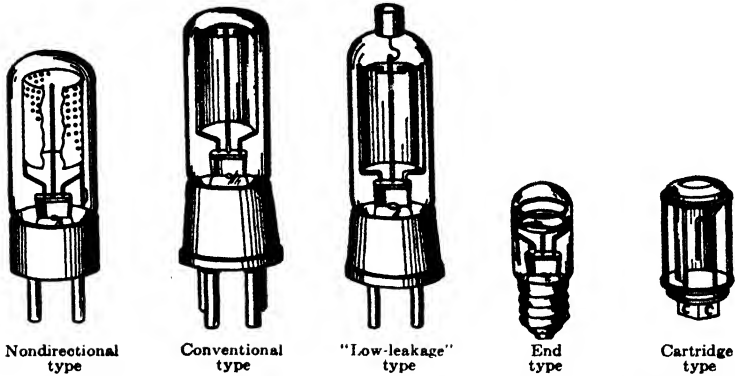
FIG. 15-7. A group of typical phototubes. (*Westinghouse Staff, "Industrial Electricity Reference Book," courtesy of John Wiley & Sons, Inc., New York, 1948.*)

incident upon the cathode. Many of the present-day phototubes consist of a semicylindrical metallic cathode on which the photosensitive substance has been evaporated. The anode is a straight wire that is practi-
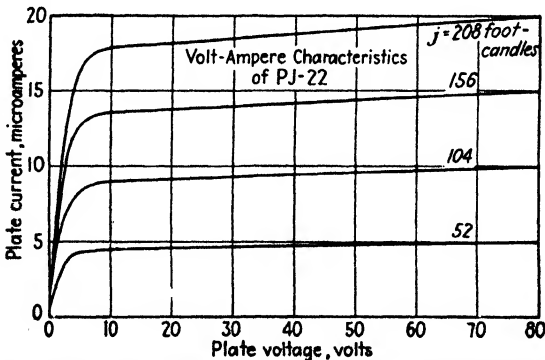


FIG. 15-8. Volt-ampere characteristics of a vacuum phototube with light intensity as a parameter.

cally coaxial with the cathode. Phototubes of modern design are shown in Fig. 15-7.

The glass bulb either may be highly evacuated or may contain an inert gas at low pressure. The volt-ampere characteristics of the ordinary vacuum phototubes are shown in Fig. 15-8. The current that exists at zero

accelerating potential results from the initial velocities of the electrons. Note that a retarding potential must be applied in order to reduce the current to zero.

As the anode-cathode potential is increased, the current to the anode increases very rapidly at first, the nonsaturation resulting from the possible space-charge effects, and also from the fact that some electrons are missing the wire anode on their journey from the cathode, since the attractive field is small at these low potentials. The current very soon reaches a saturation value, for the field becomes sufficient to attract all the electrons lib-
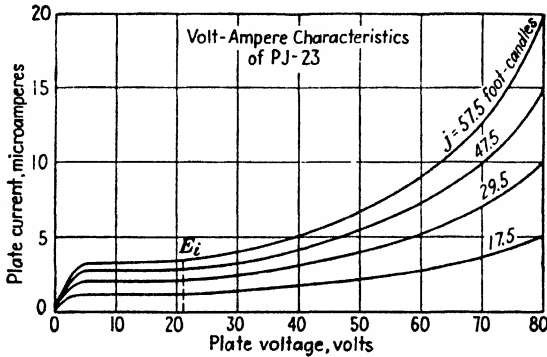


FIG. 15-9. Volt-ampere characteristics of a gas-filled phototube with light intensity as a parameter.

erated from the cathode under the influence of the incident light. The continued increase in photocurrent as the anode potential is increased results partly from the more complete collection of the electrons, and possibly from the reduction of the work function of the material of the cathode surface as a consequence of the presence of the applied electric field at the cathode (the Schottky effect; see Sec. 4-11).

By filling the glass envelope with an inert gas, such as neon or argon, at low pressure of the order of 0.5 mm, the current yield for a given intensity of illumination is greatly increased, as illustrated in Fig. 15-9. As described in Sec. 10-2, the increased current is produced by the field-intensified or Townsend discharge. It is important never to raise the potential across the tube to the point where a glow discharge occurs, for this will cause cathode sputtering with a consequent permanent damage to the cathode surface.

For purposes of comparison, the outputs of the vacuum and the gas-filled phototubes (specifically, the results obtained under identical experimental conditions with the General Electric PJ-22 and PJ-23 phototubes) are illustrated in Fig. 15-10. These tubes are identical in all respects, except for the fact that one is a vacuum tube and the other is an argon-filled tube. It is observed that the curves have the same shape until

ionization by collision occurs in the gas tube. Beyond this point, the sensitivity of the gas tube continues to increase with increased anode potentials, owing to the contribution to the current by the electrons resulting from the ionization of the gas. The gas amplification ratio of a gas-filled tube is of the order of 7 for an anode voltage of 80 volts. From Fig. 15-10, this ratio is seen to depend upon the amount of incident light flux and upon the applied anode voltage. If an attempt is made to obtain
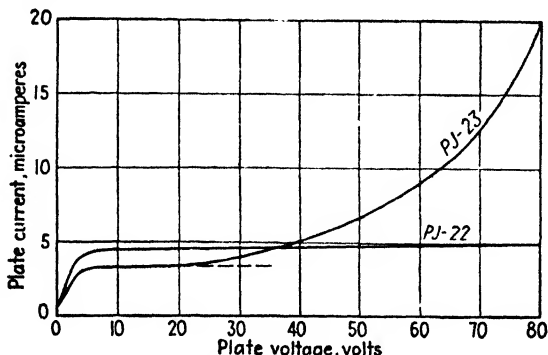


Fɪɢ. 15-10. Comparison of the volt-ampere characteristics of vacuum and gas-filled tubes with similar photosurfaces. The same light intensity was used for both tubes.

amplification ratios larger than about 10, a glow discharge usually takes place.

**15-4. Sensitivity of Phototubes.** According to the above discussion, the *static* sensitivity of a phototube may be defined as the ratio of the d-c anode current to the incident radiant flux of constant value. In a similar way, the *dynamic* sensitivity of a phototube is the ratio of the alternating component of the anode current to the alternating component of the incident radiant flux. The dynamic characteristics are very important in sound projectors, television, and facsimile systems and in fact in any system which depends for its operation on changes in light intensity.

For a vacuum phototube, the two sensitivities are equal. For a gas-filled photocell, there is a falling off of sensitivity with an increase in the frequency at which the incident light is interrupted or modulated. The average sensitivity characteristic of an RCA type 868 gas-filled phototube is shown in Fig. 15-11. The explanation of the shape of this curve lies in the amplification characteristics of the gas contained in the cell. It takes a negligible time for the photoelectrons to appear after the cathode surface has been illuminated. However, a finite amount of time is required for these electrons to build up the steady-state amplified current in the cell. This results from two causes. The one of minor importance is the time required [10] for the relatively slow ions that are formed by the electron collision process to travel to the cathode. As mentioned in Sec. 10-4, these

times are measured in microseconds and would not cause any difficulties. The more important limitation is imposed by the drift or diffusion times of the metastable atoms, which move slowly, as they are electrically neutral and are not affected by the applied field. These metastable ions
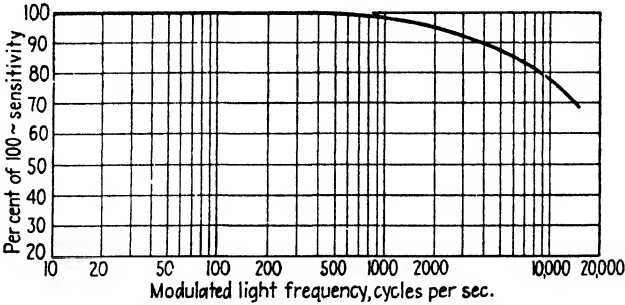


Fig. 15-11. Average sensitivity characteristic of an RCA 868 gas-filled phototube. The anode voltage is 90 volts. (*Courtesy of RCA*.)

bombard the cathode and emit electrons. Such metastable diffusion times are of the order of milliseconds, so that for modulating frequencies above 1,000 cps the sensitivity will drop, as indicated in Fig. 15-11. This feature restricts such gas tubes to applications for lower frequencies of light inter-
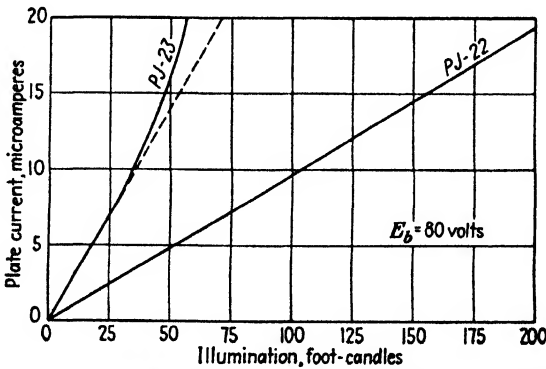


Fig. 15-12. Photocurrent as a function of illumination for a vacuum cell (PJ-22) and a gas-filled cell (PJ-23).

ruption, since a sudden change in light intensity is not accompanied by corresponding instantaneous change in photocurrent.

Another disadvantage of gas-filled cells is the lack of linearity of current with incident flux. The photocurrent increases more rapidly than the illumination for anode voltages that are higher than the ionization potential of the gas. This is illustrated in Fig. 15-12 for the PJ-23 tube. The

linear response of the vacuum cell is also shown. Gas phototubes are used primarily for on-off (relay) operation or with sound-reproduction equipment, where the slight nonlinearity and the variation of dynamic sensitivity with frequency are not too important.

It should be kept in mind that the curves shown in this chapter, and also those supplied by the phototube manufacturers, are typical of the operation, rather than specific for any particular tube type. Large variations may exist in the characteristics of phototubes manufactured under presumably identical conditions. This results from the fact, already noted,
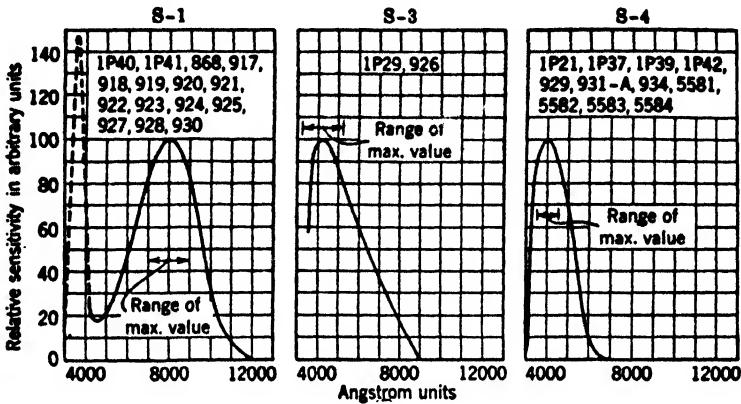


FIG. 15-13. The relative response of three commercial surfaces as a function of wave length. The phototubes using these surfaces are listed on each diagram. (*From Zworykin and Ramberg, "Photoelectricity and Its Applications," courtesy of John Wiley & Sons, Inc.*)

that the number of photoelectrons emitted for a given illumination varies appreciably for even slight changes in the surface preparation of the cathode. For the same reason, it is often found that different portions of the same emitting surface may possess different sensitivities. It is advisable, therefore, to illuminate a large part of the cathode uniformly whenever possible, rather than to focus the light source on only a portion of the photoemissive surface.

In any particular application, careful consideration must be given to the choice of the light source as well as to the photocell characteristics. For example, it is desirable that the source emit strongly in the frequency range in which the photocell is most sensitive, if large photocurrents are to be obtained. Commercial phototubes are now available with photoelectric yields [11] that have peaks in various portions of the visible spectrum. Figure 15-13 shows the spectral response of the three most common photosurfaces.

Surface *S*-1 consists of a composite silver-cesium oxide-cesium surface. Such a surface is sensitive throughout the entire visible region and has a

fair sensitivity in the infrared. As a result, this composite surface is used extensively in commercial phototubes. Surface $S$-3 is a silver-rubidium oxide-rubidium surface which has a sensitivity largely confined to the visible region, although it has its greatest sensitivity in the blue end of the spectrum. Surface $S$-4 shows the response of an antimony-cesium surface that is very sensitive to the green, blue, and near ultraviolet and is insensitive to red and infrared radiation. This is the most sensitive surface commercially available today. The tube has a sensitivity of 120 $\mu$a/lumen when daylight is used as the source. When a tungsten lamp, operating at a filament temperature of 2870°K, is the source of light, the sensitivity of the $S$-1 surface is 20, of the $S$-3 surface 6.5, and of the $S$-4 surface 45 $\mu$a/lumen.
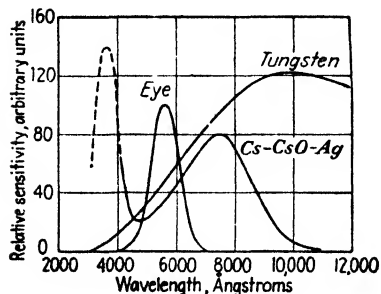


Fig. 15-14. The relative sensitivity of the eye, a Cs-CsO-Ag photosurface, and the energy distribution from an incandescent tungsten lamp as a function of wave length.

Figure 15-14 contains curves showing the spectral sensitivities of the Cs-CsO-Ag photosurface, the eye, and the energy distribution curve of the output from an incandescent tungsten lamp for purposes of comparison. It will be noticed that the Cs-CsO-Ag photosurface is sensitive over a wide range of wave lengths and so very well adapted for use with lamps as the source of illumination. Of course, if a photocell is desired that has a response somewhat resembling that of the human eye, the Cs-CsO-Ag surface would not be suitable.

In order to provide reasonably large currents from phototubes that are used with very low light levels, units that incorporate electrostatic electron multipliers are available.[12] Figure 3-19$b$ gives the electrode structure of one such type of tube. These tubes provide currents of almost 1 ma when the luminous flux on the cell is only 1 millilumen. By comparing this figure with the sensitivity of the PJ-23 phototube, it is seen that the multiplier increases the sensitivity by approximately 500,000 times.



Fig. 15-15. The basic phototube circuit.

**15-5. Phototube Applications.** The basic circuit employing a phototube is shown in Fig. 15-15. As the luminous flux that is incident on the cell varies, the output current changes and a changing voltage appears across the load resistor $R_l$. Although the basic circuits are the same, there are three important types of application of phototubes: (1) A definite fixed amount of illumination is to be measured. (2) Rapid

variations in light intensity are to be faithfully reproduced.  (3) A definite large change in light intensity is to be detected.

The field of photometry and colorimetry offers many examples of the first type of application.  In such cases, $R_l$ might simply be the internal resistance of the indicating instrument.  If the incident light is too small to be measurable directly, a d-c amplifier might be used.  In this case, $R_l$ will be the input resistance of the amplifier.  The light beam of varying intensity that has been modulated by the sound track of a motion-picture film or by the scanning process in a television tube is of the second class. Applications of the third type are exemplified by "on" and "off" circuits. In such cases the phototube is used in conjunction with a relay so that some circuit is either energized or deenergized when the light intensity exceeds or falls below some preassigned value.  Many of the common applications of the "electric eye" belong to this third class.  A few illustrations are the counting or sorting of objects on a conveyer belt; the automatic opening of a door as it is approached; devices for the protection of human life; and fire-alarm systems.[18]

As a specific illustration of one type of "on" and "off" control, the following demonstration experiment is offered: An electric lamp is to be "lit" by means of a match and is to be "extinguished" by a wave of the hand.  The circuit for accomplishing this action is shown in Fig. 15-16.  The phototube load resistor is the grid resistor of a thyratron.
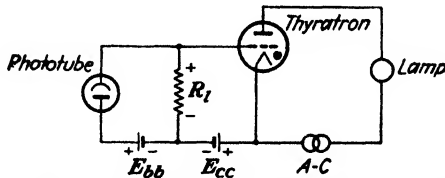


FIG. 15-16.  Phototube control of a thyratron circuit.

The rating of the thyratron must be high enough to supply the lamp current.  The plate circuit of the thyratron is fed from the a-c line.

As long as the phototube is dark, no voltage appears across $R_l$ and the bias voltage $E_{cc}$ is sufficiently negative to prevent the thyratron from firing.  However, if a match is lit in front of the lamp so that the light rays strike the photocell, a voltage is developed across $R_l$.  The net cathode-grid voltage is greater than the critical grid value of the thyratron for some point on the alternating plate voltage wave.  The lamp lights; and if its luminous flux is intercepted by the photocell, the lamp remains on.  If the hand is interposed between the photocell and the lamp so that the cell is dark for a brief interval of time (long enough for the lamp filament to cool), the lamp will remain extinguished. This action follows from the fact that once the plate voltage wave passes through zero at the end of a half cycle the grid again regains control and the thyratron will not fire again.

In order to determine the current that will flow in a phototube circuit for a given light flux, battery voltage, and load resistance, it is necessary

to use the volt-ampere tube characteristics. The straight line, expressed by the relation

$$e_b = E_{bb} - i_b R_l \qquad (15\text{-}8)$$

is superposed on this set of static characteristics. This is the same load line that was discussed in connection with the diode rectifier in Sec. 12-1. It is drawn through the point $i_b = 0$, $e_b = E_{bb}$, and with a slope determined by the load resistor $R_l$, as shown in Fig. 15-17.

The intersection of the load line with each volt-ampere curve gives the current output at the value of intensity for which that curve was con-
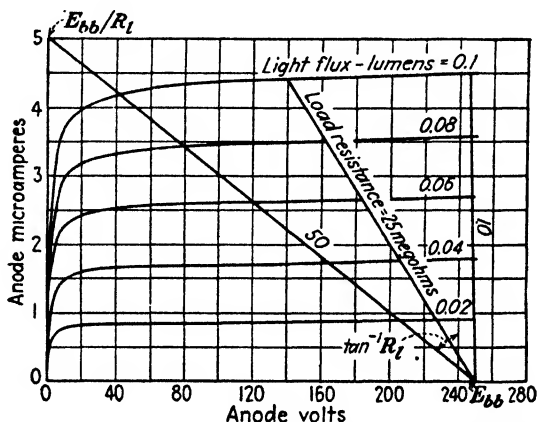


Fig. 15-17. Volt-ampere characteristics of an RCA 929 vacuum phototube. The load lines for a plate supply voltage of 250 volts and resistances of 1.0, 25, and 50 megohms are also shown. (*Courtesy of RCA Manufacturing Co.*)

structed. In this way a curve of current vs. intensity or flux for each value of load resistance can be found. The curves for $R_l = 10$, 25, and 50 megohms and $E_{bb} = 250$ volts for the RCA 929 vacuum phototube are reproduced in Fig. 15-18. It is noted that these curves are practically linear and almost independent of the load resistance. This results from the fact that the volt-ampere characteristics of Fig. 15-17 are essentially horizontal lines that are equally spaced for equal intervals of light flux. Since, for a given light intensity, the plate current is nearly independent of voltage, except for small voltages, *the vacuum photocell may be considered to be a constant current generator*. This characteristic is made use of in certain applications.

The output-voltage curves corresponding to these current curves are shown in Fig. 15-19. The voltage drop across the load resistor is plotted as a function of light flux. It is observed that a given change in light flux results in larger changes in voltage for the higher load resistances. This would seem to favor very large values of load resistances. The practical

upper limit to the resistance that can be used in such a circuit is set by the leakage currents that are ever present. Thus, $R_l$ must be kept considerably below the resistance between the cathode and anode of the phototube
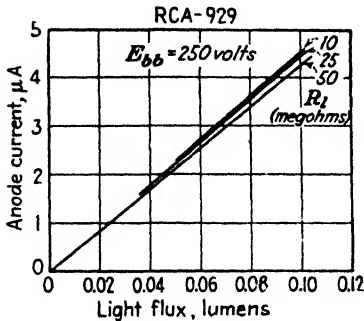


FIG. 15-18. Photocurrent as a function of light flux (dynamic curves) for load resistances of 10, 25, and 50 megohms. The plate supply voltage is 250 volts.
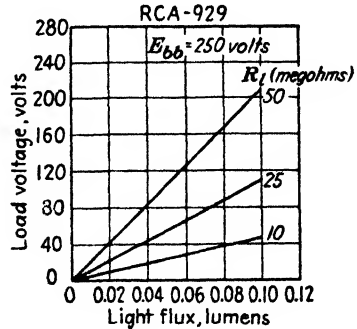


FIG. 15-19. Load voltages as a function of light flux corresponding to the current curves of Fig. 15-18.

(and also below that between the cathode and grid of the associated amplifier, if one is used). It is desirable, therefore, to reduce the leakage currents as much as possible. In certain tubes a long leakage path is provided by bringing the leads from the electrodes through opposite ends of the glass envelope (see Fig. 15-7). For very sensitive measurements, it is desirable to clean the glass surface carefully and then coat the bulb with a thin layer of ceresin wax. This reduces the leakage currents to very small values.

If the load resistance is too high, or if the plate supply voltage is low, the load line will intersect the volt-ampere curves for the higher intensities in the region near the origin, where the curves are close together. Under these circumstances, a curve of current vs. light intensity or light flux will no longer be linear. In fact, it will show a saturation value, as indicated in Fig. 15-20. Where modulated light is to be translated into proportional electrical pulses, this condition is to be avoided. However, such a characteristic may be highly desirable in certain special applications.
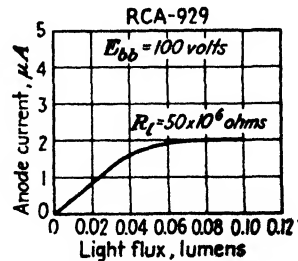


FIG. 15-20. The nonlinear dynamic curve resulting from operation at too low a supply voltage or too high a load resistance.

The analysis of a gas-phototube circuit is performed in exactly the same manner as above. Since the volt-ampere characteristics of such cells (see

Fig. 15-9) are no longer horizontal equidistant lines for equal intervals of light flux, the output will not, in general, be proportional to the light-flux variations, except for small variations. Furthermore, as discussed in Sec. 15-4, the graphical construction considered above is valid only at low frequencies, since the dynamic sensitivity of a gas cell decreases with the frequency.

**15-6. Photovoltaic Cells.** Photovoltaic cells depend for their operation on the fact that an emf is internally generated by the action of the radiant energy. This causes a current to flow in an external circuit *even though no external battery is applied.* Such cells may be electrolytic in nature,[14] consisting of two electrodes immersed in a suitable electrolyte, or they may be electronic in nature,[15] consisting of suitable materials in physical contact. It should be noted that these cells possess many features in common with the metallic disk rectifiers discussed in Sec. 7-7. There are two types of electronic photovoltaic cell that are commercially important, *viz.*, the iron-selenium Photronic cell,[16] and the copper–copper-oxide Photox cell.[15, 17]

In one type of copper–copper-oxide cell, the blocking layer, which is the point of origin of the photoelectrons, exists between the copper base plate and the cuprous oxide, precisely as in the rectifying disk. A semitransparent metallic layer is placed over the oxide layer. The copper base plate is one electrode, and a thin copper ring in contact with the semitransparent metallic layer is the other electrode. Under illumination the photoelectrons travel from the cuprous oxide to the copper, *i.e.*, from the semiconductor to the metal through the barrier layer. This is illustrated in Fig. 15-21*a*. These cells are called *back wall cells*, since the barrier layer is remote from the point of incidence of the light.

A *front wall cell* is also possible in which the barrier layer lies at the boundary between the top electrode and the cuprous oxide. In these cells the direction of current flow is also from the semiconducting cuprous oxide to the metal (the front electrode) through the barrier layer. In consequence of this, the direction of electron flow in the external circuit is opposite to that in the back wall cell, as illustrated in Fig. 15-21*b*. The sensitivity of the front wall cell is greater than that of the back wall cell because the light reaches the barrier layer without first being partially absorbed by the ineffective cuprous oxide. The spectral-response characteristics are also different, the peak of the front wall cell occurring at about 5,000 A, that of the back wall cell occurring at about 6,000 A.

The Photronic cell is a front wall cell and consists essentially of a base plate of iron on which is placed a thin layer of iron selenide, which is covered with a semitransparent layer of silver.[18] A copper ring in contact with the base plate acts as one electrode, and a copper ring in contact with the silver layer is the other electrode. The variation of the emf with the intensity of illumination is illustrated in Fig. 15-22. In this case the gen-
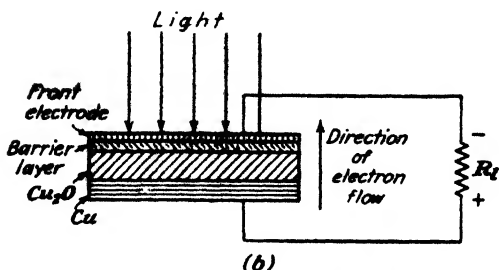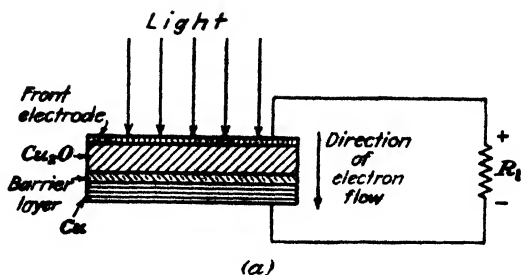
FIG. 15-21. Cuprous oxide "sandwich" photovoltaic cells of the (a) back wall and (b) front wall types.
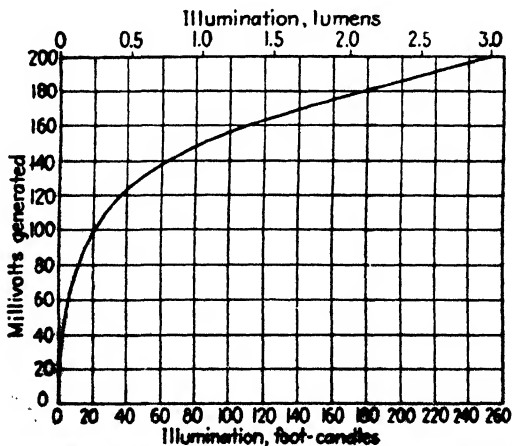


FIG. 15-22. Generated voltage in a Photronic cell as a function of light intensity. (*Courtesy of Weston Electrical Instrument Corp.*)

erated voltage may be related to the intensity of the incident light by an empirical expression of the form

$$E_g = kj^n \qquad (15\text{-}9)$$

where $n$ is a constant of the order of 0.4.

The maximum current that flows in the external circuit occurs under the short-circuit conditions. Experiment shows that the short-circuit current varies linearly with the light intensity, as indicated. Unlike the vacuum
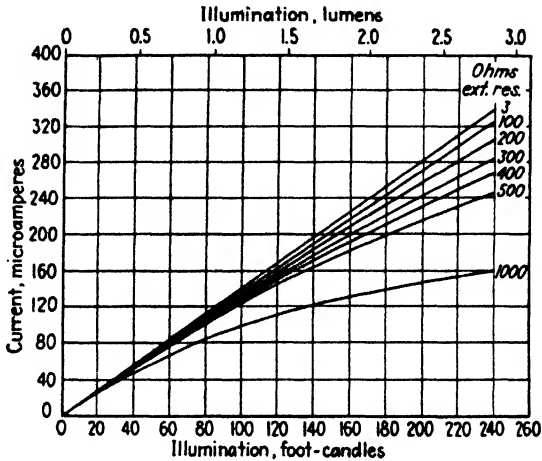


Fig. 15-23. The output current of a Photronic cell as a function of light intensity with load resistance as a parameter. (*Courtesy of Weston Electrical Instrument Corp.*)

phototubes, the output current is a function of the load resistor, and the linearity no longer applies for other than short-circuit conditions. This variation is shown in Fig. 15-23.

The response of these cells to modulated light is very bad. With equal off-on intervals of light at 60 per second, the off-response current is about 50 per cent of the on-response value, instead of the expected zero value. At 1,000 per second, the off-response is about 97 per cent of the on-response value. Because of this feature, these cells are used primarily in photometric measurements or with low-resistance relays for "on-off" operation. Since for linearity of response a low-resistance load must be used (so that the short-circuit current is obtained), it is difficult to amplify the output from such a unit.

**15-7. Photoconductive Cells.** Photoconductive cells are those for which the electrical resistance of the cell decreases with increases of illumination. Selenium is widely used for such cells. The electrical resistance of such a cell may decrease by a factor of 5 when placed in a strong light. The re-

sponse of such a cell to interrupted light is poor, there being an appreciable time lag between change in light and change in current. Since such cells have an appreciable "dark" current and a bad time lag, photoemissive tubes are much preferred. As a consequence, the photoconductive cell is of small commercial importance.

## PROBLEMS

**15-1.** Find the maximum speed with which the photoelectrons will be emitted (if at all) when radiation of wave length 5,893 A falls upon

*a.* A cesium surface, for which the work function is 1.8 volts.

*b.* A platinum surface, for which the work function is 6.0 volts.

*c.* Repeat parts *a* and *b* if the surfaces are illuminated with neon resonance radiation (743 A) instead of the yellow sodium line.

**15-2.** What is the minimum energy, expressed in joules and in electron volts, required to remove an electron from the surface of metallic potassium, the photoelectric threshold wave length of which is 5,500 A?

**15-3.** A cesium surface, for which the work function is 1.8 volts, is illuminated with argon resonance radiation (1,065 A). What retarding potential must be applied in order that the plate current in this photocell drop to zero? Assume that the contact potential is 0.50 volt, with the plate negative with respect to the cathode.

**15-4.** When a certain surface is irradiated by the 2,537-A mercury line, it is found that no current flows until at least 0.54 volt accelerating potential is applied. Assume that the contact potential is 1.00 volt, the cathode being positive with respect to the anode.

*a.* What is the work function of the surface?

*b.* What is the threshold wave length of the surface?

**15-5.** A certain photosurface has a spectral sensitivity of 6 ma/watt of incident radiation of wave length 2,537 A. How many electrons will be emitted photoelectrically by a pulse of radiation consisting of 10,000 photons of this wave length?

**15-6.** The photoelectric sensitivity of a PJ-22 photocell is 14 $\mu a$/lumen, when the anode potential is 90 volts. The window area of the photocell is 0.9 in.$^2$. A 100-watt electric-light bulb has a mean horizontal candle power of 120 cp. What will be the photocurrent if the cell is placed 3 ft from the lamp?

**15-7.** The energy distribution curve of a light source is known. The spectral sensitivity curves of several of the commercially available photosurfaces are supplied by the tube manufacturer and are shown in Fig. 15-13. Explain exactly how to determine which tube should be used with this particular light source in order to obtain the maximum photocurrent.

**15-8.** Devise a circuit for determining automatically the correct exposure time in the photographic printing process. Use a photocell, a relay, and any other auxiliary apparatus needed. The blackening of a photographic emulsion is determined by the product of the luminous intensity falling on the plate and the time of exposure. The instrument must trip the relay at the same value of this product regardless of what light source is used.

**15-9.** Plot curves of photocurrent vs. light intensity for a PJ-23 for load resistances of 1 and 10 megohms, respectively. The supply voltage is held constant at 80 volts.

**15-10.** *a.* The intensity of illumination on a 929 phototube is constant at 0.1 lumen. An adjustable voltage supply in series with a 25-megohm resistance is applied to the tube. Plot a curve of anode current vs. supply voltage. From this curve, determine

the wave shape of the photocurrent if the impressed voltage is sinusoidal and has a peak value of 250 volts.

*b.* Repeat part *a* if the load resistance is 50 megohms instead of 25 megohms.

*c.* Repeat part *a* if the impressed voltage consists of a 125-volt battery in series with an alternating voltage whose peak value is 125 volts.

**15-11.** In the circuit of Fig. 11-11 a 929 phototube is used in place of the resistor *R*. Explain clearly why this circuit will now give a linear sweep output.

Design a 1,000-cycle 50-volt sweep. The intensity of the illumination on the photocell is 0.1 lumen. Specify $E_{bb}$, $E_{cc}$, and *C*. Assume that an 885 thyratron is used for which the maintaining voltage is 15 volts and the critical grid starting characteristic is that given in Fig. 11-10.

**15-12.** In the circuit sketched the 929 phototube controls the current in the 884 thyratron. For what range of values of light flux will the thyratron be nonconducting? Explain your analysis clearly.
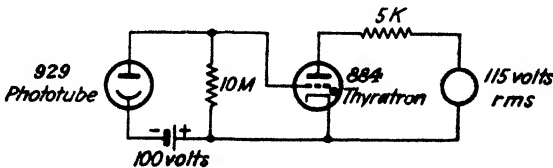


PROB. 15-12.

**15-13.** Plot a curve of the effective internal resistance vs. light intensity of the Photronic cell, the characteristics of which are given in the curves of Figs. 15-22 and 15-23.

## REFERENCES

1. HERTZ, H., *Ann. Physik*, **31**, 983, 1887.
2. HUGHES, A. L., and L. A. DuBRIDGE, more detailed information in "Photoelectric Phenomena," McGraw-Hill Book Company, Inc., New York, 1932.
3. RICHARDSON, O. W., and K. T. COMPTON, *Phil. Mag.*, **24**, 575, 1912.
4. LAWRENCE, E. O., and J. W. BEAMS, *Phys. Rev.*, **32**, 478, 1928.
5. PERKINS, T. B., *JOSA*, **29**, 226, 1939.
6. EINSTEIN, A., *Ann. Physik*, **17**, 132, 1905.
7. MILLIKAN, R. A., *Phys. Rev.*, **7**, 355, 1916.
8. FOWLER, R. H., *ibid.*, **38**, 45, 1931.
9. DuBRIDGE, L. A., and W. W. ROEHR, *ibid.*, **39**, 99, 1932.
   DuBRIDGE, L. A., *ibid.*, **39**, 108, 1932.
   APKER, L., E. TAFT, and J. DICKEY, *Phys. Rev.*, **73**, 46, 1948.
10. SKELLET, A. M., *J. Applied Phys.*, **9**, 631, 1938.
11. GLOVER, A. M., and R. B. JANES, *Electronics*, **13**, 26, August, 1940.
12. LARSEN, C. C., and H. SALINGER, *Rev. Sci. Instruments*, **11**, 226, 1940.
    RAJCHMAN, J. A., and R. L. SNYDER, *Electronics*, **13**, 20, December, 1940.
    WINANS, R. C., and J. R. PIERCE, *Rev. Sci. Instruments*, **12**, 269, 1941.
13. HENNEY, K., "Electron Tubes in Industry," 2d ed., McGraw-Hill Book Company, Inc., 1937.
14. BECQUEREL, E., *Compt. rend.*, **9**, 145, 561, 1839.
15. GRONDAHL, L. O., *Revs. Modern Phys.*, **5**, 141, 1933.
    GRONDAHL, L. O., and P. A. GEIGER, *Elec. Eng.*, **46**, 215, 1927.

Wilson, A. H., "Semi-conductors and Metals," Cambridge University Press, London, 1939.

16. Monographs B-8 and B-18-A, "The Photronic Photoelectric Cell," Weston Electrical Instrument Corp., Newark, N.J., 1935.

MacGregor-Morris, J. T., and R. M. Billington, *IEEJ*, **79**, 435, 1936.

17. Wilson, E. D., *Electronics*, **12**, 15, January, 1939.

18. Bergmann, L., *Physik. Z.*, **32**, 286, 1931.

### General References

Dow, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

Fielding, T. J.: "Photo-electric and Selenium Cells," Chapman & Hall, Ltd., London, 1935.

Henney, K.: "Electron Tubes in Industry," 2d ed., McGraw-Hill Book Company, Inc., 1937.

Hughes, A. L., and L. A. DuBridge: "Photoelectric Phenomena," McGraw-Hill Book Company, Inc., New York, 1932.

Walker, E. C.: "Photo-electric Cell Applications," Sir Isaac Pitman & Sons, Ltd., London, 1933.

Zworykin, V. K., and E. D. Wilson: "Photocells and Their Applications," John Wiley & Sons, Inc., New York, 1934.

Zworykin, V. K., and E. G. Ramberg: "Photoelectricity and Its Application," John Wiley & Sons, Inc., New York, 1949.

# CHAPTER 16

# TRIODE CHARACTERISTICS

$A$N EQUATION that relates the space-charge current in a diode with the potential existing between the anode or plate and the cathode has already been developed. This is the three-halves-power equation which is given by Eq. (7-16), *viz.*,

$$i_b = ke_b^{\frac{3}{2}} \tag{16-1}$$

where $e_b$ is the effective plate-cathode potential. If, therefore, the effective potential were changed by any means whatsoever, this change would be manifest as a change in plate current. In the case of the rectifier systems, it was found that the application of an a-c potential between the plate and cathode gives rise to a varying current in the output circuit. Evidently, the application of a constant voltage to the plate will give rise to a constant current in the output. Also, the application of a voltage of the form
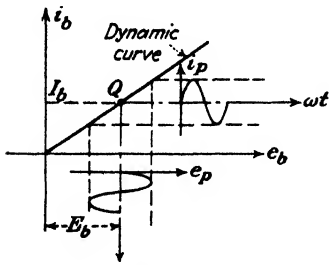
Fig. 16-1. The current output from a diode when the supply voltage consists of a d-c and an a-c voltage in series.

$$e_b = E_b + \sqrt{2}E \sin \omega t \tag{16-2}$$

to the diode will give rise to a varying current in the output, the form of which can be obtained graphically by the point-by-point method from the dynamic curve of the circuit. The results are illustrated in Fig. 16-1. If the method is not clear, refer to the discussion in connection with the construction of Fig. 12-4.

It is seen from this diagram that the current consists of a d-c component $I_b$ and an a-c variation $i_p$. It is noted that the effect of the voltage $E_b$ has been to shift the operating point from the origin to that marked $Q$ on the diagram. Evidently, with the presence of the large d-c term indicated, the tube no longer acts as a rectifier if the dynamic curve is a straight line.

**16-1. The Grid.** Suppose now that the mechanical structure of the tube is altered by inserting a metallic grid between the cathode and the plate.

This converts the tube into a triode. A schematic arrangement of the electrodes in a triode having cylindrical symmetry is given in Fig. 16-2. With a fixed d-c potential between the plate and the cathode and an a-c potential impressed between the grid and the cathode, the over-all result is to simulate the effects of a voltage of the form (16-2) applied to a diode. Under these circumstances, other conditions being comparable, the current that flows in the output or plate circuit is again of the form illustrated in Fig. 16-1. This action will be demonstrated quantitatively in the next chapter, when the use of a triode as a circuit element will be studied.
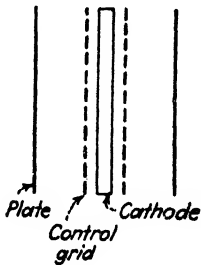


Fig. 16-2. Schematic arrangements of the electrodes in a triode. The tube has cylindrical symmetry.
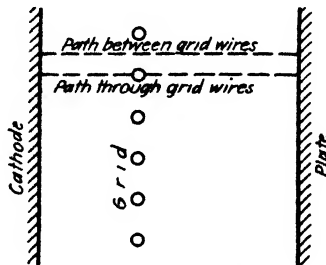
Fig. 16-3. Sketch of a plane-electrode triode, showing the paths for the potential profiles given in Fig. 16-5.

A study of the potential distribution in a triode is very instructive. For simplicity, consider a plane cathode and a parallel anode each of infinite extent. The grid is assumed to consist of parallel equidistant wires lying in a plane parallel to the cathode. The diameter of the wires is small compared with the distance between wires. Such an arrangement is shown in Fig. 16-3. If it is assumed that the cathode is so cold that it emits no electrons, then the potential at any point in the tube can be found by an electrostatic analysis. The results of such a calculation are shown in Fig. 16-4, where equipotential surfaces are indicated for various values of grid voltage. Since the electrodes are assumed to be of infinite extent, then it is only necessary to plot the equipotentials over a distance corresponding to the spacing between grid wires. Each picture is to be imagined repeated indefinitely to the right and left.

It should be noted, in particular, that the grid structure does not produce an equipotential plane at the position of the grid. If it did, there could never be plate current for *any* value of negative grid voltage because the electrons would find themselves in a retarding field as soon as they left the cathode. (This assumes, for the moment, that the cathode is heated but that the electrons leave with zero initial velocity.) Because of the influ-
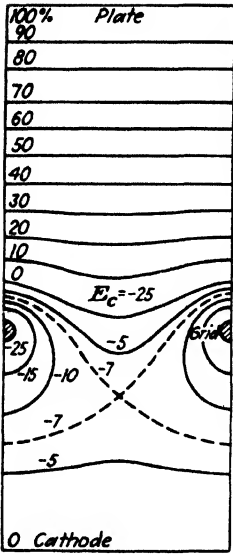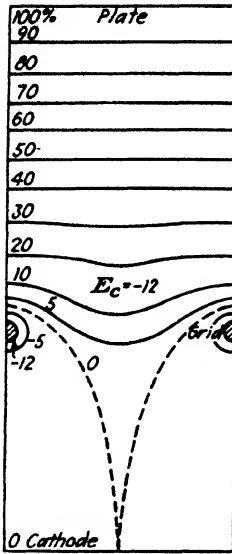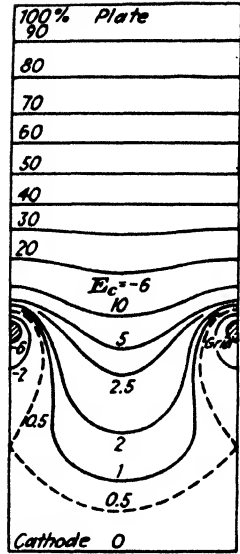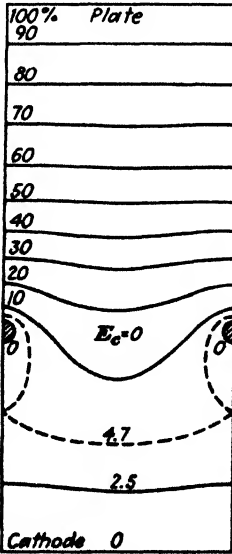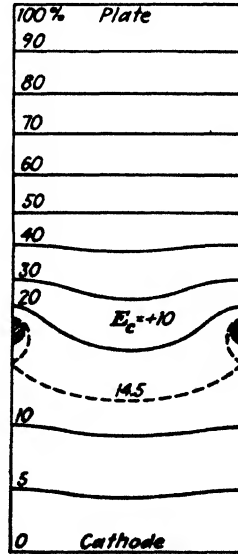
Fig. 16-4.  Equipotential contours in the plane-electrode triode: (a) grid beyond cutoff potential; (b) grid at cutoff potential; (c) grid negative at half cutoff value; (d) grid at zero potential; (e) grid positive.  (From K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, Inc.)

ence of the positive plate potential, it is possible for an electron to find a path between grid wires such that it does not collide with a potential energy barrier (provided that the grid is not too highly negative). Thus, the potential variation between cathode and anode depends upon the path. The potential vs. distance curves (called "profile presentations") corresponding to Fig. 16-4 are given in Fig. 16-5 for the two extreme conditions, a path midway between grid wires (upper curve) and a path directly through the grid wires (lower curve).

If an electron finds itself in a retarding field regardless of what part of the cathode it comes from, then it certainly cannot reach the anode. This is the situation pictured in (a) of Figs. 16-4 and 16-5 and corresponds to conditions beyond cutoff. In (b) are shown the conditions just at cutoff, where the electric-field intensity at the cathode is everywhere zero. Actually, cutoff is obtained at a grid voltage slightly less than this value so that the field at the cathode is somewhat negative and hence repels all the emitted electrons. It should be clear from a study of these figures that the current distribution is not constant along paths at different distances from the grid wires. If the grid is made sufficiently negative, then cutoff will occur throughout the entire region. This is the situation for all grid voltages more negative than that indicated in (b). If the grid voltage is made more positive than this cutoff value, then as shown in (c), current will flow only in the region midway between the grid wires, because any electrons starting out toward a grid will be repelled. This is the situation with the usual operating voltages in a triode voltage amplifier. In (d) the grid is at cathode potential, and in (e) it is held positive with respect to the cathode. Under these conditions, electrons can reach the anode by all paths unless they collide with a grid wire and are collected as grid current.

It should be emphasized that these diagrams represent space-charge-free conditions. In Chap. 7 it was shown that under space-charge conditions, the electric-field intensity at the cathode is reduced to zero. Hence, for a hot cathode, the diagrams (c), (d), and (e) must be modified (lowered) somewhat and must have zero slope at the cathode.

**16-2. The Electrode Current.** From the qualitative discussion already given, it follows that the current should depend upon the space-charge-free cathode field intensity. This, in turn, is a linear function of the grid and plate potentials. Since the grid is much closer to the cathode than the plate, a given change in potential of the grid has a much greater effect on the field intensity at the cathode than does the same change in potential of the anode. For example, if the plate voltage is changed slightly in Fig. 16-5, then it will affect the slope of the potential curve at the cathode very little. If the grid voltage is altered the same amount, the slope will change by a very much larger amount. In view of this discussion and the known three-halves-power law for diodes, it is anticipated that the plate
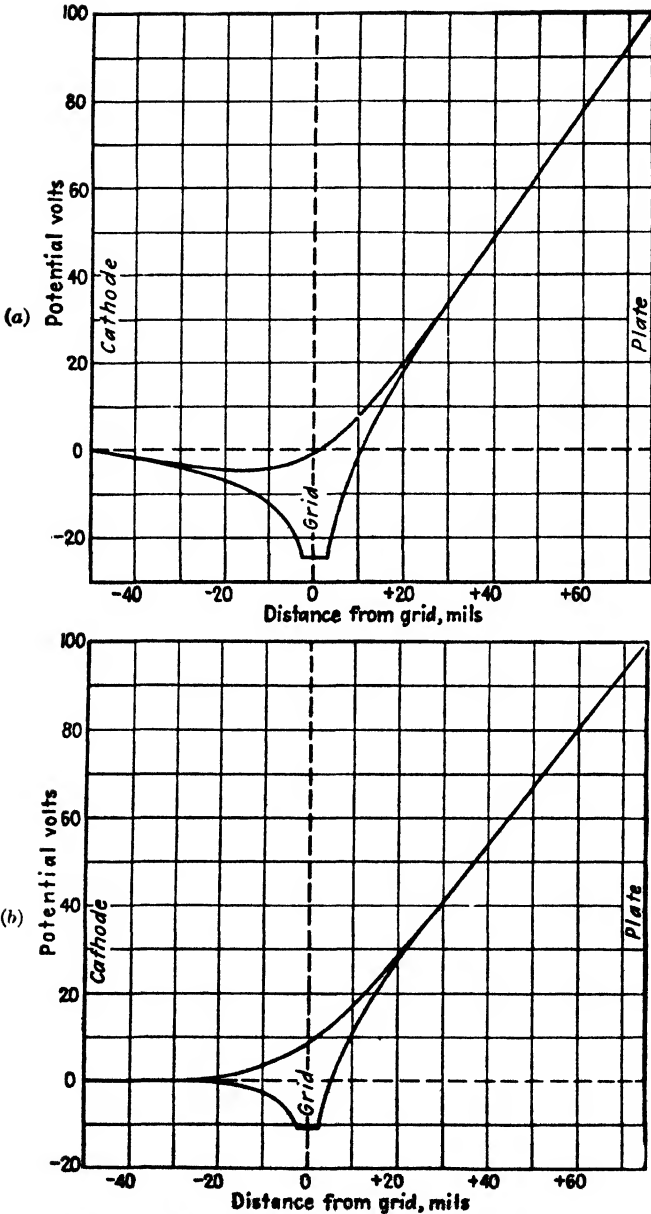
FIG. 16-5. Potential profiles of a plane-electrode triode: (a) with grid at twice the cutoff value of potential; (b) with grid at the cutoff value of potential. (*From K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, Inc.*)
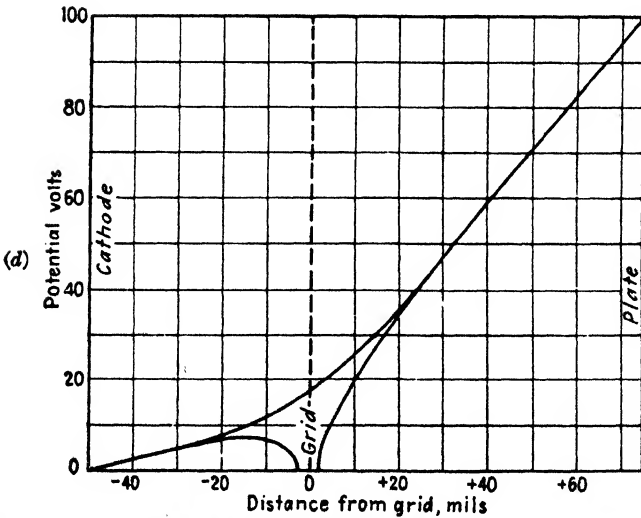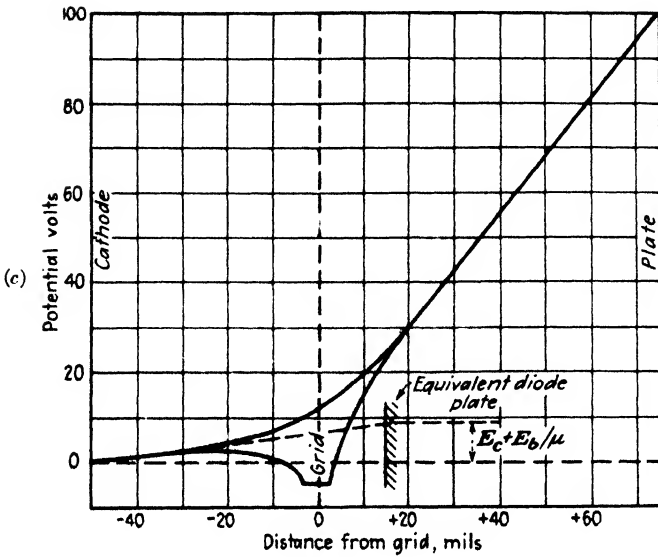
FIG. 16-5.  Potential profiles in a plane-electrode triode: (c) with grid negative at half the cutoff value of potential; (d) with grid at zero potential.  (*From K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, Inc.*)
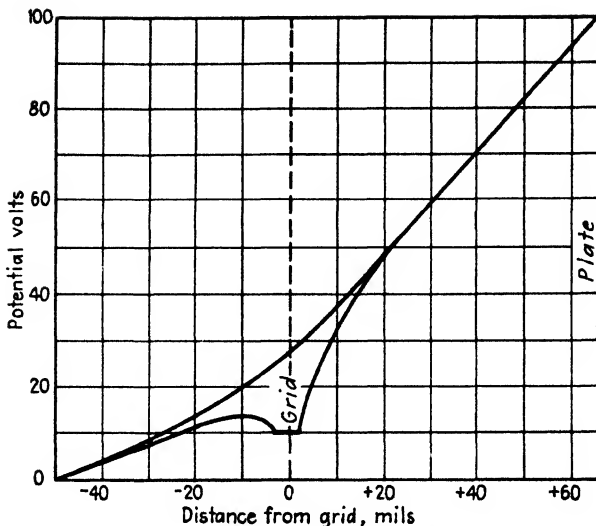
Fig. 16-5e. Potential profiles in a plane-electrode triode, with grid at a positive potential.

current may be represented approximately by the equation

$$i_b = k\left(e_c + \frac{e_b}{\mu}\right)^n \qquad (16\text{-}3)$$

where $e_b$ is the plate potential, $e_c$ is the grid potential, and where the factor $\mu$ is a measure of the relative effectiveness of the grid and plate potentials. The parameter $\mu$ is known as the "amplification factor" and is substantially constant and independent of current. The exponent $n$ is approximately equal to $\frac{3}{2}$. The constant $k$ is called the "permeance."

The validity of Eq. (16-3) has been verified experimentally for many triodes. No rigorous theoretical derivation of this equation exists, even for a triode of relatively simple geometry. In particular, the exact dependence of the constant $k$ upon the dimensions of the tube is not known.[1] However, the value of the amplification factor $\mu$ can be calculated with a fair degree of accuracy from equations that are based on certain electrostatic considerations.[2] Semiempirical relationships that are suitable for design calculations are available.[3]

Deviations from Eq. (16-3) exist for the same reasons that deviations from the three-halves-power law exist for diodes, *viz.*, the effects of contact potential, initial velocities, nonuniform structure, etc. (see Sec. 7-4). Fortunately, it is seldom necessary to know the exact mathematical relationship that exists among $i_b$, $e_b$, and $e_c$. Hence the principal emphasis in what

follows will be on the experimentally determined relationships among these variables.

If the grid potential be made positive, the electron stream will increase because of the combined action of both the grid and the plate. With a positive potential on the grid, some of the electrons will be attracted to it and a current into the grid will result. Unless the grid is designed to dissipate the power represented by the current to it, when maintained at a positive potential, the grid structure may be seriously damaged. The grid is generally maintained negative, although positive-grid triodes for power-amplifier applications are available.

The variations of the plate and grid currents with variations of grid voltage are illustrated in Fig. 16-6. In this diagram, the plate potential is assumed to be constant. It is seen that cutoff occurs if the grid potential is sufficiently negative. As the grid potential is made less negative, the plate current follows a smooth curve, this variation being expressed analytically with fair approximation by Eq. (16-3). As the grid potential is made positive with respect to the cathode, a grid current is obtained. This current increases rapidly with increasing potentials.
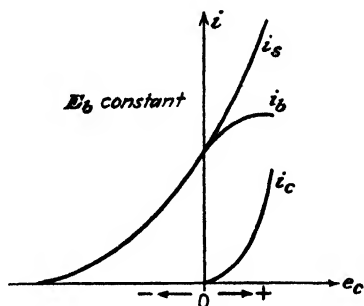


Fig. 16-6.   Plate, grid, and total space current in a triode as a function of grid voltage for a fixed plate voltage.

Once the grid potential is made positive so that a grid current $i_c$ exists, Eq. (16-3) is no longer applicable, since only a portion of the total space-charge current is collected by the anode. Equation (16-3) does give a good representation of the total space current $i_s = i_b + i_c$. Hence, as the grid potential is made increasingly positive, the plate current increases rapidly, but not according to Eq. (16-3). The plate current will ultimately fall, as shown, since, for a constant available space current and an increasing grid current, the current to the plate must decrease. Because the electrons are emitted with finite initial velocities, some of these will be collected by the grid even when its potential is slightly negative. This grid current may be of the order of a few microamperes when the grid voltage is, say, $\frac{1}{2}$ volt negative.

**16-3. Triode Characteristics.**   The plate current depends upon the plate potential and upon the grid potential and may be expressed mathematically by the functional relationship

$$i_b = f(e_b, e_c) \tag{16-4}$$

This is read "$i_b$ is some function $f$ of $e_b$ and $e_c$." This relationship is some-

times written as $i_b = i_b(e_b, e_c)$, the quantities in the parentheses designating the variables upon which the function $f$ (or $i_b$) depends. If it is assumed that the grid current is zero, then the approximate explicit form of this function is that expressed by Eq. (16-3). Of course, the plate current also depends upon the heater current; but as this is usually maintained constant at the rated value, and this is such as to provide perhaps five to ten



(a)-Plate characteristics     (b)-Transfer characteristics
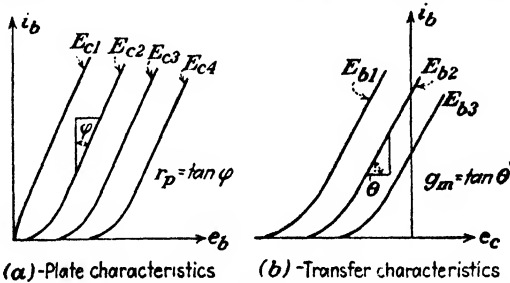
Fig. 16-7.  (a) Plate and (b) transfer characteristic curves of a triode.
$$E_{c1} > E_{c2} > E_{c3}.  \quad E_{b1} > E_{b2} > E_{b3}.$$

times the normal required space-charge current, this term does not enter into the functional relationship.

The important variables that give a complete description of the characteristics of the triode are, as already noted, $i_b$, $e_b$, and $e_c$. By plotting these three variables on a three-dimensional system of axes, a space diagram is obtained. The traces of these surfaces on the three coordinate
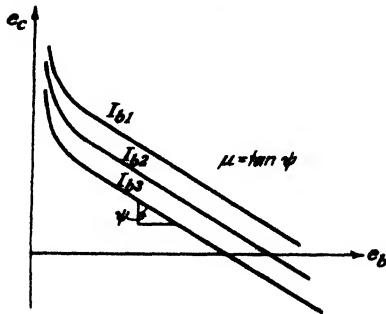


Fig. 16-8.  Constant-current characteristic curves of a triode.
$$I_{b1} > I_{b2} > I_{b3}.$$

planes (and on planes parallel to these) give three families of characteristic curves. These curves, easy to visualize, are given in Figs. 16-7 and 16-8.

Figure 16-7a shows a family of characteristic curves known as the "plate characteristics," since they give the variation of the plate current with the plate potential for various values of grid potential. The only effect of

making the grid more negative is to shift the curves to the right without changing the slopes appreciably. If the grid potential is made the independent variable, the family of curves known as the "mutual" or "transfer" characteristics, illustrated in Fig. 16-7b, is obtained. The effect of making the plate potential less positive is to shift the curves to the right, the slopes again remaining substantially unchanged. These conditions are readily evident if it is remembered that the sets of curves in these
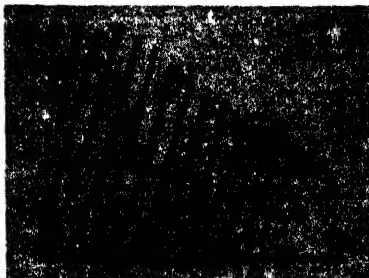


FIG. 16-9. Oscillogram of the plate characteristics of a 6C5 triode. The grid-bias values vary from +4 to −22 volts in 2-volt intervals.

FIG. 16-10. Oscillogram of the transfer characteristics of a 6C5 triode for plate voltages of 100, 200, 300, and 400 volts.

diagrams are plots of Eq. (16-3) with either $e_c$ or $e_b$ maintained constant as a parameter.

The simultaneous variation of both the plate and the grid potentials so that the plate current remains constant gives rise to a third group of characteristics, known as the "constant-current" characteristics, illustrated in Fig. 16-8. These show the effects of the plate and grid voltages on the plate current of the tube.

The most important family of characteristics is the plate family, and these are supplied in convenient form in data books supplied by the tube manufacturers. The plate characteristics for several representative tubes are reproduced in Appendix IX. It will be found in the following chapters that all operating characteristics may be deduced from this one set of curves. An oscillogram of the plate characteristics of a type 6C5 tube is given in Fig. 16-9. The transfer characteristics of this tube are shown in the oscillogram of Fig. 16-10.

**16-4. Triode Parameters.** In the analysis of networks using triodes as circuit elements (Chap. 17), it is found necessary to make use of the slopes of the characteristic curves of Figs. 16-7 and 16-8. Hence it is convenient to introduce special symbols and names for these quantities. This will now be done.

The slope of the constant-current characteristic gives the amplification factor $\mu$ introduced in Eq. (16-3). That is, the amplification factor is defined as the ratio of the change in plate voltage to the change in grid voltage for a constant plate current. Mathematically, $\mu$ is given by the relation

$$\mu = -\left(\frac{\partial e_b}{\partial e_c}\right)_{I_b} \tag{16-5}$$

The subscript $I_b$ denotes that the plate current remains constant in performing the indicated partial differentiation. In order that $\mu$ be a positive number, the minus sign is necessary because an increasing plate voltage will require a decreasing grid potential, if the current is to remain unchanged. The reciprocal of the amplification factor is called the "durchgriff" or the "penetration factor."

The quantity $(\partial e_b/\partial i_b)_{E_c}$, which expresses the ratio of an increment of plate potential to the corresponding increment of plate current when the grid potential is kept constant, has units of resistance and is known as the "plate resistance" of the tube. It is designated by the symbol $r_p$. It is noted that the plate resistance is the reciprocal of the slope of the plate characteristics of Fig. 16-7a. It should be recalled that the plate resistance of a diode was defined in a similar manner. The reciprocal of the plate resistance is called the "plate conductance," $g_p \equiv 1/r_p$.

The quantity $(\partial i_b/\partial e_c)_{E_b}$, which gives the ratio of an increment of plate current to the corresponding increment in grid potential for constant plate potential, has the units of conductance. This quantity is known as the "plate-grid transconductance" and represents the change of current in the plate circuit owing to the change in potential of the grid. The plate-grid transconductance is frequently referred to simply as the "mutual conductance" and is designated by the symbol $g_m$. The quantity $g_m$ is the slope of the mutual characteristic curves of Fig. 16-7b.

To summarize: The triode coefficients, parameters, or "constants," which are characteristic of the tube, are

$$\left.\begin{aligned}\left(\frac{\partial e_b}{\partial i_b}\right)_{E_c} &\equiv r_p, \text{ plate resistance}\\[2mm]\left(\frac{\partial i_b}{\partial e_c}\right)_{E_b} &\equiv g_m, \text{ mutual conductance}\\[2mm]-\left(\frac{\partial e_b}{\partial e_c}\right)_{I_b} &\equiv \mu, \text{ amplification factor}\end{aligned}\right\} \tag{16-6}$$

Since there is only one equation (16-4) relating the three quantities $i_b$, $e_b$, and $e_c$, the three partial derivatives cannot be independent. The interre-

lationship may be shown to be (see page 505)

$$\mu = r_p g_m \qquad (16\text{-}7)$$

The variations of these parameters for a particular value of plate potential for the 6C5 tube are shown in Fig. 16-11. The variations for other values of plate potential may be some-what different, although the general trends will be the same. It is noticed that the plate resistance varies over rather wide limits. It is very high at zero plate current and varies approximately inversely as the one-third power of the plate current (see Prob. 16-2). The transconductance increases with plate current from zero at zero plate current and varies directly as the one-third power of the plate current. The amplification factor is observed to remain reasonably constant over a wide range of currents, although it falls off rapidly at the low currents.
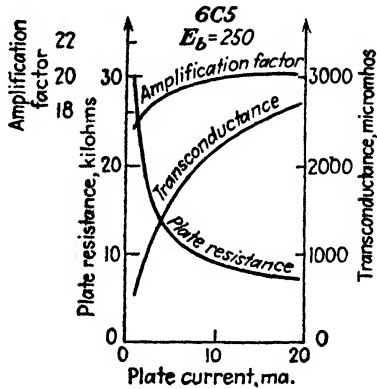


Fig. 16-11. The parameters $\mu$, $r_p$, and $g_m$ of a 6C5 triode as a function of plate current. (*Courtesy of Raytheon Production Corp.*)

The usual order of magnitude of the tube parameters for conventional triodes is approximately as follows:

$\mu$: from 2.5 to 100.
$r_p$: from 500 to 100,000 ohms.
$g_m$: from 500 to 6,000 micromhos.

Among the most commonly used triodes are the general-purpose twin triodes designated by the manufacturers as 6SL7 and 6SN7. These contain two triode units in one envelope, and each section has the following parameters at the recommended operating point:

|      | $\mu$ | $r_p$, ohms | $g_m$, micromhos |
|------|------|--------|-----------|
| 6SL7 | 70   | 44,000 | 1,600     |
| 6SN7 | 20   | 7,700  | 2,600     |

The 6J5 is a single triode whose characteristics are identical with each section of the 6SN7.

*Example.* Find approximate values of $r_p$, $\mu$, and $g_m$ directly from the plate characteristics of a 6SN7 at the operating point $E_b = 200$ and $E_c = -6$.

*Solution.* The plate curves are given in Appendix IX and are reproduced for convenience in Fig. 16-12. The operating point is designated $A$ and is seen to correspond to 7.6 ma. To find the tube parameters, the definitions summarized in Eq. (16-6) are used.
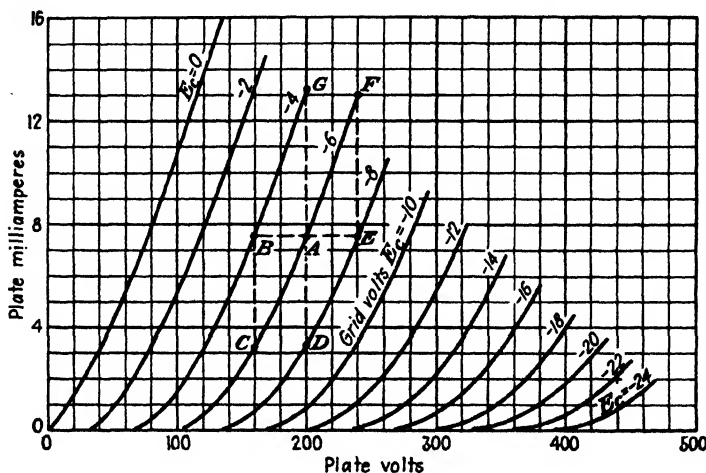


FIG. 16-12. The plate characteristics of a 6SN7 dual triode.

Approximate values of the partial derivatives are obtained by taking finite differences instead of differentials. Thus,

$$r_p = \left(\frac{\text{change in plate voltage}}{\text{change in plate current}}\right)_{\text{grid voltage const}}$$

$$= \left(\frac{\Delta e_b}{\Delta i_b}\right)_{E_c} = \frac{AB}{BC}$$

The point $B$ is characterized by $E_b = 160$, $I_b = 7.6$, and $E_c = -4$. The point $C$ is specified by $E_b = 160$, $I_b = 3.2$, and $E_c = -6$. Hence $AB = 200 - 160 = 40$ volts, and $BC = 7.6 - 3.2 = 4.4$ ma.

$$r_p = \frac{40}{4.4 \times 10^{-3}} = 9,100 \text{ ohms}$$

Similarly,

$$g_m = \left(\frac{\text{change in plate current}}{\text{change in grid voltage}}\right)_{\text{plate voltage const}}$$

$$= \left(\frac{\Delta i_b}{\Delta e_c}\right)_{E_b} = \frac{AD}{\Delta e_c}$$

The point $D$ is specified by $E_b = 200$, $I_b = 3.4$, and $E_c = -8$. Hence $AD = 7.6 - 3.4 = 4.2$ ma, and $\Delta e_c = -6 - (-8) = 2$ volts.

$$g_m = \frac{4.2 \times 10^{-3}}{2} = 2.1 \times 10^{-3} \text{ mho} = 2,100 \text{ micromhos}$$

Similarly,

$$\mu = -\left(\frac{\text{change in plate voltage}}{\text{change in grid voltage}}\right)_{\text{plate current const}}$$

$$= -\frac{AB}{\Delta e_c} = -\frac{40}{-6-(-4)} = \frac{40}{2} = 20$$

The product of $r_p$ and $g_m$ is $r_p g_m = (9,100)(2.1 \times 10^{-3}) = 19$, which checks well with $\mu = 20$, considering the large increments in voltage and current that were taken, whereas, theoretically, infinitesimal changes are indicated.

These parameters were evaluated at the point $A$ for decreasing currents and voltages. If positive increments are taken, a new set of values is obtained. Thus,

$$r_p = \frac{EA}{FE}; \qquad g_m = \frac{GA}{\Delta e_c} \qquad \text{and} \qquad \mu = -\frac{EA}{\Delta e_c}$$

For point $E$, $E_b = 240$, $I_b = 7.6$, and $E_c = -8$
For point $F$, $E_b = 240$, $I_b = 13.0$, and $E_c = -6$
For point $G$, $E_b = 200$, $I_b = 13.2$, and $E_c = -4$

Hence

$$r_p = \frac{240-200}{13.0-7.6} = \frac{40}{5.4 \times 10^{-3}} = 7,400 \text{ ohms}$$

$$g_m = \frac{(13.2-7.6) \times 10^{-3}}{-4-(-6)} = \frac{5.6 \times 10^{-3}}{2} = 2,800 \text{ micromhos}$$

$$\mu = -\frac{240-200}{-8-(-6)} = \frac{40}{2} = 20$$

$$r_p g_m = (7,400)(2.8 \times 10^{-3}) = 20.8$$

which checks well with $\mu = 20$.

It should be noted that, for increasing currents, the plate resistance decreased, the transconductance increased, and the amplification remained essentially constant (actually increased slightly). This is consistent with Fig. 16-11.

If $r_p$ were constant, then the slope of the plate characteristics would everywhere be constant. In other words, these curves would be parallel lines. If $\mu$ were constant, then the horizontal spacing of the lines ($AB$ or $EA$ in the above example) would be constant. This assumes that the characteristics are drawn with equal increments in grid voltage (as they always are). If $r_p$ and $\mu$ are constant, so also is $g_m = \mu/r_p$. Hence, an important conclusion can be drawn: *If over a portion of the $i_b$-$e_b$ plane, the characteristics can be approximated by parallel lines which are equidistant for equal increments in grid voltage, then the parameters $\mu$, $r_p$, and $g_m$ can be considered constant over this region.* It is shown in the next chapter that if the tube operates under this condition (tube parameters sensibly constant) then the analysis of the behavior of the tube as a circuit element is quite simple.

**16-5. Power Triodes.** High-power triodes are extensively used in radio transmitters. The grids of such tubes have very large signals impressed

on them.   During one part of the grid driving cycle, an appreciable grid
current flows.   During a second part of the grid driving cycle, the grid
causes plate-current cutoff to occur.   Because of this class of operation,
the grid structures have been designed to withstand the resulting heating



Fig. 16-13.   Photograph of two transmitting tubes.   (*Courtesy of RCA Manufacturing Co.*)

of the grid without damage to the tube.   The analysis of the operation of
triodes in these applications is beyond the scope of this book, but it is of
interest to note that such an analysis may be accomplished with either the
plate characteristics or the constant-current characteristics.

Photographs of two power triodes, one designed for water cooling of the
plate, the other for air cooling of the plate, are given in Fig. 16-13.   The
892 tube measures approximately 20 in. long and has a maximum allow-
able plate dissipation of 7.5 kw.   The 893A-R measures approximately

27 in. long and has a maximum allowable plate dissipation of 20 kw. The power output of these tubes depends on the character of the operation, but in a common application, the corresponding output powers are 6 kw and 18 kw, respectively.

**16-6. Triodes at Ultrahigh Frequencies.** The curve of Fig. 16-6, which is representative of triodes, shows that the grid current is appreciable only when the grid is positive with respect to the cathode. For the usual negative-grid triode, the grid current seldom exceeds $10^{-8}$ amp. If the electrodes, including the heater, are operated at reduced potential, the grid current may be as little as $10^{-12}$ amp. In fact, special electrometer tubes are available which have grid currents as low as $10^{-15}$ amp. It would appear from this that the effective grid loading that results when a tube is connected in a circuit is very low. Actually, as will be discussed below, the grid loading increases with frequency of the applied signal, and the usual triode is limited in applications to frequencies of 200 megacycles or less.

To understand the physics of grid loading, consider an electron that has left the cathode and is approaching the grid on its way to the anode. Assume also that the grid potential is negative with respect to cathode, so that no electrons will be collected by the grid. As the electron approaches the grid, a changing image charge density will be induced on the grid (see Sec. 4-7 for more details). This changing charge on the grid represents an instantaneous grid current, and the current is in such a direction as to charge the grid bias battery. The power for this charging process is supplied by the approaching electron, and it is decelerated as a result. Once the electron passes beyond the grid, the process is reversed and the electron will take energy from the grid, and it is repelled thereby. The amount of energy lost by the electron as it approaches the grid is just equal to that which it gains as it recedes from the grid, and the net energy is zero. Under these conditions the net grid loading is zero.

The situation of negligible grid loading will exist over a wide range of frequencies of voltage applied to the grid. This follows from considerations of Chap. 2, where it was shown that the velocity acquired by an electron when it is accelerated by even a few volts is quite high, with a consequent small transit time in passing from cathode to plate through the interelectrode space. If the transit time is very small compared with the period of the voltage applied to the grid, then in so far as any electron is concerned, the situation is just as though a d-c potential existed on the grid.

At those frequencies of applied voltage to the grid when the transit time of the electron is no longer small compared with the period of the applied voltage, the grid loading is no longer negligible.[4] Now it is possible for the energy that is supplied to the grid by the electron on its passage from cathode toward the grid to exceed the amount of energy that is returned

by the grid, as the electron continues from the grid to the anode, with a resultant net energy loss in the grid circuit. Since this energy must be supplied by the grid driving source, it represents a loading on this driving source. This loading increases with increased grid driving frequency and ultimately reaches the point where the tube can no longer be used.

The foregoing discussion may be expressed in terms of a circuit point of view, by noting that at the lower frequencies, the current that is induced in the grid by the moving electron is 90 deg out of phase with the grid voltage, with a consequent zero net power loss. At the higher frequencies a net inphase component exists, and this represents a resistive component of input impedance. This effective input resistance acts as a shunt on the high-impedance signal applied to the grid.
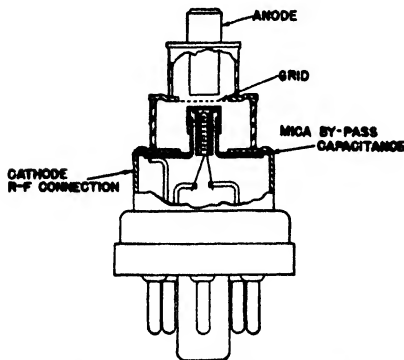


FIG. 16-14. The construction of a light-house or disk-seal triode. (*From K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, Inc.*)

In addition to the electron-transit-time grid loading just discussed, there is reactive loading of the input because of interelectrode capacitances and lead inductances. Although these values are small (of the order of a few micromicrofarads and 0.01 microhenry, respectively), yet at frequencies above 100 megacycles they become important. Thus, there may be an appreciable loss in radio-frequency voltage across the lead inductance. Series resonance may occur between the inductance and the capacitance at some high frequency, and it now will act as a short circuit on the input. Even if the inductances are neglected, the capacitances will be small shunting reactances which load the input and output circuits.

Because of these considerations, tubes for high-frequency operation are made in a manner to reduce transit time, interelectrode capacitances, and lead inductances. This means very close spacing of electrodes (sometimes only a few thousandths of an inch between cathode and grid). It also requires that the physical dimensions of the electrodes be small. Among these high-frequency tubes are the so-called "acorn," "doorknob," and "disk-seal" or "lighthouse" tubes with upper-frequency limits of about 2,000, 1,700, and 3,500 megacycles, respectively. The names of these tubes are descriptive of the physical shape of their envelopes. The first two are built like conventional cylindrical triodes, except for the close spacing, small electrode size, and the fact that the leads are brought out of the envelope at widely spaced points so as to cut down capacitances.

The lighthouse tube [5] has essentially plane-parallel electrode construction but with the connections to the electrodes brought out in the form of disks. When these tubes are inserted into a coaxial-line resonator, lead inductance is unimportant because the electrodes are integral parts of the resonant cavities. Radiation losses are also eliminated by this construction and use. The construction of a lighthouse tube is shown in Fig. 16-14.

## PROBLEMS

**16-1.** *a.* From the plate characteristics of the 6C5 triode (see Appendix IX), obtain the mutual characteristics for $E_b = 80, 160, 240,$ and 320 volts.

*b.* Obtain the constant-current characteristics of the 6C5. Plot $e_b$ vs. $e_c$ for $I_b = 4, 8, 12,$ and 16 ma.

*c.* Repeat part *a* for the 6SF5 high-$\mu$ triode.

*d.* Obtain the constant-current characteristics for the 6SF5 high-$\mu$ triode. Plot $e_b$ vs. $e_c$ for $I_b = 0.2, 0.6, 1.0,$ and 1.4 ma.

**16-2.** If the plate current in a triode can be represented by Eq. (16-3), show that $r_p$ is proportional to $i_b^{-m}$ and that $g_m$ is proportional to $i_b^m$, where $m = (n-1)/n$. Note that if the three-halves-power law is valid, then $m = \frac{1}{3}$.

**16-3.** The plate resistance of a 6J5 triode or one unit of a 6SN7 is 7,700 ohms, and the transconductance is 2,600 micromhos.

*a.* If the plate voltage is increased by 50 volts, what is the increase in plate current? The grid voltage is maintained constant.

*b.* What change in grid voltage will bring the plate current back to its former value? The plate voltage is maintained at the value to which it was raised in part *a*.

**16-4.** The plate resistance of a 6C5 triode is 10,000 ohms, and the amplification factor is 20. The tube is operated at the quiescent point $E_b = 250$ volts, $E_c = -8$ volts, and $I_b = 8$ ma.

*a.* To what value must the grid voltage be changed if the plate current is to change to 12 ma? The plate voltage is maintained at 250 volts.

*b.* To what value must the plate voltage be changed if the plate current is to be brought back to its previous value? The grid voltage is maintained constant at the value found in part *a*.

Compare the values calculated above with those which can be read directly from the plate characteristics of the 6C5 (see Appendix IX).

**16-5.** The plate current of a 6F6 tube connected as a triode can be expressed approximately by the equation $i_b = 41(e_b + 7e_c)^{1.41} \times 10^{-6}$ amp. The tube is operated with a grid bias of $-20$ volts, and the plate voltage is 250 volts. Calculate the following:

*a.* The tube current.

*b.* The plate resistance.

*c.* The transconductance.

**16-6.** *a.* Show that the volt-ampere curve for $E_c = -8$ volts of a type 6C5 triode (see Appendix IX) can be expressed by an equation of the form

$$i_b = k(e_b + 20e_c)^n$$

Determine $k$ and $n$ from a logarithmic plot.

Plot the curve expressed by this equation and the experimental curve on the same sheet of paper, and compare.

*b.* Calculate the values of $r_p$ and $g_m$ from this equation at the conditions $E_c = -8$ volts and $E_b = 250$ volts, and compare with the values given in Fig. 16-11.

**16-7.** *a.* Calculate $\mu$, $r_p$, and $g_m$ from the plate characteristics of the 6C5 tube (see Appendix IX) at the quiescent point $E_b = 250$ volts, $E_c = -8$ volts.

*b.* Plot $\mu$, $r_p$, and $g_m$ for a 6C5 as a function of $I_b$, with $E_b = 250$ volts.

*c.* Plot $\mu$, $r_p$, and $g_m$ for a 6C5 as a function of $E_c$, with $E_b = 250$ volts.

**16-8.** Given the transfer characteristics of a triode. Explain clearly how to determine $r_p$, $\mu$, and $g_m$ at a specified quiescent point.

**16-9.** Show that if the triode plate characteristic can be approximated by straight lines the equation of these lines is

$$i_b = \frac{1}{r_p}(\mu e_c + e_b - \epsilon)$$

where $\epsilon$ is the voltage intercept at zero current of the $e_c = 0$ curve.

**16-10.** *a.* Starting with the definitions of $g_m$ and $r_p$, show that if two identical tubes are connected in parallel $g_m$ is doubled and $r_p$ is halved. Since $\mu = r_p g_m$, then $\mu$ remains unchanged.

*b.* If the two tubes are not identical show that

$$g_m = g_{m1} + g_{m2}$$

that

$$\frac{1}{r_p} = \frac{1}{r_{p1}} + \frac{1}{r_{p2}}$$

and that

$$\mu = \frac{\mu_1 r_{p2} + \mu_2 r_{p1}}{r_{p1} + r_{p2}}$$

**16-11.** *a.* The circuit shown is used to measure grid current. Prove that the grid current $I_c$ with the switch $S$ open is given by

$$I_c = \frac{\Delta I_b}{g_m R_g}$$

where $\Delta I_b$ is the change in plate current as the switch is closed and $g_m$ is the transconductance of the tube.

*b.* The foregoing method depends upon a knowledge of $g_m$. This limitation is removed as follows: The plate current $I_b$ is noted with $S$ open. The switch is then closed,
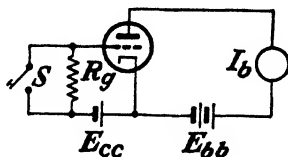


PROB. 16-11.

and the bias is adjusted until the plate current is its previous value. Explain how this procedure allows the determination of $I_c$.

**16-12.** When the grid is insulated from the cathode and the plate of a vacuum triode, it is said to be "floating." Describe an experimental method of determining this floating potential.

### REFERENCES

1. Dow, W. G., *Proc. IRE*, **28**, 548, 1940.

FREMLIN, J. H., *Phil. Mag.*, **27**, 709, 1939.

RODDA, S., *ibid.*, **29**, 601, 1940.

SPANGENBERG, K. R., "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.

2. OLLENDORF, F., *Elektrotech. u. Maschinenbau*, **52**, 585, 1934.

Dow, W. G., "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

SPANGENBERG, *op. cit.*

MILLER, J. M., *Proc. IRE*, **8**, 64, 1920.

VOGDES, F. B., and F. R. ELDER, *Phys. Rev.*, **24**, 683, 1924; **25**, 255, 1925.

KING, R. W., *ibid.*, **15**, 256, 1920.

3. KUSUNOSE, Y., *Proc. IRE*, **17**, 1706, 1929.

4. LLEWELLYN, F. B., "Electron Inertia Effects," Cambridge Physical Tract, Cambridge University Press, London, 1941.

BRONWELL, A. B., and R. E. BEAM, "Theory and Application of Microwaves," McGraw-Hill Book Company, Inc., New York, 1947.

SPANGENBERG, *op. cit.*

5. MCARTHUR, E. D., *Electronics*, **18**, 98, 1945.

HAMILTON, D., J. K. KNIPP, and J. B. H. KUPER, "Klystrons and Microwave Triodes," Radiation Laboratory Series, Vol. VII, McGraw-Hill Book Company, Inc., New York, 1948.

### General References

CHAFFEE, E. L.: "Theory of Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1933.

Dow, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.

EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," 3d ed., McGraw-Hill Book Company, Inc., New York, 1949.

GLASGOW, R. S.: "Principles of Radio Engineering," McGraw-Hill Book Company, Inc., New York, 1936.

REICH, H. J.: "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1944.

SPANGENBERG, K. R.: "Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1948.

TERMAN, F. E.: "Radio Engineering," 3d ed., McGraw-Hill Book Company, Inc., New York, 1947.

# CHAPTER 17

# TRIODES AS CIRCUIT ELEMENTS

THE analysis of the behavior of a vacuum tube in a circuit may be accomplished by two different methods, both of which are important, and both of which will be examined in some detail. Initially, it will be shown that the behavior of a triode as a circuit element may be obtained graphically from a knowledge of the static plate characteristic curves. This is essentially the same procedure that was followed in treating the diode as a circuit element, except that the diode had but two active terminals and one characteristic curve, whereas the triode has three active terminals and a family of curves. The three terminals are marked $P$ (plate), $K$(cathode), and $G$ (grid).

The second method, which is applicable when the triode parameters $\mu$, $r_p$, and $g_m$ are essentially constant over the operating range, is an analytical one. The tube will then be replaced by a generator and a resistor, and conventional linear-circuit theory will be used in the analysis.

**17-1. Symbols and Terminology.** The simplest circuit in which the triode acts as an amplifier is that shown in Fig. 17-1. Before proceeding with an analysis of this circuit, it is necessary to explain the meanings of
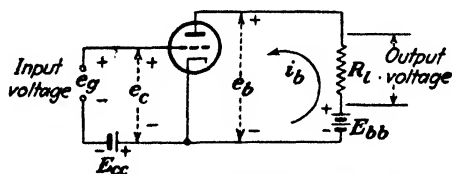


Fig. 17-1. The basic circuit of a triode used as an amplifier.

the symbols and the terminology to be used in this and subsequent analyses.

The input circuit of an amplifier refers to all elements of the circuit that are connected between grid and cathode terminals of the tube. Similarly, the output or plate circuit usually refers to the elements that are connected between the plate and cathode terminals. In the circuit illustrated the output, or plate, circuit contains a d-c supply voltage which is in series with a load resistor $R_l$. The input, or grid, circuit consists of a d-c supply voltage in series with the input voltage. The input voltage

may have any wave shape whatsoever, but it is usually chosen, for convenience in analysis, to be a sinusoidally varying voltage.

Owing to the fact that a variety of potentials and currents, both d-c and a-c, are involved simultaneously in a vacuum-tube circuit, it is necessary that a precise method of labeling be established, if confusion is to be avoided.

In what follows, lower-case letters will be used to designate instantaneous values, and capital letters will be used to denote either d-c values or rms values of sinusoids. The subscripts $c$ and $g$ will refer to the grid circuit, and the subscripts $b$ and $p$ will refer to the plate circuit. Examples of the notation follow:

$E_{cc}$ = constant d-c grid voltage, called the $C$ bias or grid bias.
$E_{bb}$ = constant d-c plate voltage, called the $B$ supply or the plate supply voltage.

It will be assumed in everything that follows that the cathode is the reference point, and all potentials will be measured with respect to it. Then,

$e_c$ = instantaneous potential difference of the grid with respect to the cathode and is positive when the grid is positive with respect to the cathode.
$e_b$ = instantaneous voltage of the plate with respect to the cathode and is positive when the plate is positive with respect to the cathode.
$i_b$ = instantaneous plate current and is positive in the direction from the cathode to the plate *through the load* (see Fig. 17-1).
$e_g$ = instantaneous value of the grid input signal voltage and is positive if the grid terminal is positive with respect to the second input terminal.

If, as is often the case, the input is a sine wave, then

$E_g$ = rms value of the a-c input excitation voltage or input signal.
$E_p$ = rms value of the a-c output voltage.

For example, if the input signal voltage is sinusoidal and of the form

$$e_g = \sqrt{2}E_g \sin \omega t$$

then the net grid voltage in Fig. 17-1 is

$$e_c = -E_{cc} + \sqrt{2}E_g \sin \omega t$$

where $E_{cc}$ is the *magnitude* of the bias voltage.

Table 17-1 summarizes the notation introduced above. In the table are also listed some symbols not yet defined, but which will be used in later sections. This table should serve as a convenient reference until the reader is thoroughly familiar with the notation.

TABLE 17-1

TRIODE SYMBOLS

|  | Grid voltage with respect to cathode | Plate voltage with respect to cathode | Current in direction toward plate through the load |
|---|---|---|---|
| Instantaneous total value..... | $e_c$ | $e_b$ | $i_b$ |
| Quiescent value............. | $E_c$ | $E_b$ | $I_b$ |
| Instantaneous value of varying component............ | $e_g$ | $e_p$ | $i_p$ |
| Effective value of varying component.................. | $E_g$ | $E_p$ | $I_p$ |
| Amplitude of varying component...................... | $E_{gm}$ | $E_{pm}$ | $I_{pm}$ |
| Supply voltage.............. | $E_{cc}$* | $E_{bb}$* |  |

* These are positive numbers giving the *magnitude* of the voltages.

**17-2. Graphical Analysis of the Circuit.** Suppose for the moment that
no grid signal is applied in Fig. 17-1, so that $e_g = 0$. It must not be sup-
posed that there will be no plate current. This might be true if the bias
is very negative. However, in general, a definite d-c current will exist
when the input signal is zero. The value of this current may be found
graphically in the same way as that used to determine the instantaneous
plate current in the diode circuit of Fig. 12-1 for a given instantaneous
transformer voltage.

Because of the presence of the load resistor $R_l$, the potential that exists
between the plate and the cathode will depend upon both the magnitude
of the battery supply and the magnitude of the current in the load resistor.
It follows from Fig. 17-1 that

$$e_b = E_{bb} - i_b R_l \qquad (17\text{-}1)$$

Here, as in the case of the diode, this one equation is not sufficient to deter-
mine the current corresponding to any voltage $E_{bb}$ because there are two
unknown quantities in this expression, $e_b$ and $i_b$.

A second relation between these two variables is given by the plate
characteristics of the triode (see Fig. 16-7a). The straight line represented
by Eq. (17-1) is plotted on the plate characteristics of Fig. 17-2. This line
is obviously independent of the tube characteristics, for it depends only
upon elements external to the tube itself. The intersection of this line with
the curve for $e_c = -E_{cc}$ is called the "operating point" or the "quiescent

point," $Q$. The current in the external circuit is $I_b$, and the corresponding plate potential is $E_b$.

The simplest method of drawing the load line is to locate two points of this line and to connect these with a straightedge. One such point is the intersection with the horizontal axis, namely, $i_b = 0$ and $e_b = E_{bb}$. Another is the intersection with the vertical axis, namely, $e_b = 0$ and $i_b = E_{bb}/R_l$. These are illustrated in Fig. 17-2. Sometimes this latter point falls off the printed plate characteristics supplied by the manufacturer, the current $E_{bb}/R_l$ being considerably greater than the rated tube current. In such a situation any value of current, say $i_0$, that is given on the plate characteristics is



FIG. 17-2. The operating point $Q$ is located at the intersection of the load line and the plate characteristic for the bias $e_c = -E_{cc}$.

chosen and the corresponding plate voltage is found from Eq. (17-1), namely, $E_{bb} - i_0R_l$.

*Example.* A 6C5 triode is operated at a bias of $-8$ volts and a plate supply of 360 volts. If the load resistance is 20,000 ohms, what are the quiescent current and voltage values?

*Solution.* The plate characteristics of the 6C5 are given in Fig. A9-3 of Appendix IX. One point on the load line is $i_b = 0$ and $e_b = 360$. Corresponding to $e_b = 0$ the value of $i_b$ is $E_{bb}/R_l = 360/20 = 18$ ma, whereas the largest current on Fig. A9-3 is 14 ma. Hence a second point on the load line is found by choosing $i_b = 10 = i_0$, and then

$$e_b = E_{bb} - i_0R_l = 360 - (10 \times 10^{-3})(20,000) = 360 - 200 = 160$$

The load line is now drawn through the pair $(i_b, e_b)$ of points $(0, 360)$ and $(10, 160)$ on Fig. A9-3. This line is found to intersect the plate curve corresponding to $E_c = -8$ at a plate current of 6.4 ma $= I_b$ and a plate voltage of 233 volts $= E_b$. [The plate voltage corresponding to 6.4 ma is actually $E_b = 360 - (6.4)(20) = 232$ volts, in good agreement with the 233 volts read directly from the graph.]

If the voltage of the grid with respect to the cathode in the above example is changed to $-14$ volts, then the plate current is found to be approximately 3.0 ma (this is the intersection of the load line with $E_c = -14$ volts). If the grid voltage is changed to $-2$ volts, then the plate current is found to be 10.5 ma (the reader should check these values). A change in input voltage of $-2 - (-14) = 12$ volts has resulted in a change of plate current of $10.5 - 3.0 = 7.5$ ma and a corresponding change in output voltage across the 20,000-ohm load of $(20)(7.5) = 150$ volts. Hence, the output is $\frac{150}{12} = 12.5$ times as great as the input. This illustrates the amplifying action of the tube.

The above method of finding the output current corresponding to a given input voltage will now be discussed in more detail. Suppose that the grid potential is

$$e_c = -E_{cc} + \sqrt{2}E_g \sin \omega t$$

The maximum and minimum values of $e_c$ will be $-E_{cc} \pm \sqrt{2}E_g$, which indicates that the grid swings about the point $-E_{cc}$. Consequently, the
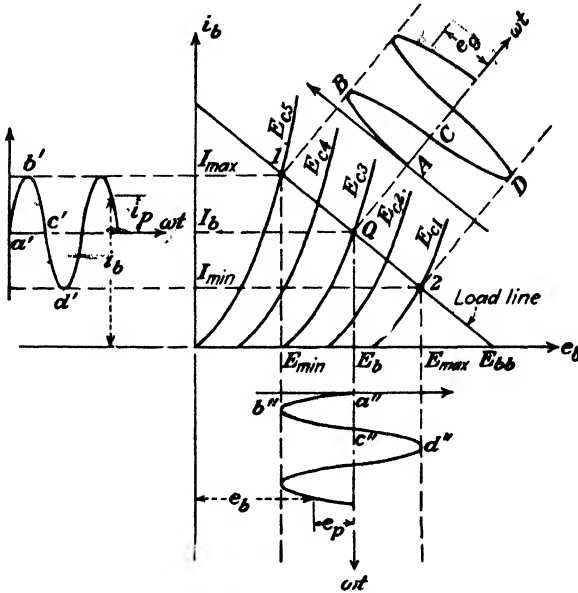


FIG. 17-3. The output current and voltage wave forms for a given input-grid signal are determined from the plate characteristics and the load line.

plate current and the plate voltage will then swing about the values $I_b$ and $E_b$, respectively. The graphical construction showing these conditions is illustrated in Fig. 17-3. *For any given value of $e_c$, the corresponding values of $i_b$ and $e_b$ are located at the intersection of the load line and the $i_b$-$e_b$ curve corresponding to this value of $e_c$.* This construction is valid *for any input wave form* and is not restricted to sinusoidal voltages. This permits the wave form of the output current and the output voltage to be determined graphically. The points $a'$, $b'$, $c'$, etc., of the output current, and the points $a''$, $b''$, $c''$, etc., of the output voltage wave correspond, respectively, to the points $A$, $B$, $C$, etc., of the input grid-voltage wave form.

A word is in order about the manner in which the grid voltage is drawn on the diagram. The time axis is shown normal to the load line. This is done to show that the signal voltage $e_g$ causes the grid voltage to vary

about the value $Q$ or $E_{c3}$ from $E_{c5}$ at one extreme of the signal voltage swing to $E_{c1}$ at the other extreme of this swing. This representation has no real significance but does serve to clarify the operating path.

**17-3. Variations from Quiescent Values.** Suppose that in Fig. 17-1 $e_g$ represents the output from a microphone and that $R_l$ is the effective resistance of a loud-speaker. There is no particular interest in the quiescent current, which is the current to the speaker when no one talks into the microphone. (Actually, the speaker would be transformer coupled into the plate circuit, and the current in the secondary under quiescent conditions would be zero.) The principal interest is in the speaker output for a given microphone output. Thus the variations in current and voltage with respect to the quiescent values are most important.

If the load is a resistor and not a speaker, and if the output from this resistor is taken through a coupling capacitor (as is usually the case), then under zero input conditions the capacitor will charge up to the quiescent voltage $E_b$. The voltage with respect to cathode from the side of the capacitor not connected to the plate is zero under these conditions. If a varying grid voltage is now added to the bias, the output will again represent voltage variations about the quiescent value.

It is evident that the significant quantities are the currents and voltages with respect to their quiescent values. To examine this matter in some detail, refer to the oscillograms of Figs. 17-4 and 17-5. These oscillograms show curves of $e_c$ and the corresponding values of $i_b$ and $e_b$ for a particular triode circuit. In the curves of Fig. 17-5, the variations of the voltage and current about the $Q$ point are indicated separately. It is seen that the output current, defined by the equation

$$i_p = i_b - I_b \tag{17-2}$$

is simply the current variation about the quiescent-point current $I_b$. The output voltage $e_p$, which is similarly defined, represents the potential variations about the $Q$ point. Consequently, if the input signal is a pure sinusoidal wave and if the tube characteristics are equidistant lines for equal intervals of $e_c$, $i_p$ will also be a sinusoidal wave. If the characteristic curves are not equidistant lines over the range 1-2 for equal intervals of $e_c$, the wave form of $i_p$ will differ from that of the input-signal wave form. The latter condition will give rise to harmonics, since a nonsinusoidal wave may be expressed as a Fourier series in which some of the higher harmonic terms are appreciable. These considerations should be clear if reference is made to Figs. 17-3 and 17-4.

Corresponding to Eq. (17-2) the variables $e_p$ and $e_g$ are defined by the equations

$$e_p = e_b - E_b \qquad e_g = e_c + E_{cc} \tag{17-3}$$

If the symbol $\Delta$ is used to denote a change from the quiescent value, then

$$\Delta e_b \equiv e_p \qquad \Delta e_c \equiv e_g \qquad \Delta i_b \equiv i_p \qquad (17\text{-}4)$$

An inspection of Fig. 17-5 reveals the following extremely significant result: If $i_p$ is a sine wave, then $i_p$ and $e_p$ are exactly 180 deg out of phase
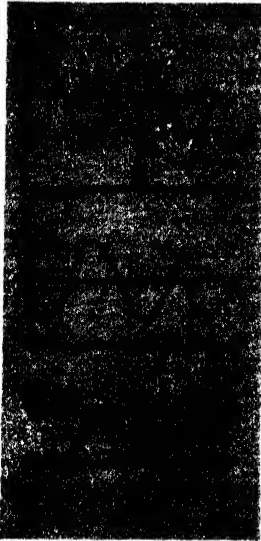


Fig. 17-4.                    Fig. 17-5.

Fig. 17-4.   Oscillogram of the curves $e_b$, $i_b$, and $e_c$ corresponding to Fig. 17-3.

Fig. 17-5.   Oscillogram of the signal voltage $e_g$ and the corresponding output current $i_p$ and output voltage $e_p$. This is the same oscillogram as in Fig. 17-4 except that the axes have been shifted so that the curves now represent variations about the $Q$ point, thus

$$e_p = e_b - E_b$$
$$i_p = i_b - I_b$$
$$e_g = e_c + E_{cc}$$

with each other. This result follows from the fact that $i_p$ is a maximum, whereas $e_p$ is a minimum at the point 1. The reverse is true at the point 2 (see Fig. 17-3).

**17-4. The Dynamic Transfer Characteristic.** The "dynamic" characteristic of the triode may be constructed from the plate characteristics of the tube and the load line. This is simply the curve relating $i_b$ with $e_c$. By comparing Fig. 17-6 with Fig. 17-3, it will be seen that the same points on each are similarly marked. That is, the points $a'$, $b'$, $c'$, etc., of the output current correspond to the points $A$, $B$, $C$, etc., of the input grid-voltage wave $e_g$.

The dynamic curve has the same utility as the dynamic curve of the diode, *viz.*, it permits the output wave for a given input wave to be deduced. For triodes the input is in the grid circuit, whereas, for a diode, it is in the plate circuit. It should be emphasized that a different dynamic curve must be drawn for each value of load resistance.
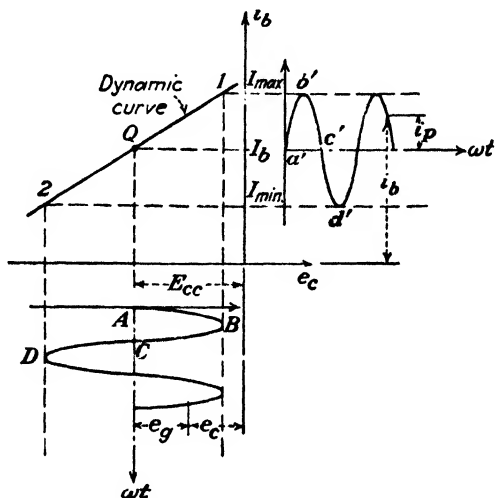


Fig. 17-6. The dynamic transfer characteristic is used to determine the output-wave shape for a given input signal.

**17-5. Equivalent Voltage-source Representation of a Triode.** As already discussed in Sec. 17-3, one is ordinarily interested only in the variation in voltage and current about the $Q$ point, rather than in the total values of these quantities. That is, the primary interest is in determining the values of $e_p$ and $i_p$ for a given value of $e_g$. This may be referred to as the "a-c response" of the tube, since $e_g$ will usually be a periodically varying voltage.

The graphical methods of the foregoing sections are tedious to apply and often are very inaccurate. Certainly if the input signal is very small, say 0.1 volt or less, then values cannot be read from the plate characteristic curves with any degree of accuracy. But for such small input signals, the parameters $\mu$, $r_p$, and $g_m$ will remain substantially constant over the small operating range. Under these conditions it is possible to replace the graphical method by an analytical one. This is often called the *small-signal method*, but it is applicable even for large signals provided only that the tube parameters are constant over the range of operation. The constancy of the parameters is judged by an inspection of the plate characteristics. If these are straight lines, equally spaced for equal intervals of grid bias over the operating range, then the parameters are constant. Under these con-

ditions it will be found that the tube may be replaced by a simple linear system. The resulting circuit may then be analyzed by the general methods of a-c circuit analysis.

In order to find the equivalent linear representation of the tube, the variation in current $\Delta i_b$ about the quiescent value is considered. If the grid voltage remains constant but the plate voltage changes by an amount $\Delta e_b$, then the change in current equals the rate of change of current with plate voltage times the change in plate voltage, or

$$\Delta i_b = \left(\frac{\partial i_b}{\partial e_b}\right)_{E_c} \Delta e_b$$

The subscript indicates the variable held constant in performing the partial differentiation. This is illustrated in Fig. 17-7 and is seen to be
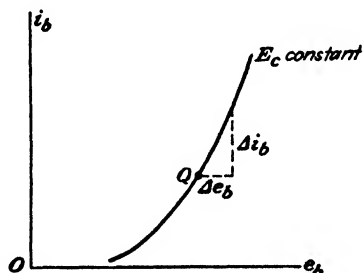


FIG. 17-7. If the grid voltage is constant, then $\Delta i_b = (\text{slope})(\Delta e_b) = \left(\dfrac{\partial i_b}{\partial e_b}\right)_{E_c} \Delta e_b$.

strictly true only if the slope of the plate characteristic is constant for the assumed change in current. Similarly, if the plate voltage remains constant but the grid voltage changes by $\Delta e_c$, then the change in current is given by

$$\Delta i_b = \left(\frac{\partial i_b}{\partial e_c}\right)_{E_b} \Delta e_c$$

If both the grid and plate voltages are varied, then the plate-current change is the sum of the two changes indicated above, or

$$\Delta i_b = \left(\frac{\partial i_b}{\partial e_b}\right)_{E_c} \Delta e_b + \left(\frac{\partial i_b}{\partial e_o}\right)_{E_b} \Delta e_c \qquad (17\text{-}5)$$

As mentioned above, this expression is only approximate. It is, in fact, just the first two terms of the Taylor's series expansion of the function

$i_b(e_b, e_c)$.  In the general case,

$$\Delta i_b = \left(\frac{\partial i_b}{\partial e_b}\right)_{E_c} \Delta e_b + \left(\frac{\partial i_b}{\partial e_c}\right)_{E_b} \Delta e_c + \frac{1}{2}\left(\frac{\partial^2 i_b}{\partial e_b{}^2}\right)_{E_c}(\Delta e_b)^2 + \frac{1}{2}\left(\frac{\partial^2 i_b}{\partial e_c{}^2}\right)_{E_b}(\Delta e_c)^2$$

$$+ \frac{\partial^2 i_b}{\partial e_b \, \partial e_c} \Delta e_b \, \Delta e_c + \cdots \quad (17\text{-}6)$$

Consider the third term in this expansion.  Since from Eq. (16-6) the plate resistance is given by $1/r_p = (\partial i_b/\partial e_b)_{E_c}$, this term equals

$$\frac{1}{2}\left[\frac{\partial(1/r_p)}{\partial e_b}\right]_{E_c}(\Delta e_b)^2$$

Similarly, the fourth-, fifth-, and higher-order terms in Eq. (17-6) represent derivatives of $r_p$ and $g_m$ with respect to plate and grid voltages.

The present method of analysis is based on the assumption that the tube parameters are sensibly constant over the operating range $\Delta e_b$ and $\Delta e_c$. Under these conditions a satisfactory representation of the variations in plate current about the quiescent point is given by Eq. (17-5).  This expression may be written in the following form, by virtue of Eqs. (16-6):

$$\Delta i_b = \frac{1}{r_p} \Delta e_b + g_m \, \Delta e_c \quad (17\text{-}7)$$

Since $g_m = \mu/r_p$,* this becomes

$$r_p \, \Delta i_b = \Delta e_b + \mu \, \Delta e_c \quad (17\text{-}8)$$

where, as before, the $\Delta$'s denote changes about the quiescent point.  Using the notation of Eq. 17-4, this becomes

$$i_p r_p = e_p + \mu e_g$$

or

$$e_p = -\mu e_g + i_p r_p \quad (17\text{-}9)$$

This expression may be given a very significant physical interpretation.

* This follows from Eq. (17-7).  Thus, if the plate current is constant so that $\Delta i_b = 0$, then

$$0 = \frac{\Delta e_b}{r_p} + g_m \, \Delta e_c$$

or

$$-\frac{\Delta e_b}{\Delta e_c} = g_m r_p$$

But since the plate current has been taken to be constant, then $-\Delta e_b/\Delta e_c$ is by definition [Eq. (16-5)] the amplification factor.  Hence, $\mu = g_m r_p$.

It shows that the varying voltage $e_p$ with respect to the $Q$ point is made up of two components: one is a generated emf which is $\mu$ times as large as the grid-cathode voltage variation $e_g$; the second is a varying voltage across the tube resistor $r_p$ that results from the varying load current $i_p$ through it.

The result of this discussion is best illustrated diagrammatically, the circuit that gives rise to Eq. (17-9) being shown in Fig. 17-8. This diagram also includes a schematic of the tube itself in order to stress the correspondence between it and its equivalent representation. For greater preciseness in the representation, double subscripts have been employed in this di-



(a)                              (b)
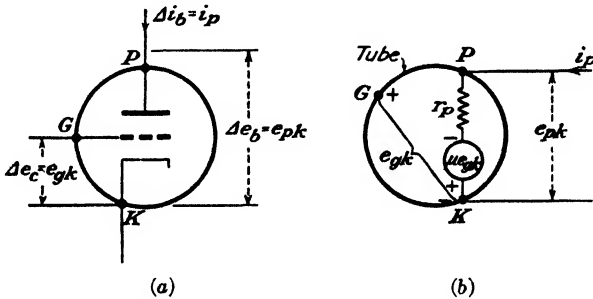
FIG. 17-8. A triode $(a)$ and its equivalent voltage-source representation $(b)$. The voltage drop from the point $P$ to the point $K$, $e_{pk}$, is seen from the diagram to equal $-\mu e_{gk} + i_p r_p$. The brace from $G$ to $K$ indicates that the quantity $e_{gk}$ is to be evaluated by traversing the circuit from $G$ to $K$ and adding all the voltage *drops* on the way.

agram. Note that the tube is replaced by a fictitious generator having a generated emf equal to $\mu e_{gk}$ and an internal resistance of $r_p$ ohms. It is seen from the diagram that the voltage drop $e_{pk}$ from plate to cathode is equal to the voltage drop in the plate resistor less the generator voltage, or

$$e_{pk} = i_p r_p - \mu e_{gk}$$

This is exactly Eq. (17-9), which verifies that Fig. 17-8 is the correct equivalent circuit representation of the tube.

A point of the utmost importance is that no d-c quantities are indicated on this diagram. This is so because the equivalent circuit of the tube applies only for changes about the $Q$ point. Moreover, the equivalent tube circuit representation is valid for any type of load whether it be a pure resistance, an impedance, or another tube. This is true because the above derivation was accomplished without any regard to the external circuit in which the tube is incorporated. The only restriction is that the parameters $\mu$, $r_p$, and $g_m$ must remain substantially constant over the operating range.

If sinusoidally varying quantities are involved in the circuit, and this is usually assumed to be the case, the analysis proceeds most easily if the sinors of elementary a-c circuit theory are introduced. The circuit notation used in this text is discussed in Appendix VIII, and the reader is urged to read this very carefully before proceeding further. For the case of
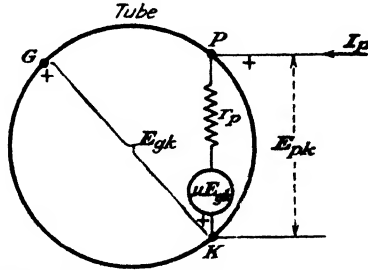


Fig. 17-9. For sinusoidally varying quantities the tube is replaced by this simple network.

sinusoidally varying quantities, the tube is replaced by the equivalent network shown in Fig. 17-9.

**17-6. Linear Analysis of the Circuit.** Based on the foregoing discussion, the amplifier circuit of Fig. 17-1 may be replaced by an equivalent form which permits an analytic determination of its a-c operation. The following simple rules should be adhered to in drawing the equivalent form of even relatively complicated amplifier circuits.

1. Draw the actual wiring diagram of the circuit neatly.
2. Mark the points $G$, $P$, and $K$ on this circuit diagram. Locate these points as the start of the equivalent circuit. Maintain the same relative positions as in the original circuit.
3. Replace the tube by its linear equivalent form (Fig. 17-9).
4. Transfer all circuit elements from the actual circuit to the equivalent circuit of the amplifier. Keep the relative positions of these elements intact.
5. Replace each d-c source by its internal resistance, or by a short circuit, if its resistance is negligible.

A point of special importance is that regardless of the form of the input circuit, the fictitious generator that appears in the equivalent representation of the tube is *always* $\mu E_{gk}$, where $E_{gk}$ is the drop from grid to cathode. The positive reference terminal of the generator is *always* at the cathode.

To illustrate the application of these rules, four examples will be given. The first is a single-mesh circuit involving resistors only, the results being given in terms of symbols rather than numerical values. The second illustration is a simple circuit which includes reactive elements and which requires a numerical result. The third is a two-mesh circuit. The fourth network contains two tubes.

*Example* 1. Find the a-c output current and voltage of the basic triode amplifier circuit illustrated in Fig. 17-10a.



(a)-Actual circuit        (b)-Equivalent circuit
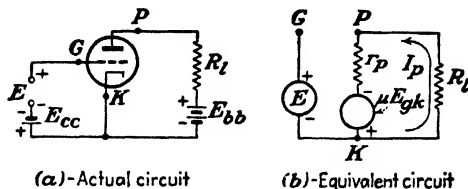
FIG. 17-10.   The schematic and equivalent circuits of a simple amplifier.

*Solution.* According to the foregoing rules, the equivalent circuit is that of Fig. 17-10b. Kirchhoff's voltage law, which requires that the sum of the voltage drops around the circuit equal zero, yields

$$I_p R_l + I_p r_p - \mu E_{gk} = 0$$

A glance at this circuit shows that the voltage drop from grid to cathode is **E**. Hence $\mathbf{E}_{gk} = \mathbf{E}$, and the output current $\mathbf{I}_p$ is

$$I_p = \frac{\mu E}{R_l + r_p}$$

The corresponding output-voltage drop is

$$E_{pk} = -I_p R_l$$

The minus sign arises because the direction from $P$ to $K$ is opposite to the positive reference direction of the current $\mathbf{I}_p$.

$$E_{pk} = \frac{-\mu E R_l}{R_l + r_p}$$

The *gain*, or *voltage amplification*, of the tube circuit is defined as the ratio of the output- to input-voltage drops. It is, for the simple amplifier of Fig. 17-10,

$$K \equiv \frac{E_{out}}{E_{in}} = \frac{E_{pk}}{E_{gk}} = \frac{E_{pk}}{E} = -\mu \frac{R_l}{R_l + r_p} = -\mu \frac{1}{1 + r_p/R_l} \quad (17\text{-}10)$$

The minus sign signifies a phase shift of 180 deg between the output and the input voltages. This is in agreement with the result obtained from the graphical analysis of the problem in the previous sections. Equation (17-10) is a very important result and one well worth remembering.
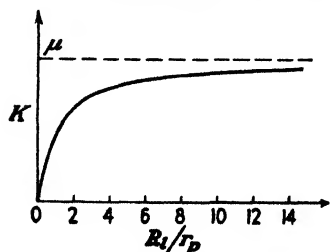


FIG. 17-11.  The gain of the amplifier of Fig. 17-9 as a function of the load resistance.  $\mu$ and $r_p$ are assumed to be constant.

The magnitude of the gain increases with the load resistance and approaches a maximum value as $R_l$ becomes much greater than $r_p$. The general form of this variation is illustrated in Fig. 17-11. Since $K$ is greater than unity, in general,

the meaning of the term "amplifying action of the tube" is clear.   It is noted that *the maximum possible gain is* $\mu$, although this can be obtained only if $R_l = \infty$.   Too large a value of $R_l$ cannot be used, however, since for a given quiescent current this would require an impractically high power supply voltage.   Nevertheless, since $K$ increases rapidly at first and then approaches $\mu$ asymptotically, a gain approaching $\mu$ may be realized with a reasonable value of $R_l$.

*Example* 2.   A 1-volt rms potential at 800 cps is impressed on the grid of a tube for which

$$\mu = 8 \qquad r_p = 5,000 \text{ ohms}$$

The load impedance is

$$Z_l = 1,000 + j6,000 = 6.08 \times 10^3 \underline{/80.5°}$$

Calculate the voltage gain of the circuit, and draw the complete sinor diagram of the system.

*Solution.*   The actual and equivalent circuits are shown in Fig. 17-12.   Evidently $E_{gk} = 1$, and the fictitious-generator voltage is $\mu E_{gk} = 8$ volts rms.   Employing the
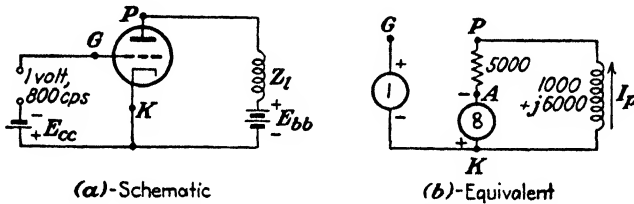


FIG. 17-12.   A numerical example.

customary analysis using complex numbers, the voltage drop from $G$ to $K$ is $1 + j0$. This voltage is chosen along the axis of reals, for convenience.   $E_{KA}$ is in phase with this voltage and equals $8 + j0$.   Therefore,

$$I_p = \frac{8 + j0}{5,000 + (1,000 + j6,000)} = 0.943 \times 10^{-3} \underline{/-45°} \text{ amp}$$

The output-voltage drop is

$$E_{pk} = -I_p Z_l = -0.943 \times 10^{-3} \underline{/-45} \times 6.08 \times 10^3 \underline{/80.5} = -5.73 \underline{/35.5}$$

$$= 5.73 \underline{/-144.5°} \text{ volts}$$

The voltage gain of the system is

$$K = \frac{\text{output-voltage drop}}{\text{input-voltage drop}} = \frac{5.73 \underline{/-144.5°}}{1 \underline{/0°}} = 5.73 \underline{/-144.5°}$$

The gain $K$ is a complex number.   Its magnitude gives the ratio of output to input voltages.   Its phase angle gives the number of degrees by which the output leads the

input. In this illustration the input leads the output by 144.5 deg. The complete sinor diagram of the circuit is shown in Fig. 17-13.
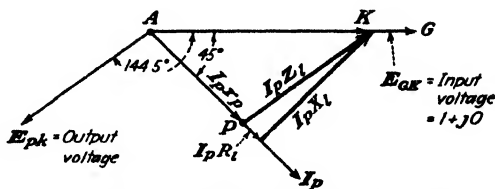


FIG. 17-13. The sinor diagram for the circuit of Fig. 17-12.

*Example* 3. In the circuit sketched in Fig. 17-14a the input signal $E$ is 0.5 volt rms at 2,000 cps. If the 12,000 ohms represents the resistance of an a-c voltmeter, what will this meter read?
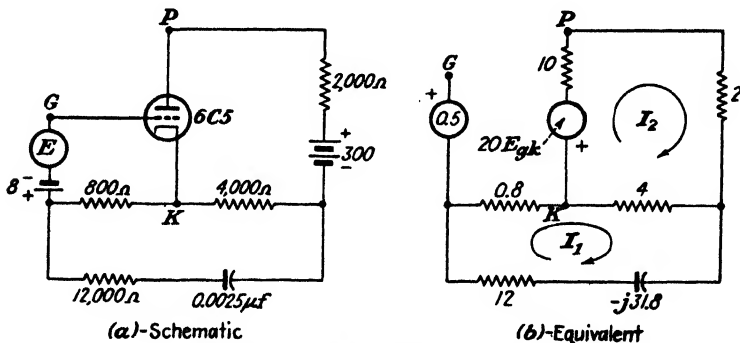


FIG. 17-14. A network with two meshes.

*Solution.* The quiescent values will first be found. Since the capacitor cannot pass d-c, then the branch of the circuit containing the capacitor is effectively open as far as the quiescent point is concerned. The grid bias is $-8$ volts since the d-c drop through the 800 ohms is zero (assuming zero grid current). A load line corresponding to a load of $2,000 + 4,000 = 6,000$ ohms and a plate supply of 300 volts gives $E_b = 250$ volts and $I_b = 8$ ma at $E_c = -8$ volts. These are just the values recommended by the manufacturers, and at this point they list $\mu = 20$ and $r_p = 10,000$ ohms (see Fig. A9-3). The equivalent circuit is shown in Fig. 17-14b. All impedance values have been divided by 1,000, for convenience, so.that the values actually represent kilohms. It is understood that the *mesh* currents $I_1$ and $I_2$ are in milliamperes. (Note that the product of kilohms and milliamperes is volts.) The magnitude of the reactance of the capacitor is

$$X_C = \frac{1}{2\pi f C} = \frac{10^6}{(2\pi)(2,000)(0.0025)} \text{ ohms} = 31.8 \text{ kilohms}$$

The reference directions for the mesh currents are completely arbitrary and have been chosen clockwise. It is most important to note that $E_{gk}$ is *not* equal to the input voltage. It can be found by traversing the network from the grid to the cathode and adding all

the voltage *drops* encountered. Any path from $G$ to $K$ may be chosen but the most direct one is usually taken since it involves the least amount of labor. Thus

$$E_{gk} = 0.5 + 0.8I_1 \qquad (17\text{-}11)$$

Adding the voltage drops in the direction of the current $I_1$ and equating the result to zero yields

$$(12 + 0.8 + 4 - j31.8)I_1 - 4I_2 = 0$$

or

$$I_2 = (4.2 - j7.95)I_1 \qquad (17\text{-}12)$$

Similarly, Kirchhoff's voltage law around mesh 2 gives

$$-4I_1 + (10 + 2 + 4)I_2 + 20E_{gk} = 0 \qquad (17\text{-}13)$$

Equations (17-11), (17-12), and (17-13) determine the solution. Thus, putting Eqs. (17-11) and (17-12) into (17-13) yields

$$-4I_1 + (16)(4.2 - j7.95)I_1 + (20)(0.5 + 0.8I_1) = 0$$

or

$$I_1 = -0.0354 - j0.0567 \text{ ma}$$

Since the capacitor blocks the quiescent current from getting to the meter, it will read only the product of its resistance and the a-c current through it. It is this a-c current which we have just calculated from the equivalent circuit (which is valid only for variations from the quiescent value). The rms voltage is therefore

$$(12)[(0.0354)^2 + (0.0567)^2]^{\frac{1}{2}} = 0.80 \text{ volt}$$

*Example* 4. Draw the equivalent circuit for the network of Fig. 17-15a, and set up the equations from which the a-c currents and voltages may be found. Since there
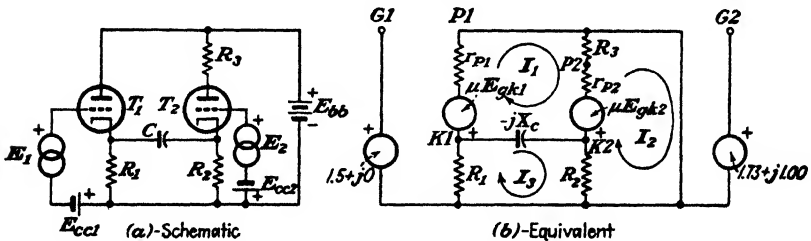


Fig. 17-15. A two-tube problem.

are two generators shown in the network, the problem is not uniquely specified unless the phase between these two is given. Assume, therefore, that with the reference polarities indicated, $E_2$ leads $E_1$ by 30 deg and that the rms value of $E_1$ is 1.50 volts and that of $E_2$ is 2.00 volts.

*Solution.* The phasor $E_1$ is arbitrarily chosen along the horizontal axis, so that

$$E_1 = 1.50 + j0$$

and

$$E_2 = (2.00)\underline{/30°} = (2.0)(\cos 30° + j \sin 30°) = 1.73 + j1.00$$

The points $K$, $P$, and $G$ are indicated for each tube. The numbers 1 or 2 are used to designate which tube is under consideration. For example, $P2$ is the plate of the second

tube; $\mathbf{E}_{gk1}$ represents the voltage drop from grid to cathode of tube 1. The rules given at the beginning of this section are followed for each tube separately, and the resultant equivalent circuit is shown in Fig. 17-15b.

The positive reference directions for the mesh currents are arbitrarily taken clockwise. For any network the fictitious generators $\mu \mathbf{E}_{gk}$ of each tube *must* have the positive reference polarity at its respective cathode independent of any other voltage or current polarities in the circuit.

It should be noted that the battery $E_{bb}$ has been replaced on the equivalent network by a short circuit from the plate of $P1$ to the junction of $R_1$ and $R_2$. (This assumes that the battery has zero internal resistance.)

The most systematic way of setting up the equations is as follows:

1. Traverse each mesh in the current reference direction.
2. Find the voltage drops due to $I_1$ flowing through all the passive impedances in this mesh. Use a plus sign for the voltage drops if traversing the circuit in the current reference direction and a minus sign if traveling in the opposite direction.
3. Repeat for $I_2$ and all other currents.
4. Find the voltage drops of all the generators (in the chosen traversing direction).
5. Set the sum of all these equal to zero.

Applying these rules to mesh 1 yields

$$+ (r_{p1} + R_3 + r_{p2} - jX_C)I_1 - (R_3 + r_{p2})I_2 - (-jX_C)I_3 + \mu \mathbf{E}_{gk_1} - \mu \mathbf{E}_{gk_2} = 0 \quad (17\text{-}14)$$

Similarly for mesh 2, there is obtained

$$-(R_3 + r_{p2})I_1 + (R_2 + r_{p2} + R_3)I_2 - R_2I_3 + \mu \mathbf{E}_{gk2} = 0 \quad (17\text{-}15)$$

and, for mesh 3, there results

$$-(-jX_C)I_1 - R_2I_2 + (R_1 - jX_C + R_2)I_3 = 0 \quad (17\text{-}16)$$

Before it is possible to solve for the currents, $\mathbf{E}_{gk_1}$ and $\mathbf{E}_{gk_2}$ must be found. Adding drops from $G1$ to $K1$ gives

$$\mathbf{E}_{gk_1} = 1.50 + j0 + R_1I_3 \quad (17\text{-}17)$$

Adding drops from $G2$ to $K2$ yields

$$\mathbf{E}_{gk_2} = 1.73 + j1.00 + R_2I_2 - R_2I_3 \quad (17\text{-}18)$$

Substitute from Eqs. (17-17) and (17-18) into Eqs. (17-14) and (17-15). The resulting equations and Eq. (17-16) will then be a set of three for the three unknowns $I_1$, $I_2$, and $I_3$ in terms of the circuit constants.

**17-7. Measurement of Triode Coefficients.**[1] The values of $\mu$, $r_p$, and $g_m$ illustrated in Fig. 16-11 may be determined graphically from the static characteristic curves given in Figs. 16-7a, 16-7b, and 16-8 by drawing tangents to these curves and determining the slopes at the points in question. The accuracy with which these determinations may be made is generally low, and it is usually more desirable to measure these quantities dynamically by means of suitable a-c bridge networks. The analysis of the circuits employed for these measurements is considerably simplified through the use of the equivalent circuit.

The amplification factor $\mu$ is readily determined by means of the circuit shown in Fig. 17-16a. The equivalent circuit of this network is given in Fig. 17-16b. The operations involved in balancing this bridge consist simply in varying $R_1$ and $R_2$ until no a-c signal from the oscillator is heard

in the earphones. This requires, of course, that the current $I_1$ be zero. When this condition is satisfied, $\mu = R_2/R_1$. This can be proved as follows: The voltage drop from grid to cathode is $\mathbf{E}_{gk} = IR_1$. By applying
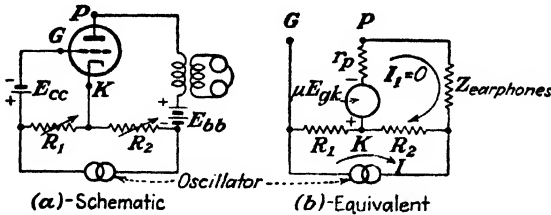


Fig. 17-16. The Miller bridge circuit for determining the amplification factor of a triode under operating conditions.

Kirchhoff's law to the plate circuit and remembering that $I_1$ must be zero under the conditions of balance, then

$$+ \mu\mathbf{E}_{gk} - IR_2 = 0$$

or

$$\mu\mathbf{E}_{gk} = IR_2 = \mu IR_1$$

from which it follows that

$$\mu = \frac{R_2}{R_1} \tag{17-19}$$

This is an extremely convenient method for measuring the amplification factor of the tube. This may be done for any desired value of d-c plate current $I_b$ in the tube simply by adjusting the grid bias $E_{cc}$. The telephone is preferably connected to the secondary of a small transformer so that the d-c current does not pass through the telephone and thereby polarize it.
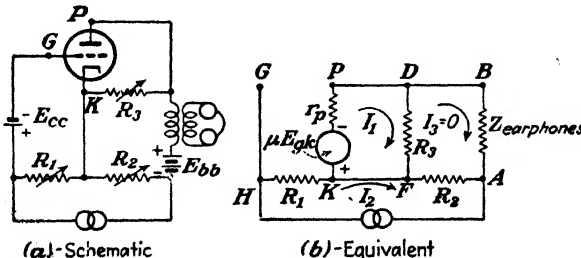


Fig. 17-17. The Miller bridge circuit for determining the transconductance of a triode under operating conditions.

In order to measure the transconductance $g_m$, the foregoing circuit is modified slightly. This modification consists in inserting the resistor $R_3$ between the cathode and the plate. The schematic and equivalent circuits of this modified network are shown in Fig. 17-17   To make this measure-

ment, the resistances are varied until no sound is heard in the telephone. The current in mesh 3 is zero, since this is the current in the telephone at balance. Kirchhoff's law is applied to mesh 1. The sum of the voltage drops is

$$(R_3 + r_p)\mathbf{I}_1 + \mu\mathbf{E}_{gk} = 0$$

The voltage drop from grid to cathode is

$$\mathbf{E}_{gk} = \mathbf{I}_2 R_1$$

Then,

$$\mathbf{I}_1(R_3 + r_p) = -\mu\mathbf{E}_{gk} = -\mu\mathbf{I}_2 R_1 \tag{17-20}$$

Kirchhoff's law applied to mesh 3 yields

$$-\mathbf{I}_2 R_2 - \mathbf{I}_1 R_3 = 0$$

or

$$\mathbf{I}_1 R_3 = -\mathbf{I}_2 R_2 \tag{17-21}$$

The ratio of Eq. (17-20) to Eq. (17-21) is

$$\frac{R_3 + r_p}{R_3} = \mu \frac{R_1}{R_2}$$

from which

$$r_p = R_3 \left( \mu \frac{R_1}{R_2} - 1 \right) \tag{17-22}$$

If $\mu$ is first measured, this circuit permits $r_p$ to be calculated. This makes the value of $r_p$ dependent upon the measurement of $\mu$, however. Because of this, this bridge network is used, not to measure $r_p$, but to measure $g_m$ instead. This follows by choosing $R_1$ much greater than $R_2$, so that $\mu R_1/R_2 \gg 1$. Then, approximately, from the last equation,

$$r_p = \mu \frac{R_3 R_1}{R_2}$$

or

$$g_m = \frac{\mu}{r_p} = \frac{R_2}{R_3 R_1} \tag{17-23}$$

A transformer should be used with the telephone receiver, as in the measurement of $\mu$.

The plate resistance $r_p$ of the tube can be directly measured by incorporating the plate circuit of the tube as the fourth arm of a Wheatstone bridge, as shown in Fig. 17-18. When the bridge is balanced

$$r_p = \frac{R_2 R_3}{R_1} \tag{17-24}$$

In order to obtain perfect balance in the bridge circuits of Figs. 17-18, 17-17, and 17-16, it is sometimes necessary to provide means for balancing the capacitive effects of the tube. Basically, however, the circuits are those given above.
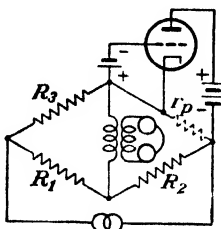


FIG. 17-18. A Wheatstone bridge arrangement for determining the plate resistance of a triode under operating conditions.
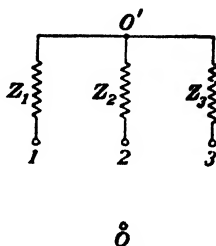
FIG. 17-19. Equation (17-25) applies to this network.

**17-8. Two Network Theorems.** Once the equivalent linear network of a given tube circuit has been drawn, the problem becomes one of network analysis to find either the currents in, or the voltages across, particular elements of the network. This may be done by the use of any of the standard methods of network analysis. For example, best suited to the solution of the particular problem may be a direct application of the mesh or loop method of analysis or an application of the nodal method of analysis. Frequently the application of Thévenin's theorem is of some assistance in the subsequent analysis, and in some cases an application of Norton's theorem simplifies the analysis. A special theorem has been shown by Millman [2] to be extremely useful in the analysis of electron-tube circuits. A generalized form of Norton's theorem also proves to be useful in some problems. These latter two are discussed below as Theorems I and II, respectively.

*Theorem* I. Consider the network of Fig. 17-19. The impedances $Z_1$, $Z_2$, and $Z_3$ terminate in a common junction $0'$. The opposite ends of these impedances are numbered 1, 2, and 3, whereas 0 is any other point in the network. It is not necessary to know the network interconnections between the points 0, 1, 2, and 3. *The voltage drop $E_{0'0}$ from the point $0'$ to the point 0 is given by*

$$E_{0'0} = \frac{E_{10}Y_1 + E_{20}Y_2 + E_{30}Y_3}{Y_1 + Y_2 + Y_3} \qquad (17\text{-}25)$$

*where $E_{s0}$ ($s = 1, 2,$ or 3) is the voltage drop from the sth impedance to the*

point 0 *and where* $Y_s \equiv 1/Z_s$ *is the admittance corresponding to the sth impedance.* In this equation the voltages and impedances are complex numbers, in general.

*Proof.* The voltage drop across $Z_1$ is

$$E_{0'1} = E_{0'0} + E_{01} = E_{0'0} - E_{10}$$

The current through $Z_1$ is, therefore,

$$I_{0'1} = \frac{E_{0'1}}{Z_1} = E_{0'1}Y_1 = (E_{0'0} - E_{10})Y_1$$

Similarly, the currents in the other two branches are given by

$$I_{0'2} = (E_{0'0} - E_{20})Y_2$$

and

$$I_{0'3} = (E_{0'0} - E_{30})Y_3$$

From Kirchhoff's point law, the sum of the three currents $I_{0'1}$, $I_{0'2}$, and $I_{0'3}$ leaving the point 0' must be zero. Then

$$(E_{0'0} - E_{10})Y_1 + (E_{0'0} - E_{20})Y_2 + (E_{0'0} - E_{30})Y_3 = 0$$

from which Eq. (17-25) is readily obtained.

This result is not restricted to three impedances, as shown in Fig. 17-19. The proof given above can easily be generalized to include any number $n$ of impedances terminating in the common junction 0'. The result is

$$E_{0'0} = \frac{\sum_{s=1}^{n} E_{s0}Y_s}{\sum_{s=1}^{n} Y_s} \tag{17-26}$$

where, as usual, $\sum_{s=1}^{n}$ denotes the summation of $n$ terms. Equation (17-25) is a special case of Eq. (17-26) with $n = 3$.

It is emphasized that no restrictions are placed upon the type of network between the points 0, 1, 2, ..., $n$. Any of these points may be connected together with short-circuiting links or through active or passive elements in any manner whatsoever. In order to apply the theorem, all that is necessary is that the voltages $E_{s0}$ be known.

*Theorem* II. *The voltage between two points in a linear network equals the product of the current* I' *which flows in a short circuit placed between these terminals by the impedance* Z' *between these points.* In calculating the impedance, each generator is replaced by its internal impedance, which is zero if the generator impedance may be neglected. This is not the usual form in which Norton's theorem is stated [3] but is extremely useful as herein given.

Several examples will be carried through to show the application of the two theorems.

*Example* 1. A type of network that frequently arises in both power and communication circuits is that of two generators feeding a common load, in the manner shown in Fig. 17-20. The impedances $Z_1$, $Z_2$, and $Z_3$ might be the elements of an equivalent T
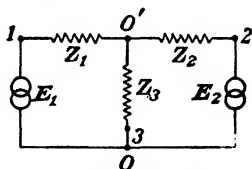


Fig. 17-20. Two generators feeding the common load $Z_3$.

network of a complicated system interconnecting the two generators. Find the current through the load impedance $Z_3$.

*Solution.* The solution will be carried out using each theorem. By Theorem I, the result is given directly by

$$I_{0'0} = \frac{E_{0'0}}{Z_3} = E_{0'0}Y_3$$

where

$$E_{0'0} = \frac{E_1Y_1 + E_2Y_2}{Y_1 + Y_2 + Y_3}$$

$E_{10} = E_1$ and $E_{20} = E_2$ are the potentials of the generators. $E_{30}$ does not appear in this expression since a short-circuiting link connects points 3 and 0.

Note that the individual generator currents are easily found. For example,

$$I_{0'1} = (E_{0'0} - E_1)Y_1$$

The solution according to Theorem II follows. If a short circuit is placed between $0'$ and $0$, the current in this connection from $0'$ to $0$ due to $E_1$ is $E_1/Z_1$ and that due to $E_2$ is $E_2/Z_2$. By introducing the admittances, $Y_s = 1/Z_s$, the total short-circuit current is

$$I' = E_1Y_1 + E_2Y_2$$

With the generators removed from the circuit (each is assumed to be a zero impedance source), the three impedances are in parallel between $0'$ and $0$, and the resultant admittance is the sum of the three admittances. The resultant impedance is

$$Z' = \frac{1}{Y_1 + Y_2 + Y_3}$$

By Theorem II, the voltage from $0'$ to $0$ is $I'Z'$, or

$$E_{0'0} = \frac{E_1Y_1 + E_2Y_2}{Y_1 + Y_2 + Y_3}$$

as before.

It should be recalled that the formal method of solving this problem is to introduce two mesh currents $I_1$ and $I_2$; to write Kirchhoff's law for each branch; and to solve from the currents $I_1$ and $I_2$. The load current is the

sum or difference of these two currents (depending upon the relative directions assigned to $I_1$ and $I_2$).

*Example* 2. Find the voltage drop from $0'$ to $0$ in the network shown in Fig. 17-21. Also find the current in the 5-ohm resistor.
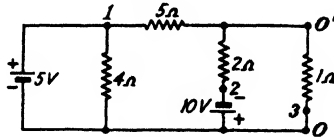


FIG. 17-21. A numerical problem.

*Solution.* The solution is carried out with the use of Theorems I and II. To apply Theorem I it is convenient to write

$$E_{10} = +5 \text{ volts} \qquad E_{20} = -10 \text{ volts} \qquad E_{30} = 0$$
$$Z_1 = 5 \text{ ohms} \qquad Z_2 = 2 \text{ ohms} \qquad Z_3 = 1 \text{ ohm}$$

Then

$$E_{0'0} = \frac{\frac{5}{5} - \frac{10}{2}}{\frac{1}{5} + \frac{1}{2} + \frac{1}{1}} = \frac{-4}{\frac{17}{10}} = -2.35 \text{ volts}$$

But to find the current in the 5-ohm resistor, the voltage $E_{10'}$ is required. This is seen to be

$$E_{10'} = E_{10} + E_{00'} = 5 + 2.35 = 7.35 \text{ volts}$$

Hence the current from 1 to $0'$ is $E_{10'}/5 = 7.35/5 = 1.47$ amp.

To apply Theorem II, the current through a short circuit connected between $0'$ and $0$ is determined.

The current due to the 5-volt generator is 5 volts/5 ohms $= 1$ amp.

The current due to the 10-volt generator is $-10$ volts/2 $= -5$ amp.

The total short-circuit current $I' = 1 - 5 = -4$ amp.

The impedance between $0'$ and $0$ is obtained by considering the network with each battery replaced by its internal resistance. This puts a short circuit across the 4-ohm resistor and yields a parallel group of resistors. The effective resistance is obtained from

$$\frac{1}{Z'} = \frac{1}{5} + \frac{1}{2} + \frac{1}{1} = \frac{17}{10} \text{ mhos}$$

Hence

$$E_{0'0} = I'Z' = -\frac{4}{\frac{17}{10}} = -2.35 \text{ volts}$$

which agrees with the above.

**17-9. Interelectrode Capacitances in a Triode.** It was assumed in the foregoing discussions that with a negative bias on the grid the current taken from the input source was negligible. It was also assumed that changes in the plate circuit were not reflected into the grid circuit. These assumptions are not strictly true, as will now be shown.

The grid, plate, and cathode elements are conductors separated by a dielectric (a vacuum), and hence, by elementary electrostatics, there exist capacitances between pairs of electrodes. Clearly, the input current can-

not be zero because this source must supply current to the grid-cathode capacitance and to the grid-plate capacitance. Furthermore the input and output circuits are no longer isolated, but there is coupling between them through the grid-plate capacitance. Although these capacitances are small, and of the order of 10 $\mu\mu$f, yet at the upper audio frequencies and above they produce appreciable loading of the input source and they also cause output-to-input feedback. They must therefore be taken into account.

A more complete circuit and its equivalent circuit, which includes the interelectrode capacitances, are given in Fig. 17-22. In this circuit, $C_{gp}$
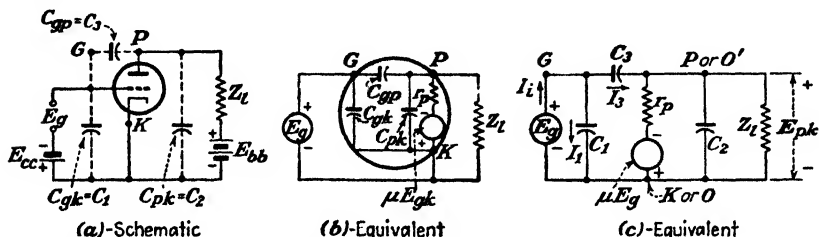


FIG. 17-22. The schematic and equivalent circuits of an amplifier taking into account the interelectrode capacitances. The equivalent circuit (b) shows the elements of the tube, whereas (c) is the more customary form of the equivalent circuit.

represents the capacitance between the grid and the plate; $C_{gk}$ is the capacitance between the grid and the cathode; and $C_{pk}$ is the capacitance between the plate and the cathode. The inclusion of these tube capacitances [shown dotted in the schematic diagram and shown explicitly in the equivalent amplifier circuits of (b) and (c)] yields results that are more precise than those resulting from the analysis of the simple circuit of Fig. 17-10. It will be noted that the same procedure outlined in Sec. 17-6 has been followed in order to obtain the equivalent circuit of the amplifier. It is evident that $\mathbf{E}_{gk} = \mathbf{E}_g$, and so $\mu\mathbf{E}_g$ has been written for the emf of the fictitious generator in the equivalent circuit (c).

If the mesh-current equations for the analysis of this network were set up in the conventional form, their solution would represent a formidable problem. With the aid of the theorems of the foregoing section, however, the output voltage across the load impedance $\mathbf{Z}_l$ can be written down immediately. In order to apply Theorem I, the point 0 in Eq. (17-26) is taken as the cathode terminal $K$, and the point $0'$ as the plate terminal $P$. Four branches ($s = 4$) must be considered between the points $P$ and $K$ of the network. These are the load impedance $\mathbf{Z}_l$ with zero voltage, the capacitance $C_2$ with zero voltage, the voltage rise $\mu\mathbf{E}_g$ in series with $r_p$, and the voltage drop $\mathbf{E}_g$ in series with $C_3$. The capacitance $C_1$ which exists across the input $\mathbf{E}_g$ does not enter the theorem. This follows from the fact

that has already been emphasized that the circuit connections between the points 0, 1, 2, and 3 of Fig. 17-19 need not be specified, it being necessary to know only the voltage drops between these points.

For the present circuit, Eq. (17-26) becomes

$$E_{pk} = \frac{-\mu E_\iota Y_p + E_\iota Y_3}{Y_p + Y_l + Y_2 + Y_3} \tag{17-27}$$

where $Y_p = 1/r_p$ is the admittance corresponding to $r_p$

$Y_2 = j\omega C_2$ is the admittance corresponding to $C_2$

$Y_3 = j\omega C_3$ is the admittance corresponding to $C_3$

$Y_l = 1/Z_l$ is the admittance corresponding to $Z_l$

$E_{pk}$ is the voltage drop across the load impedance, or the output-voltage drop. It follows from this that the complex voltage gain is

$$K = \frac{\text{output-voltage drop}}{\text{input-voltage drop}} = \frac{E_{pk}}{E_\iota} \tag{17-28}$$

which may be written in the form

$$K = \frac{-g_m + Y_3}{Y_p + Y_l + Y_2 + Y_3} \tag{17-29}$$

In this expression, use has been made of the fact that $g_m = \mu/r_p$. The reader should verify that Theorem II leads to exactly this same result.

In this analysis, it has been assumed that no conduction or leakage currents exist between terminals. This leakage current depends upon many variable factors, such as the spacing between electrodes, the type of glass of the envelope, the base of the tube, the conditions of the surface of these, and perhaps the surface leakage between connecting wires. Usually the error is small in assuming infinite resistances (zero conductances) between electrodes. If this assumption is not true, the analysis is still correct provided that the interelectrode admittances are considered to be of the form $Y_\iota = g_\iota + j\omega C_\iota$ instead of merely $j\omega C_\iota$. In addition, interwiring and stray capacitances may be taken into account by considering them to be in parallel with $C_1$, $C_2$, and $C_3$, their major effect being to increase the values of the interelectrode capacitances.

It is interesting to see that Eq. (17-29) reduces to the expression already developed for the case where the interelectrode capacitances are neglected. Under these conditions, $Y_2 = Y_3 = 0$, and Eq. (17-29) reduces to

$$K = \frac{-g_m}{Y_p + Y_l} = \frac{-g_m}{\dfrac{1}{r_p} + \dfrac{1}{Z_l}} = -\mu \frac{1}{1 + \dfrac{r_p}{Z_l}} \tag{17-30}$$

This is a generalization of Eq. (17-10) which applies for the case where the load is $Z_l$. Of course, if $Z_l = R_l$, this reduces to Eq. (17-10).

It is a simple matter to show that the error made in the calculation of the gain is very small when the interelectrode capacitances are neglected for frequencies covering the entire audio-frequency range. These interelectrode capacitances are seldom as large as 15 $\mu\mu f$, which corresponds to an admittance of only about 2 micromhos at 20,000 cps. Since the transconductance $g_m$ of a triode is generally several hundred micromhos or more, $Y_3$ may be neglected in comparison with $g_m$. Furthermore, if $Y_p$ is greater than 100 micromhos ($r_p < 10,000$ ohms), the terms $Y_3 + Y_2$ may be neglected in comparison with $Y_p + Y_l$. Under these conditions the gain is that given by the simple expression (17-30).

Since the interelectrode capacitances have a relatively minor effect on the audio gain of an amplifier, why is it important to make note of them? The answer is to be found in the input impedance of the tube (the loading of the tube on the input circuit) and in the feedback between output and input circuits. Also, if the amplifier is to be used beyond the audio range, say as a video (television or radar) amplifier, then the capacitances may seriously affect the gain and the exact expression, Eq. (17-29), must be used. These effects will be examined.

**17-10. Input Admittance of a Triode.** An inspection of Fig. 17-22 reveals that the grid circuit is no longer isolated from the plate circuit. The input signal must supply a current $I_i$. In order to calculate this current, it is observed from the diagram that

$$I_1 = E_g Y_1$$

and

$$I_3 = E_{gp} Y_3 = (E_g + E_{kp}) Y_3$$

Since, from Eq. (17-28), $E_{kp} = -KE_g$, then the total input current is

$$I_i = I_1 + I_3 = [Y_1 + (1 - K)Y_3]E_g \qquad (17\text{-}31)$$

By definition, the admittance of any circuit element is the ratio of the current through the element to the voltage drop across it in the direction of the current. Consequently, the input admittance is given by

$$Y_i = \frac{I_i}{E_g} = Y_1 + (1 - K)Y_3 \qquad (17\text{-}32)$$

This is the explicit expression for the input admittance of the triode, all factors upon which it depends being clearly indicated. Thus, for the system to possess a negligible input admittance over a wide range of frequencies, the cathode-grid and the grid-plate capacitances must be negligible.

Consider a triode with a pure resistance load. Within the audio-frequency range, the gain is given by the simple expression

$$\mathbf{K} = -\frac{\mu R_l}{r_p + R_l}$$

for the reasons outlined. In this case Eq. (17-32) becomes

$$\mathbf{Y}_i = j\omega \left[ C_1 + \left( 1 + \frac{\mu R_l}{r_p + R_l} \right) C_3 \right] \qquad (17\text{-}33)$$

Thus the input admittance is that arising from the presence of a capacitance from the grid to the cathode of magnitude $C_i$, where

$$C_i = C_1 + \left( 1 + \frac{\mu R_l}{r_p + R_l} \right) C_3 \qquad (17\text{-}34)$$

This increase in input capacitance $C_i$ over the capacitance from grid to cathode $C_1$ is known as the *Miller effect*. The maximum possible value of this expression is $C_1 + (1 + \mu)C_3$, which, for large values of $\mu$, may be considerably larger than any of the interelectrode capacitances. The presence of this input capacitance may prove detrimental in certain circuits.

This input capacitance is important in the operation of cascaded amplifiers. In such a system the output from one tube is used as the input to a second tube. Hence, the input impedance of the second stage acts as a shunt across the load of the first stage. In those cases for which the foregoing is valid, the load is being shunted by the capacitance $C_i$. Since the reactance of a capacitor decreases with increasing frequencies, the resultant output impedance of the first stage will be correspondingly low for the high frequencies. This will result in a decreasing gain at the higher frequencies.

*Example.* A 6SL7 has a load resistance of 100,000 ohms and operates at 20,000 cps at the quiescent point recommended by the manufacturers. Calculate the gain of this tube as a single stage and then as the first tube in a cascaded amplifier consisting of two identical stages.

*Solution.* The pertinent information for the 6SL7 is obtained from a tube data book.

$$g_m = 1,600 \text{ micromhos} \qquad r_p = 44,000 \text{ ohms} \qquad \mu = 70$$

$$C_1 = 3.0 \qquad C_2 = 3.8 \qquad C_3 = 2.8 \ \mu\mu f$$

Then, at 20,000 cps,

$$\mathbf{Y}_1 = j\omega C_1 = j2\pi \times 2 \times 10^4 \times 3.0 \times 10^{-12} = j3.76 \times 10^{-7} \text{ mho}$$

$$\mathbf{Y}_2 = j\omega C_2 = j4.77 \times 10^{-7} \text{ mho}$$

$$\mathbf{Y}_3 = j\omega C_3 = j3.52 \times 10^{-7} \text{ mho}$$

$$Y_p = \frac{1}{r_p} = 2.27 \times 10^{-5} \text{ mho}$$

$$Y_l = \frac{1}{R_l} = 10^{-5} \text{ mho}$$

$$g_m = 1.60 \times 10^{-3} \text{ mho}$$

The gain of a one-stage amplifier is given by Eq. (17-29),

$$K = \frac{-g_m + Y_3}{Y_p + Y_l + Y_2 + Y_3} = \frac{-1.60 \times 10^{-3} + j3.52 \times 10^{-7}}{3.27 \times 10^{-5} + j8.29 \times 10^{-7}}$$

It is seen that the $j$ terms (arising from the interelectrode capacitances) are negligible in comparison with the real terms. If these are neglected, then $K = -49.0$. This can be checked by using Eq. (17-30), which neglects interelectrode capacities. Thus

$$K = \frac{-\mu}{1 + r_p/R_l} = \frac{-70}{1 + 0.44} = -48.7$$

Since the gain is a real number, then the input impedance consists of a capacitor whose value is given by Eq. (17-34),

$$C_i = C_1 + \left(1 + \frac{\mu R_l}{r_p + R_l}\right) C_3 = 3.0 + (1 + 49)(2.8) = 143 \ \mu\mu f$$

Consider now a two-stage amplifier, each stage consisting of a 6SL7 operating as above. The gain of the second stage is that just calculated. However, in calculating the gain of the first stage it must be remembered that *the input impedance of the second stage acts as a shunt on the load resistance of the first stage*. Thus the plate load now consists of 100,000 ohms resistance in parallel with 143 $\mu\mu f$. To this must be added the capacitance from plate to cathode of the first stage since this is also in shunt with the plate load. Furthermore any stray capacitances due to wiring should be taken into account. For example, for every 1 $\mu\mu f$ capacitance between the leads going to the plate and grid of the second stage there is 50 $\mu\mu f$ effectively added across the load resistor of the first tube! This clearly indicates the importance of making connections with very short direct leads in high-frequency amplifiers. Let it be assumed that the input capacitance taking into account the various factors just discussed is 200 $\mu\mu f$ (probably a conservative figure). Then the load admittance is

$$Y_l = \frac{1}{R_l} + j\omega C_i = 10^{-5} + j2\pi \times 2 \times 10^4 \times 200 \times 10^{-12}$$

$$= 10^{-5} + j2.52 \times 10^{-5} \text{ mho}$$

The gain is given by Eq. (17-30),

$$K = \frac{-g_m}{Y_p + Y_l} = \frac{-1.6 \times 10^{-3}}{2.27 \times 10^{-5} + 10^{-5} + j2.52 \times 10^{-5}}$$

$$= -30.7 + j23.7 = 38.8 \underline{/143.3°}$$

Thus, the effect of the capacitances has been to reduce the magnitude of the gain from 49.0 to 38.8 and to change the phase angle between the output and input from 180 to 143.3 deg.

If the frequency were higher, the gain would be reduced still further.   For example, this circuit would be useless as a video amplifier, say to a few megacycles per second, since the gain would then be less than unity.   This variation of gain with frequency is called *frequency* distortion.

If the load circuit of the amplifier is an impedance instead of a pure resistance, then **K** is a complex number in general and the input admittance will consist of two terms, a resistive and a reactive term.   Let **K** be written in the general form

$$\mathbf{K} = k_1 + jk_2 \tag{17-35}$$

Then Eq. (17-32) becomes

$$\mathbf{Y}_i = \omega C_3 k_2 + j\omega[C_1 + (1 - k_1)C_3] \tag{17-36}$$

The expression indicates that the equivalent grid input circuit comprises
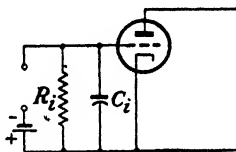


FIG. 17-23.   The equivalent grid input circuit may be represented as a resistance $R_i$ in parallel with a capacitance $C_i$.

a resistance $R_i$ in parallel with a capacitance $C_i$ as shown in Fig. 17-23.

$$\mathbf{Y}_i = \frac{1}{R_i} + j\omega C_i \tag{17-37}$$

It follows from this that the input admittance consists of a resistance

$$R_i = \frac{1}{\omega C_3 k_2} \tag{17-38}$$

in parallel with a capacitance

$$C_i = C_1 + (1 - k_1)C_3$$

Since no restrictions have been placed on the system, it is possible for the term $k_2$ to be negative and the effective input resistance to be negative.   It is interesting to note that an effective negative input resistance is possible only when the load is inductive, with the inductance in a definite range.[4]

The presence of a negative resistance in a circuit can only mean that some power is being generated, rather than being absorbed.   Physically, this means that power is being fed back from the output circuit into the grid circuit through the coupling provided by the grid-plate capacitance. If this feed-back feature reaches an extreme stage, the system will lose its entire utility as an amplifier, becoming in fact a self-excited amplifier, or oscillator.

*Example.*  Calculate the input admittance of a 6J5 triode working into a load consisting of a 25-millihenry coil whose resistance is 2,000 ohms.  Assume that the tube is operated under recommended conditions and at a frequency of 10,000 cps.

*Solution.*  The pertinent information for the 6J5 triode is obtained from a tube data book.

$$g_m = 2,600 \text{ micromhos} \qquad r_p = 7,700 \text{ ohms} \qquad \mu = 20$$

$$C_1 = 3.4 \qquad C_2 = 3.6 \qquad C_3 = 3.4 \ \mu\mu\text{f}$$

At 10,000 cps,

$$Y_1 = Y_3 = j\omega C_1 = +j2.13 \times 10^{-7} \text{ mho}$$

$$Y_2 = j\omega C_2 = j2.26 \times 10^{-7} \text{ mho}$$

$$Y_p = \frac{1}{r_p} = 1.3 \times 10^{-4} \text{ mho}$$

$$Z_l = R_L + j\omega L = 2,000 + j1,570 \text{ ohms}$$

$$Y_l = \frac{1}{Z_l} = (3.08 - j2.42) \times 10^{-4} \text{ mho}$$

All wiring capacitances are neglected in this example.  The gain of the amplifier is

$$K = -\frac{g_m}{Y_p + Y_l} = \frac{-26 \times 10^{-4}}{1.3 \times 10^{-4} + (3.08 - j2.42)10^{-4}} = -4.55 - j2.51$$

whence the input admittance becomes

$$Y_i = Y_1 + (1 - K)Y_3 = (-5.35 + j13.9) \times 10^{-7}$$

If the input circuit is supposed to consist of a resistance and capacitance in parallel, the constants are, according to Eq. (17-38);

$$R_i = \frac{1}{-5.35 \times 10^{-7}} = -1.87 \text{ megohms}$$

$$C_i = \frac{13.9 \times 10^{-7}}{2\pi \times 10^4} \text{ f} = 22.2 \ \mu\mu\text{f}$$

## 17-11. Second Harmonic Distortion in Triodes.

The method of construction by which the output wave shape for any input wave-shape signal could be obtained graphically from the plate characteristics was discussed in Sec. 17-2.  An inspection of Fig. 17-6 shows that the application of a sinusoidal input voltage to a linear dynamic curve results in an output wave the wave shape of which is also sinusoidal.  This is, in fact, the approximation that is employed in the equivalent-circuit method of analysis.

In general, however, the dynamic characteristic will not be a straight line but will contain a slight curvature as seen in the oscillogram of Fig. 17-24.  The amount of this curvature decreases with increasing values of load resistance.  This nonlinearity arises because the $i_b$-$e_b$ static characteristics are not equidistant lines for constant $e_c$ intervals in the range 1-2 of Fig. 17-3.  Because of this nonlinear characteristic of the dynamic curve over the operating range, the wave form of the output wave differs slightly

from that of the grid-exciting-voltage wave shape. Distortion of this type is called "nonlinear" or "amplitude" distortion.

In order to investigate the magnitude of the distortion that results from the nonlinearity of the dynamic curve, it will be assumed that the dynamic curve of a triode with respect to the point $Q$ can be represented by a para-

bolic rather than by a linear curve. Thus, instead of relating the alternating plate current $i_p$ with the grid voltage $e_g$ by the relation

$$i_p = \left(\frac{\mu}{r_p + R_l}\right) e_g = A e_g \qquad (17\text{-}39)$$

that arises from the equivalent circuit of the triode, it is assumed that the relationship between $i_p$ and $e_g$ is given more accurately by the expression

$$i_p = A_1 e_g + A_2 e_g{}^2 \qquad (17\text{-}40)$$

where the $A$'s are constants. Actually, these are simply the first two terms of a power series expansion of $i_p$ as a function of $e_g$.

If the input potential is sinusoidal and of the form

$$e_g = E_{gm} \cos \omega t$$

the substitution of this expression in (17-40) leads to

$$i_p = A_1 E_{gm} \cos \omega t + A_2 E_{gm}{}^2 \cos^2 \omega t$$

Since $\cos^2 \omega t = \frac{1}{2} + \frac{1}{2} \cos 2\omega t$, the expression for the instantaneous total current reduces to the form

$$i_b = I_b + i_p = I_b + B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t \qquad (17\text{-}41)$$

The physical meaning of this equation is evident. It shows that the application of a sinusoidal signal on a parabolic dynamic characteristic will result in an output current that contains, in addition to a term of the same frequency as the input, a second harmonic term and also a constant term. The constant term, which appears in Eq. (17-41) adds to the original d-c value $I_b$, yielding a total d-c component of current $I_b + B_0$ in the output.

This term arises from the partial rectification that takes place as a consequence of the curvature of the dynamic curve.

The amplitudes $B_0$, $B_1$, and $B_2$ for a given load resistor are readily determined from either the static or the dynamic characteristics of the particu-
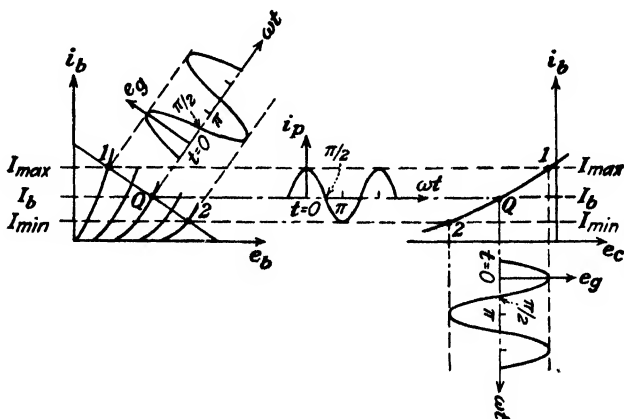


FIG. 17-25.   The output current of an amplifier has zero-axis symmetry provided only that the input is assumed to be of the form $e_g = E_{gm} \cos \omega t$.

lar tube circuit.   This allows the amount of second harmonic distortion that exists to be calculated.   It is observed from Fig. 17-25 that when

$$\omega t = 0, \qquad i_b = I_{\max}$$

$$\omega t = \frac{\pi}{2}, \qquad i_b = I_b$$

$$\omega t = \pi, \qquad i_b = I_{\min}$$

By substituting these values in Eq. (17-41) there results

$$I_{\max} = I_b + B_0 + B_1 + B_2$$

$$I_b = I_b + B_0 - B_2$$

$$I_{\min} = I_b + B_0 - B_1 + B_2$$

This is a set of three equations for the three unknowns $B_0$, $B_1$, and $B_2$.   It follows from the second of this group that

$$B_0 = B_2 \qquad\qquad (17\text{-}42)$$

By subtracting the third equation from the first, there results

$$B_1 = \frac{I_{\max} - I_{\min}}{2} \qquad\qquad (17\text{-}43)$$

With this value of $B_1$, the value for $B_2$ may be evaluated from either the first or the last equation. This yields

$$B_2 = B_0 = \frac{I_{\max} + I_{\min} - 2I_b}{4} \qquad (17\text{-}44)$$

The percentage of second harmonic distortion is defined as

$$D_2 \equiv \frac{100|B_2|}{|B_1|} \quad \% \qquad (17\text{-}45)$$

The quantities $I_{\max}$, $I_{\min}$, and $I_b$ appearing in these equations are obtained directly from the characteristic curves and the load line.

*Example.* Calculate the d-c current, the fundamental gain, and the second harmonic distortion for a 6C5 triode operating at a bias of $-8$ volts, a plate supply of 360 volts, and a load resistance of 6,000 ohms. The input signal has a peak value of 6 volts.

*Solution.* The maximum current corresponds to a grid voltage of $-8 + 6 = -2$ volts and the minimum current to a bias of $-8 - 6 = -14$ volts. This is the same amplifier, and the same operating conditions, as that discussed in the example in Sec. 17-2. The values found in that section are

$$I_{\max} = 21.5 \qquad I_b = 12 \qquad I_{\min} = 4.8 \text{ ma}$$

Hence

$$B_2 = B_0 = \frac{I_{\max} + I_{\min} - 2I_b}{4} = \frac{21.5 + 4.8 - 24}{4} \doteq 0.6 \text{ ma}$$

This represents the current rectification due to the nonlinearity of the tube characteristics. The total d-c current is $I_b + B_0 = 12 + 0.6 = 12.6$.

The peak fundamental current is

$$B_1 = \frac{I_{\max} - I_{\min}}{2} = \frac{21.5 - 4.8}{2} = 8.35 \text{ ma}$$

The peak fundamental output voltage is this value multiplied by the 6,000-ohm load resistance, or 50 volts. To obtain the fundamental gain this must be divided by the peak input-voltage swing of 6 volts, or the gain is $\frac{50}{6} = 8.3$. This is the same result as obtained in Sec. 17-2, and it should now be clear that it was the *fundamental* voltage gain which was calculated previously.

The per cent of second harmonic distortion is

$$D_2 = \frac{100B_2}{B_1} = \frac{(100)(0.6)}{8.35} = 7.2\%$$

**17-12. Harmonic Generation in a Tube.** The analysis in the foregoing section indicates that the application of a sinusoidal voltage to the input of a tube circuit, the dynamic curve of which may be represented as a parabola about the quiescent point, results in the presence of a d-c component and a second harmonic term in addition to a term of fundamental frequency. This approximation is usually adequate to represent normal

operation of such a tube. It is inadequate to represent the conditions when the triode may operate with a positive grid potential over a portion of its swing. It is also invalid when the tube is operated with such a bias that the very curved portions of the plate characteristics must be employed. It is likewise true that the parabolic approximation may not be sufficiently accurate to describe the behavior of tetrodes and pentodes as circuit elements.

It is frequently necessary, therefore, to express the dynamic curve with respect to the $Q$ point by a power series of the form

$$i_p = A_1 e_g + A_2 e_g{}^2 + A_3 e_g{}^3 + A_4 e_g{}^4 + \cdots \tag{17-46}$$

If it is assumed, as above, that the input wave is a simple cosine function of time, of the form

$$e_g = E_{gm} \cos \omega t \tag{17-47}$$

then the output current may be shown to have the form

$$i_b = I_b + B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \cdots \tag{17-48}$$

This equation results when Eq. (17-47) is inserted in Eq. (17-46) and the proper trigonometric transformations are made.

That the output-current wave form must be expressible by a relationship of this form is made evident from an inspection of Fig. 17-25. It is observed from this figure that the plate-current curve must possess zero-axis symmetry. This means that the current is an *even* function of time, or expressed mathematically $i(\omega t) = i(-\omega t)$. Physically, it means that the wave shape for every quarter cycle of the plate-current curve as the operating point moves from the point $Q$ to the point 1 is similar to the shape of the curve that is obtained as the operating point moves back from the point 1 to the point $Q$. Similarly, the wave shape of the current generated by the operating point as it moves from the point $Q$ to the point 2 is symmetrical with that generated as it moves from the point 2 back to the point $Q$. These conditions are true regardless of the curvature of the characteristics. Since $i$ is an even function of time, the Fourier series representing a periodic function possessing this symmetry contains only cosine terms. (If any sine terms were present they would destroy the symmetry since they are *odd* and not *even* functions of time. A mathematical proof is given in Sec. 14-2.)

If in Eq. (17-46) one assumes, as is frequently done in the literature, that the excitation voltage is a sine instead of a cosine function of the time, the resulting output current is no longer expressed by a series of cosine terms only. Though a sine function differs from a cosine function only in the shift of the axis by an amount $\omega t = \pi/2$, nevertheless such a shift destroys the above-noted zero-axis symmetry. It is found in this case that

the Fourier series representing the output current contains odd sine components and even cosine components.

Any one of a number of methods [5] may be used in order to obtain the coefficients $B_0$, $B_1$, $B_2$, etc. These include the standard procedures employed in conventional Fourier analysis, *viz.*, the schedule method or the Fischer-Hinnen average selected ordinate method. The method due to Espley, which is a combination of the two standard procedures and which is simply an extension of the method of procedure of the last section, will be described here. It was assumed in the foregoing section that only three terms, $B_0$, $B_1$, and $B_2$, of the Fourier series were different from zero. These three values were evaluated in terms of three measured values of current, $I_{max}$, $I_{min}$, and $I_b$. As the next approximation, it will be assumed that only five terms, $B_0$, $B_1$, $B_2$, $B_3$, $B_4$, exist in the resulting Fourier series. In order to evaluate these five coefficients, the values of the currents at five different values of $e_g$ are needed. These are chosen at equal intervals in the grid swing. Thus, $I_{max}$, $I_{\frac{1}{2}}$, $I_b$, $I_{-\frac{1}{2}}$, and $I_{min}$ correspond, respectively, to the following values of $e_g$: maximum positive value, $\frac{1}{2}$ the maximum positive value, zero, $\frac{1}{2}$ the maximum negative value, and the maximum negative value. These values are illustrated in Fig. 17-26.

FIG. 17-26. The values of signal voltage and the corresponding values of plate current that are used in the five-point schedule for determining the Fourier components $B_0$, $B_1$, $B_2$, $B_3$, and $B_4$ of the current.

By assuming that the grid voltage has the form

$$e_g = E_{gm} \cos \omega t$$

as illustrated, then when

$$\omega t = 0, \qquad i_b = I_{max}$$

$$\omega t = \frac{\pi}{3}, \qquad i_b = I_{\frac{1}{2}}$$

$$\omega t = \frac{\pi}{2}, \qquad i_b = I_b$$

$$\omega t = \frac{2\pi}{3}, \qquad i_b = I_{-\frac{1}{2}}$$

$$\omega t = \pi, \qquad i_b = I_{min}$$

By combining these conditions with Eq. (17-46), five equations containing five unknowns are obtained. The solution of these equations yields

$$
\left.
\begin{aligned}
B_0 &= \tfrac{1}{6}(I_{\max} + 2I_{\frac{1}{2}} + 2I_{-\frac{1}{2}} + I_{\min}) - I_b \\
B_1 &= \tfrac{1}{3}(I_{\max} + I_{\frac{1}{2}} - I_{-\frac{1}{2}} - I_{\min}) \\
B_2 &= \tfrac{1}{4}(I_{\max} - 2I_b + I_{\min}) \\
B_3 &= \tfrac{1}{6}(I_{\max} - 2I_{\frac{1}{2}} + 2I_{-\frac{1}{2}} - I_{\min}) \\
B_4 &= \tfrac{1}{12}(I_{\max} - 4I_{\frac{1}{2}} + 6I_b - 4I_{-\frac{1}{2}} + I_{\min})
\end{aligned}
\right\}
\qquad (17\text{-}49)
$$

This method might be termed a five-point schedule. Higher harmonics can be obtained by employing a seven-point, a nine-point, etc., schedule.

If a large number of determinations of harmonic distortion are to be made, it might be advantageous to use specially constructed direct reading scales or certain trigonometric simplifications that are described in the literature.[6]

The percentage of harmonic distortion is defined as

$$
D_2 = \frac{100|B_2|}{|B_1|} \qquad D_3 = \frac{100|B_3|}{|B_1|} \qquad D_4 = \frac{100|B_4|}{|B_1|} \qquad \% \quad (17\text{-}50)
$$

where $D_s(s = 2, 3, 4, \ldots)$ represents the per cent distortion of the $s$th harmonic, and the *total distortion* or *distortion factor* is defined as

$$
D = \sqrt{D_2{}^2 + D_3{}^2 + D_4{}^2 + \cdots} \qquad (17\text{-}51)
$$

An example illustrating how these equations may be used to determine the distortion in multielectrode tubes is given in Sec. 18-9.

**17-13. Load Curve for a Reactive Load.** A graphical method of obtaining the operating characteristics of a triode with a resistance load is given in Sec. 17-2. It is there shown that the operating region in the $i_b$-$e_b$ plane is a straight line, which is called the "load line." However, if the load is reactive, then the work curve is no longer a straight line but attains the form of an ellipse. This follows from the fact that the plate current and the plate voltage are given by

$$
e_p = E_{pm} \sin \omega t \qquad \text{and} \qquad i_p = -I_{pm} \sin (\omega t + \theta) \qquad (17\text{-}52)
$$

which are the parametric equations of an ellipse. If the angle $\theta$ is zero, the ratio of these equations yields

$$
\frac{e_p}{i_p} = -\frac{E_{pm}}{I_{pm}} = -R_l
$$

which represents the load line for a resistance load. This load line and also the elliptical work curve for a reactive load are shown on the volt-ampere characteristics of Fig. 17-27.

The equivalent-circuit concepts introduced in Secs. 17-3 and 17-5 depend upon the tube parameters $\mu$, $r_p$, and $g_m$ being constant over the range of operation in the $i_b$-$e_b$ plane. Then, subject to the same limitation on the tube parameters, the equivalent-circuit method of analysis is valid when the work curve is an ellipse. If these parameters are not constant, the operating curve will no longer be an ellipse. No simple analysis of the output of an amplifier with a reactive load under these conditions exists.
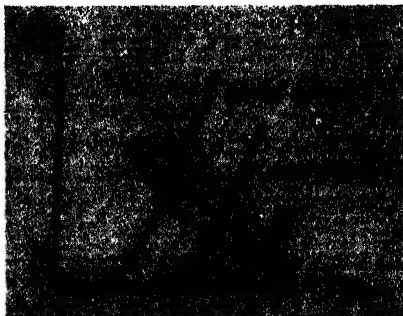


Fig. 17-27. Oscillogram of the plate characteristics of a triode, of the load line for a pure resistance load, and of the load curve (an ellipse) for a reactive load.

A graphical method is given by Preisman.[7] Also, no simple analysis exists that permits a calculation of the nonlinear harmonic distortion introduced in the output of a tube with a reactive load.

## PROBLEMS

**17-1.** An adjustable resistor $R_l$ is connected in series with the plate of a 6C5 triode and a plate supply of 240 volts. The grid is maintained at $-4$ volts with respect to the cathode. Determine by graphical methods, using the plate characteristics in Appendix IX,

a. The plate current, when $R_l$ has the following values: 0, 10, 50 kilohms.

b. The plate voltage corresponding to the resistances in part a.

c. The load resistance that will give a plate current of 10 ma.

d. The voltage drop across $R_l$ if the bias is changed to $-8$ volts and if the load resistance is 10 kilohms.

**17-2.** A 6SF5 triode is operated at the quiescent point $E_b = 240$ volts, $E_c = -2$ volts. The plate characteristics of this tube are given in Appendix IX.

a. Calculate the plate supply voltage that must be used for the following values of load resistance: 0.1, 0.2, and 0.3 megohms.

b. If the grid excitation is sinusoidal with a peak value of 2 volts, find the maximum and minimum currents obtainable with each of the load resistances of part a.

c. Plot the dynamic characteristic for $R_l = 0.2$ megohm.

**17-3.** Figure 17-11, which shows the magnitude of the gain $K$ of an amplifier as a function of load resistance $R_l$, is based upon the assumption that $\mu$ and $r_p$ remain constant as $R_l$ is varied. Suppose that instead of operating the tube at a fixed $Q$ point the plate

and grid supply voltages are maintained constant as $R_l$ is varied. The operating point and hence $\mu$ and $r_p$ will now be functions of $R_l$.

a. Explain clearly the effect this will have upon the $K$ vs. $R_l$ curve.

b. Plot the $K$ vs. $R_l$ curve for a 6C5 tube operating under the conditions $E_{bb}$ = 400 volts and $E_{cc}$ = −8 volts. Choose values of load resistance that correspond to quiescent currents of from 1 to 16 ma. Draw Fig. 17-11 on the same graph sheet, and compare the two curves. The variation of $\mu$ and $r_p$ as a function of load current are given in Fig. 16-11 for a 6C5. The plate characteristics of this tube are given in Appendix IX.

**17-4.** The plate current of a certain triode can be represented by the equation

$$i_b = 0.002(e_c + 0.1e_b)^{\frac{3}{2}} \text{ amp}$$

a. Find the plate resistance when the plate voltage is 200 volts and the grid bias is −15 volts.

b. The plate load consists of a series circuit made up of a resistance of 10 kilohms and an inductance of 2.0 henrys (of negligible resistance). What must be the value of the plate supply voltage in order that it operate under the conditions of part a?

c. If a 1-volt maximum 1,000-cps signal is impressed on the amplifier, calculate the output voltage across the inductor.

d. Calculate the phase angle between input and output voltages. Draw a sinor diagram.

**17-5.** a. Find an expression for the gain of the circuit shown. Prove that it is always less than unity. Show that the cathode voltage is in phase with the grid voltage (with respect to ground as the reference point). Neglect interelectrode capacities. This circuit is called a *cathode-follower* or a *cathode-coupled* stage.

b. Find an expression for the cathode-grid voltage.
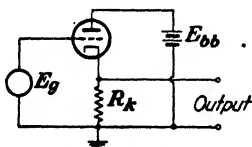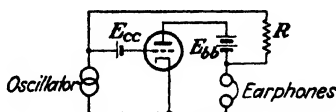


PROB. 17-5.



PROB. 17-6.

**17-6.** The circuit shown can be used to measure the transconductance of a triode. Prove that, when a balance is obtained, $g_m = 1/R$ (see Ref. 1 of this chapter).

**17-7.** A triode is operated as shown. Calculate the voltage gain at 1,000 cycles. Draw a sinor diagram. $\mu$ = 20, and $r_p$ = 10 kilohms.
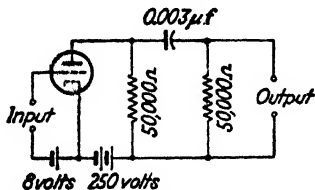


PROB. 17-7.
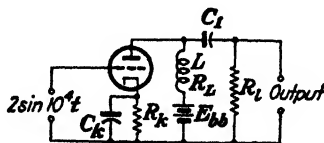


PROB. 17-8.

**17-8.** A triode is used in the circuit shown. The circuit parameters have the following values: $L$ = 10 henrys, $R_L$ = 0, $C_1$ = 0.10 $\mu$f, $C_k$ = 10 $\mu$f, $R_l$ = 10 kilohms, $E_{bb}$ = 300 volts, $\mu$ = 50, $r_p$ = 50 kilohms. It is desired to have a d-c current of 1.0 ma in the

plate circuit and that the grid voltage be $-3$ volts when the applied signal is zero. Determine the following:

   *a.* Self-bias resistor $R_k$.

   *b.* Voltage gain of the amplifier.

   *c.* Phase shift between input and output voltage. Draw a sinor diagram.

   Note that no grid battery is used in this circuit but that the voltage drop $I_b R_k$ acts as the grid bias.

**17-9.** The triode shown has a plate resistance of 2.5 kilohms and an amplification factor of 5. If the rms-reading voltmeter $V$ has a resistance of 10 kilohms and negligible reactance, what will it read? The input signal $E$ is 12 volts rms at a frequency of 1,000 cps. The reactance of the capacitor $C$ may be neglected in comparison with the voltmeter resistance.
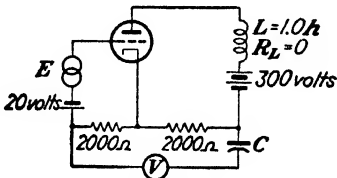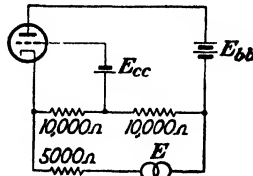


PROB. 17-9.           PROB. 17-10.

**17-10.** Draw the equivalent circuit and find the a-c plate current in the tube shown. The tube constants are $\mu = 10$ and $r_p = 5$ kilohms. The 1,000-cycle oscillator $E$ has an rms output of 0.2 volt.

**17-11.** Analyze the circuit shown, which is known as a *bootstrap* circuit. This provides full gain but without the 180-deg phase reversal of a conventional amplifier. It has the disadvantage of requiring isolation of the input-signal source from ground.
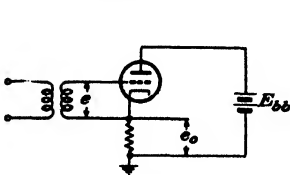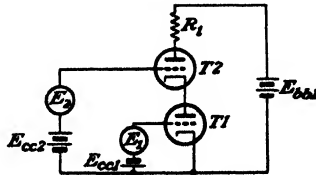


PROB. 17-11.           PROB. 17-12.

**17-12.** Find an expression for the a-c output across $R_l$. The two tubes are identical and have parameters $\mu$, $r_p$, and $g_m$.

**17-13.** Find the currents in each of the resistors. Use the method of Sec. 17-8.
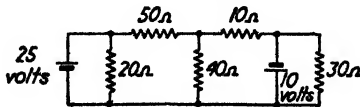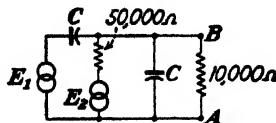


PROB. 17-13.           PROB. 17-14.

**17-14.** The alternator voltages are each 10 volts rms. The reactance of each capacitor is 100 kilohms. Find the voltage across the points $AB$ if the two alternators are

   *a.* In phase.

   *b.* 180 deg out of phase.

**17-15.** *a.* A 6SF5 triode works into a 100-kilohm resistive load. Calculate the complex voltage gain and the input admittance of the system for frequencies of 100 and 100,000 cps. Take the interelectrode capacitances (see Appendix IX) into account, and assume that the tube is operated under recommended conditions.

Compare these results with those which are obtained when the interelectrode capacitances are neglected.

*b.* Calculate the input resistance and capacitance.

**17-16.** Calculate the input admittance of a 6C5 triode at $10^3$ and $10^6$ cps when the total plate circuit impedance is

*a.* A resistance of 50 kilohms.

*b.* An inductive reactance of 50 kilohms at each frequency.

Take the interelectrode capacitances (see Appendix IX) into consideration, and assume that the tube is operated under recommended conditions. Express the results in terms of the input resistance and capacitance.

**17-17.** A 6C5 is incorporated in a simple amplifier circuit. It is to be operated at the recommended point. Starting with a zero load, how much resistance must be introduced as load to increase the input capacitance by a factor of 5?

**17-18.** In the circuit shown the triode is used as an adjustable impedance element, by adjusting the d-c bias.
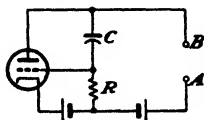


PROB. 17-18.

*a.* Assume that there is a generator $E$ between the terminals $A$ and $B$. Draw the equivalent circuit. Neglect interelectrode capacitances.

*b.* Show that the input admittance between $A$ and $B$ is

$$Y_i = Y_p + (1 + g_m R)Y_{CR}$$

where $Y_p$ is the admittance corresponding to $r_p$ and $Y_{CR}$ is the admittance corresponding to $R$ and $C$ in series.

*c.* If $g_m R \gg 1$, show that the effective input capacitance is

$$C_i = \frac{g_m \alpha}{\omega(1 + \alpha^2)}$$

and the effective input resistance is

$$R_i = \frac{(1 + \alpha^2)r_p}{1 + \alpha^2(1 + \mu)}$$

where $\alpha = \omega CR$.

*d.* At a given frequency show that the maximum value of $C_i$ (as either $C$ or $R$ is varied) is obtained when $\alpha = 1$ and

$$(C_i)_{\text{max}} = \frac{g_m}{2\omega}$$

Also show that the value of $R_i$ corresponding to this $C_i$ is

$$(R_i)_{\text{max}} = \frac{2r_p}{2 + \mu}$$

which, for $\mu \gg 2$, reduces to $(R_i)_{\text{max}} = 2/g_m$.

*e.* The tube is a 6C5 (see Fig. 16-11) operated at 5,000 cps. If the bias is adjusted so that the tube current can be varied over the range from 2 to 20 ma, over what range do $(C_i)_{\text{max}}$ and $(R_i)_{\text{max}}$ vary?

**17-19.** Solve for the analogous quantities to those asked for in Prob. 17-18, if the capacitor $C$ is replaced by an inductor $L$.

**17-20.** A type 6F6 tube is operated as a triode with a load resistance of 2.5 kilohms and a plate supply of 280 volts. The grid bias is $-20$ volts, and the peak grid signal is 20 volts.

  *a.* What is the fundamental current output?

  *b.* What is the percentage second harmonic distortion?

  *c.* What is the d-c current?

  Use the plate characteristics in Appendix IX.

**17-21.** A 6SN7 (whose characteristics are given in Fig. A9-10) is to be operated from a $B$-supply voltage of 240 volts. A gain of approximately 10 is desired. The peak-to-peak 2,000-cycle sinusoidal input voltage is 12 volts. If the grid is never to swing positively, specify

  *a.* The load resistance.

  *b.* The bias voltage.

  *c.* The quiescent current.

  *d.* The per cent second harmonic distortion.

  *e.* The fundamental gain.

**17-22.** A type 6A3 triode supplies power to a 4-kilohm load from a 300-volt source of power. The bias voltage is set at $-40$ volts, and a 40-volt-peak sinusoidal signal is applied. Using the plate characteristics in Appendix IX, plot

  *a.* One cycle of the output plate current as a function of the time for the sinusoidal input grid signal.

  *b.* On the same curve sheet, plot the calculated output curve on the assumption that the output will consist of the fundamental and a second harmonic distortion component. The magnitude of the fundamental and second harmonic terms must be found first.

**17-23.** A 6A3 triode (plate characteristics are given in Fig. A9-2) is operated as a simple power amplifier with a load resistance of 800 ohms. The grid bias is $-45$ volts.

  *a.* If the quiescent current is to be 60 ma, what must be the plate supply voltage?

  *b.* If a 1,000-cycle signal whose peak value is 4.0 volts is applied to the grid, what will be the peak value of the output voltage?

  *c.* If a 1,000-cycle signal whose peak value is 40.0 volts is applied to the grid, what will be the peak value of the 1,000-cycle output voltage?

  *d.* Under the conditions of part *c*, what will be the peak value of the 2,000-cycle component of the output voltage?

  *e.* Under the conditions of part *c*, what will be the d-c component of the plate current?

**17-24.** A 6F6 triode supplies 0.85 watt to a 4-kilohm load. The zero-signal d-c plate current is 31 ma, and the d-c plate current with signal is 34 ma. Determine the per cent second harmonic distortion.

**17-25.** The grid excitation of an amplifier is $e_g = \sqrt{2}E_g \sin \omega t$. Prove that the output current can be represented by a Fourier series which contains only odd sine components and even cosine components.

**17-26.** Supply the missing steps in the derivation of Eq. (17-49).

**17-27.** Obtain a seven-point schedule by the Espley method. Determine $B_0$, $B_1$, $B_2$, $B_3$, $B_4$, $B_5$, and $B_6$ in terms of $I_{max}$, $I_{\frac{1}{2}}$, $I_{\frac{1}{4}}$, $I_b$, $I_{-\frac{1}{4}}$, $I_{-\frac{1}{2}}$, and $I_{min}$.

**17-28.** Analyze the output current from an ideal half-wave vacuum rectifier by the five-point schedule. Compare the coefficients obtained from this analysis with the first five terms of the exact solution. The input must be assumed to be a cosine and not a sine function of the time.

**17-29.** Obtain a five-point schedule for determining $B_0$, $B_1$, $B_2$, $B_3$, and $B_4$ in terms of $I_{max}$, $I_{0.707}$, $I_b$, $I_{-0.707}$, and $I_{min}$.

**17-30.** Show that the parametric equations (17-52) are those of an ellipse.

**17-31.** A PJ-22 phototube and a 6C5 amplifier are used in the circuit shown for counting certain objects. When no object is in the path of the incident beam, the intensity on the cell is 100 ft-c. Under this condition, it is desired that there be no current in the relay. When an object interrupts the light beam, the light intensity on the cell is 5 ft-c and it is desired that the relay should close. If the relay closes at 20 ma, for what range of values of $R$ will the circuit operate properly? The relay resistance is 1 kilohm.
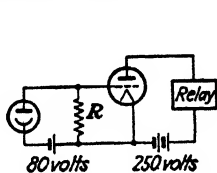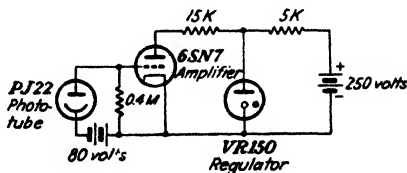


PROB. 17-31.          PROB. 17-32.

**17-32.** In the circuit shown the illumination on the phototube is 50 ft-c. Calculate the current in each resistor and the current in the VR tube.

**17-33.** In the circuit shown the illumination on the phototube controls the d-c current in the FG-27A thyratron load $R_l$. Find the d-c current in $R_l$ when the illumination is 100 ft-c.
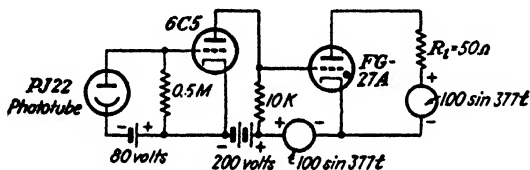


PROB. 17-33.

**17-34.** In the figure shown the phototube controls the load current of the triode. As used in this circuit, the phototube current $I_b$ as a function of the intensity $j$ (in foot-candles) is given by the equation

$$I_b = \frac{j}{10}$$

where $I_b$ is in microamperes.

a. Find the minimum illumination that will send current through the 15-kilohm load.

b. If the illumination is fixed at 150 ft-c, what is the peak instantaneous load current?
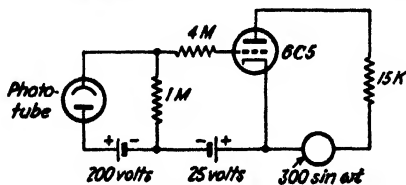


PROB. 17-34.

c. Make a rough plot of the a-c input voltage as a function of time, and (using the same abscissa) make a rough plot of the instantaneous load voltage for the fixed illumination of 150 ft-c. In particular, indicate where the output voltage drops to zero.

**17-35.** Design a photocell circuit that will close a relay if the illumination exceeds 0.1 lumen and will open the relay if the illumination falls below 0.02 lumen. A type 929 photocell, a type 6C5 amplifier, and a 1-kilohm relay that closes at 10 ma and opens at 2 ma are to be used. The voltage source for each tube is the 120-volt d-c house supply.

    *a.* Sketch the circuit.

    *b.* Calculate the resistance needed to couple the phototube to the amplifier.

    *c.* Calculate the grid bias that must be used.

## REFERENCES

1. Standards Report on Electronics, Institute of Radio Engineers, New York, N.Y., 1938.
2. MILLMAN, J., *Proc. IRE*, **28**, 413, 1940.
3. EVERITT, W. L., "Communication Networks," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937.
   LePAGE, W. R., and S. SEELY, "General Network Analysis," McGraw-Hill Book Company, Inc., New York, 1952.
4. MILLMAN, J., and S. SEELY, "Electronics," 1st ed., p. 536, McGraw-Hill Book Company, Inc., New York, 1941.
5. ESPLEY, D. C., *Proc. IRE*, **21**, 1439, 1933.
   CHAFFEE, E. L., *Rev. Sci. Instruments*, **7**, 384, 1936.
6. ESPLEY, D. C., and L. I. FARREN, *Wireless Eng.*, **11**, 183, 1934.
   BLOCH, A., *ibid.*, **16**, 592, 1939.
7. PREISMAN, A., "Graphical Constructions for Vacuum Tube Circuits," McGraw-Hill Book Company, Inc., New York, 1943.

### General References

For general references, see end of Chap. 16.

# CHAPTER 18

# MULTIELECTRODE TUBES

THE analysis in the previous chapter shows that the grid-plate interelectrode capacitance contributes the major portion of the input capacitance of a triode. Moreover, it is this capacitance that causes the feedback from the plate circuit to the grid circuit. Clearly, any reduction of this quantity is most desirable. The introduction of additional electrodes in the tube accomplishes this desired result, although the characteristics of the tube are markedly changed thereby.

**18-1. Screen-grid Tubes or Tetrodes.** The screen-grid tube was introduced commercially about 1928. In these tubes a fourth electrode is interposed between the grid and the plate. This new electrode is known as the screen grid, the shield grid, or grid 2, in order to distinguish it from the "control" grid of the three-electrode tube. It almost entirely encloses the plate, as shown schematically in Fig. 18-1.[1] Because of its design and disposition, the screen grid affords very complete electrostatic shielding between the plate and the grid. This shielding is such that the grid-plate capacitance is reduced by a factor of about 1,000 or more. However, the screen mesh is sufficiently coarse so that it does not interfere appreciably with the flow of electrons.

Fig. 18-1. Schematic arrangement of the electrodes in a screen-grid tetrode. The tube has cylindrical symmetry.

Because of the shielding action of the plate by the screen grid, the electric field produced in the neighborhood of the cathode by the anode potential is practically zero. Since, however, the total space current is determined almost wholly by the field near the cathode surface, the plate exerts little or no effect on the total space charge drawn from the cathode. There are, therefore, several significant differences between triodes and tetrodes. The plate in a triode performs two distinct functions, that of controlling the total space current, and that of collecting the plate current. In a tetrode, the plate serves only to collect those electrons which succeed in passing through the screen.

539

This passive character of the plate makes the tetrode a much better voltage amplifier than the triode. Physically, this follows from the fact that in a three-element tube an increase in signal voltage causes an increase in load current. This increased current causes a decreased plate-cathode potential, because of the increased $i_bR_l$ drop. The decreased plate-cathode potential reduces the space current below its value were there no such reduction in the plate-cathode voltage. Although the $i_bR_l$ drop still exists in the tetrode, no consequent decrease in space current occurs because the plate is effectively isolated by the screen. Hence the input signal must necessarily have a much greater effect on the output voltage in a tetrode than in a triode.

This reasoning is valid, of course, only if the change in plate current for a given change in control-grid voltage in both tubes is the same. That is, this discussion presumes that the plate-grid transconductance is the same for the two tubes. In most cases, the disposition of the cathode and grid is almost the same in the three- and four-element tubes, and the control of the electron stream by the grid is nearly alike for both types of tube.

Since changes in plate voltage have very little effect upon the plate current, it follows that the plate resistance of tetrodes must be very high. Correspondingly, the amplification factor of the tube must also be high. This follows from the definition of the amplification factor, which measures the relative effectiveness of changes of plate and grid voltage in producing equal plate-current increments. Quantitative definitions of the tube parameters of the tetrode will be given in Sec. 18-4. It can be concluded from the qualitative discussion given that the tetrode is characterized by the following features: a plate-grid capacitance which is only a few thousandths of that of a triode, a plate-grid transconductance which is roughly the same as that of a triode, and an amplification factor and plate resistance which are about ten or one hundred times that of a triode.

**18-2. Tetrode Characteristics.** As already discussed, the main purpose of the plate is to collect those electrons which succeed in passing through the screen. Also, the total space current is practically constant for given control-grid and screen-grid potentials. Hence, that portion of the space current which is not collected by the plate must be collected by the screen; *i.e.*, the two currents are complementary. Where the plate current is large, the screen current must be small, and vice versa. These features can be noted from the oscillogram of Fig. 18-2 obtained with a 6J7 tube connected as a tetrode. The screen-grid current curve has been inverted for purposes of clarity.

Although the plate voltage does not affect the total space current very markedly (a slight dip does occur in the curve of total space current at the lower plate potentials), it does determine the division of the space current

between the plate and the screen. At zero plate potential, none of the electrons has sufficient energy to reach the anode, if it is assumed that the electrons are liberated with zero initial velocities and that the effect of contact potential may be neglected. Hence, the plate current should be zero. As the plate voltage is increased, one should expect a rapid rise in plate current and a corresponding fall in the screen current. When the plate potential is very much larger than the screen potential, the plate current should approach the space current and the screen current should approach zero. This asymptotic behavior is noted on the oscillogram.

FIG. 18-2. Oscillogram of the total space current, the plate current, and the screen current (inverted for the sake of clarity) in a 6J7 connected as a tetrode.

FIG. 18-3. The approximate potential profiles in an idealized tetrode for several values of plate voltage. Two curves are shown for each plate voltage. One is for a path between grid and screen wires, and the other is for a path through the wires.

An inspection of the oscillogram indicates that the plate current rises very rapidly for the first few volts, and it is then followed by a rather anomalous behavior in the region of plate potentials from a few volts to potentials somewhat higher than the screen voltage. The plate current is seen to decrease with increasing values of plate potential. That is, the tube possesses a negative plate resistance in this region.

The general character of the curves of Fig. 18-2 may be described on the basis of the approximate potential distribution diagram of Fig. 18-3. This diagram should be compared with Fig. 16-5, which shows the potential profiles in a triode. In this diagram it is supposed that the control-grid and the screen-grid voltages are at fixed values and that the plate voltage

$E_b$ may be adjusted from zero to a value considerably in excess of the screen voltage.

It is seen that for a given grid and screen voltage, the curves for different plate voltages merge into one near the cathode. This means that the field intensity near the cathode is sensibly independent of the plate voltage and depends only on the grid and screen voltages. As a result, the space current does not depend on the plate voltage. The division of the space current between the screen and the plate does depend on their relative potentials, because of the emission of secondary electrons.



<div align="center">(a)                                  (b)</div>

FIG. 18-4. Oscillograms of the plate characteristics of two 89 tubes connected as tetrodes with grid voltage as a parameter. The plate current of tube (a) is always positive. The secondary emission is so large in tube (b) that the plate current is negative for some values of plate voltage.

Now consider the kinks, or folds, that appear in the curves of Fig. 18-2 in the region where the plate potential is lower than the screen potential. These are caused by the liberation of secondary electrons from the plate by the impact of the primary electrons with the plate. These secondary electrons are attracted to the screen. The screen current is increased, whereas the plate current is decreased. The number of secondary electrons liberated by this electron bombardment depends upon many factors and may exceed the total number of primary electrons that strike the plate and so may result in an effective negative plate current. These conditions are illustrated in the oscillograms of Fig. 18-4b, which show the plate characteristics of an 89 tube connected as a tetrode as a function of grid voltage for a fixed screen voltage. A large number of presumably identical tubes were tested; and although the majority had characteristics as illustrated in Fig. 18-4a, a number had a secondary-emission ratio greater than unity, which gives rise to the negative current over a portion of the characteristics, as in Fig. 18-4b.

In the region where the plate potential is higher than the potential of the screen, the secondary electrons that are liberated from the plate by the impact of the primary electrons are drawn back to the plate. In addi-

tion, some secondary electrons may also be liberated from the screen by the impact of the primary electrons on it. These secondary electrons from the screen are attracted to the plate, with the result that the plate current is slightly higher than it would be in the absence of secondary emission from the screen. Furthermore, the plate current continues to increase with increasing plate potentials because the collection of these secondary electrons is more complete. At the same time, the screen current tends toward zero.

Although the total space current should be a constant, in accordance with the previous discussion, Fig. 18-2 indicates that it is somewhat less than the ultimate value in the region of low plate potentials. This apparent inconsistency is readily explained in terms of the electrostatic deflection of the electrons by the screen wires at plate potentials that are less than the screen potential. The electrons that pass close to the screen wires are deflected from their path toward the plate and return toward the screen as a result of the electrostatic attraction. As they again approach the screen, they may once more suffer a deflection that will direct them toward the plate. This process may continue, the electrons oscillating back and forth between the grid wires until they are ultimately collected by the screen. Because of the electrons that may exist in the neighborhood of the screen, a partial space charge will develop. This will reduce the effec-

tive screen potential, with a consequent reduction in the total space-charge current in the tube.

**18-3. Transfer Characteristics.** As already noted, the plate of a tetrode has very little influence on the space current. It is expected, therefore, that the cathode, the control grid, and the screen grid should have characteristics similar to those of an ordinary triode. That this is actually so is seen from the curves of Fig. 18-5. The curves of Fig. 18-5 for fixed $E_{c2}$ show the effect of variations in the potential of the plate. By maintaining the screen at a fixed positive potential, the attractive field necessary to permit the flow



Fig. 18-5. Oscillogram of the transfer characteristics of an 89 tube connected as a tetrode for a fixed screen voltage and with the plate voltage as a parameter.

of electrons against the opposing force due to the space charge is supplied. Because of the slight influence of the plate, the transfer curves are much more bunched together than the transfer curves of triodes for different values of plate voltage, provided that $E_b > E_{c2}$.

It should be noted that the transfer curves for plate voltages below the screen voltage (a condition that is generally avoided in practice) become quite separated. As a matter of fact, the transfer characteristic for $E_b$

= 100 volts actually falls below that for $E_b = 50$ volts. This result follows from the method of construction of the transfer curves. In order to obtain the transfer curves from the plate characteristics (see Fig. 18-4), all that is necessary is to note the values of plate current along a vertical line (corresponding to the specified plate potential) and plot these results as a function of the grid voltage. It is evident that the curves at the higher plate potentials will be closely bunched. Furthermore, the apparently anomalous behavior of the transfer curves at plate potentials below the screen potential follows directly from this construction.

**18-4. Tube Parameters.** It has already been shown that the plate current of a triode is a function of the potentials on the grid and the plate. Similarly, for multielectrode tubes, variations in potential of the various electrodes influence the current to these electrodes. The plate current may be expressed as a function of the potential of the various electrodes by the expression

$$i_b = f(e_{c1}, e_{c2}, e_b) \tag{18-1}$$

where the symbol $e_{c1}$ denotes the voltage of the first, or control, grid, $e_{c2}$ denotes the voltage of the second, or screen, grid, and $e_b$ is the potential of the plate. This functional relationship is merely an extension of that which applies for triodes.

An approximate explicit form of the dependence is possible. This form, which is an extension of Eq. (16-3), may be written as

$$i_b = k\left(e_{c1} + \frac{e_b}{\mu_1} + \frac{e_{c2}}{\mu_2}\right)^{\!\frac{3}{2}} \tag{18-2}$$

where $\mu_1$ and $\mu_2$ are the control- and screen-grid amplification factors, respectively, and $k$ is called the "permeance."

The variation in the plate current, second and higher order terms in the expansion being neglected, is given by

$$\Delta i_b = \left(\frac{\partial i_b}{\partial e_{c1}}\right)\Delta e_{c1} + \left(\frac{\partial i_b}{\partial e_{c2}}\right)\Delta e_{c2} + \left(\frac{\partial i_b}{\partial e_b}\right)\Delta e_b \tag{18-3}$$

Generally, the screen potential is maintained constant at some appropriate value. Under these conditions, $\Delta e_{c2} = 0$, and the second term can be omitted from the equation. Note, however, that the screen grid may be used to exercise a control function in much the same way as the control grid. Owing to its disposition relative to the cathode, the amplification factor of the screen grid is less than that of the control grid. With $\Delta e_{c2} = 0$, and proceeding as for the case of triodes, the partial-differential coefficients appearing in this expression are used in order to define the tube parameters.

These are

$$\left(\frac{\partial e_b}{\partial i_b}\right)_{E_{c1},E_{c2}} \equiv r_p, \text{ plate resistance}$$

$$\left(\frac{\partial i_b}{\partial e_{c1}}\right)_{E_b,E_{c2}} \equiv g_m, \text{ plate control-grid transconductance} \qquad (18\text{-}4)$$

$$-\left(\frac{\partial e_b}{\partial e_{c1}}\right)_{I_b,E_{c2}} \equiv \mu, \text{ amplification factor}$$

The two subscripts outside of each parenthesis indicate the parameters that are maintained constant during the partial differentiation. These should likewise appear in Eq. (18-3) but have been omitted for the sake of simplicity.

It is easy to show that the relation $\mu = r_p g_m$ applies in the present case as well as for the triodes. Nominal values for the various parameters that appear in this relationship are $r_p = 0.1$ to 2 megohms, $g_m = 500$ to 3,000 micromhos, and $\mu = 100$ to 1,200 in the ordinary screen-grid tubes.

By employing the same system of notation in this case as for the triode, *viz.*,

$$\Delta i_b = i_p \qquad \Delta e_b = e_p \qquad \Delta e_{c1} = e_g$$

and if the screen voltage is maintained constant, so that $\Delta e_{c2} = 0$, as is customary, then Eq. (18-3) becomes

$$e_p = i_p r_p - \mu e_g \qquad (18\text{-}5)$$

This is precisely Eq. (17-9) for the triode.

**18-5. Equivalent Circuit of a Tetrode.** The basic equivalent-circuit representation of the tetrode is essentially that of the triode, even though



*(a)*-Schematic          *(b)*-Equivalent          *(c)*-Equivalent
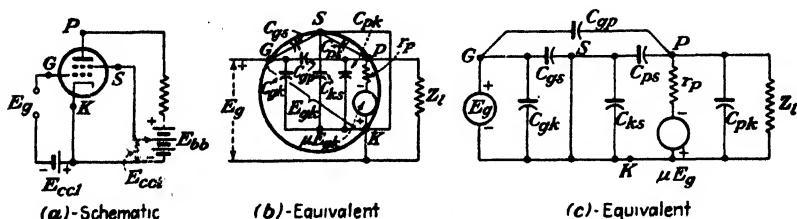
FIG. 18-6. The schematic and equivalent circuits of a tetrode connected as an amplifier.

a screen grid exists in the tetrode. The screen is connected to the cathode through a source of d-c potential which is usually less than, or at most equal to, the plate potential. The equivalent circuit of a tetrode is subject to the same limitations as the equivalent circuit of a triode. A schematic diagram of a simple amplifier circuit employing a tetrode is given in Fig. 18-6a. The equivalent circuit of Fig. 18-6b is drawn in a manner to illus-

trate the equivalent-circuit representation of the tube, including the inter-electrode capacitances.    Figure 18-6c shows the more conventional representation of such an equivalent amplifier circuit.

In drawing the equivalent amplifier circuit, the rules given in Sec. 17-6 have been appropriately extended and employed.    Thus, in addition to the points $K$, $G$, and $P$, the screen terminal $S$ is also marked.    The circuit elements of the original circuit are included in their appropriate positions between these four points, except that all d-c potentials are omitted and the tube itself is replaced by an equivalent generator $\mu E_{gk}$ having an internal resistance $r_p$, between the points $K$ and $P$. The capacitances between all pairs of the four electrodes are included, the double subscript denoting the pair of electrodes under consideration.



FIG. 18-7.   The ideal equivalent circuit of a tetrode.   The grid-plate capacitance has been assumed equal to zero.

Since the screen battery must be short-circuited in the equivalent circuit, this puts the screen at ground potential in so far as a-c variations about the $Q$ point are concerned.    Usually, the screen potential is obtained from the plate supply through a screen dropping resistor.    In this case a capacitor is connected from the screen to cathode.    This capacitor is chosen sufficiently large so that the screen potential remains constant even though the screen current may vary.    In this case, too, the screen is at a-c ground potential.    Thus, as indicated in the figure, this effectively shorts out $C_{ks}$ and puts $C_{gk}$ and $C_{gs}$ in parallel.    Let this parallel combination be denoted $C_1$.    The capacity $C_{ps}$ now appears from plate to ground and is effectively in parallel with $C_{pk}$.    Let this parallel combination be denoted $C_2$.    From the discussion of the shielding action of the screen grid, the capacitance between the plate and the control grid $C_{gp}$ has been reduced to a very small value.    If this capacitance is assumed to be negligible, then Fig. 18-6c may be redrawn more simply as shown in Fig. 18-7, where

$$C_1 = C_{gk} + C_{gs} \qquad C_2 = C_{ps} + C_{pk} \qquad (18\text{-}6)$$

Owing to the shielding action of the screen, little error will be made if $C_{pk}$ is neglected in comparison with $C_{ps}$, so that $C_2 = C_{ps}$, to a good approximation.    This capacitance acts as a shunt across the load.    The input admittance of the tube is thus seen to be simply

$$Y_i = j\omega C_1 \qquad (18\text{-}7)$$

A significant difference is seen to exist between the ideal equivalent circuit of the tetrode and the complete equivalent circuit of the triode, given in Fig. 17-22.    The idealization made here consists in the assump-

tion that the grid-plate capacitance is zero rather than a very small fraction of a micromicrofarad. The circuit of Fig. 18-7 clearly shows that under this condition the plate circuit has been isolated from the grid circuit, the possibility of adverse reactions, discussed in Sec. 17-9, being thus eliminated.

It should be pointed out, however, that the mere substitution of a tetrode for a triode will not, in general, effect any marked difference in the amplifier response. This follows because the wiring and stray capacitances between circuit elements external to the tube may provide the capacitances that the tube itself seeks to eliminate. It is necessary, therefore, that the elements of the circuit should be carefully arranged in order to permit short interconnecting leads, and generally neat wiring so as to reduce wiring capacitances. It is only if the capacitance between the grid and anode circuits external to the tube is small that the inherent possibilities of the tetrode can be utilized.

In order to calculate the approximate performance of circuits equipped with tetrodes, either the static curves of the tube or an analytical method based upon the equivalent circuit may be used. Since the equivalent circuit of the tetrode is that which was found to be valid approximately for triodes, all of the previous discussion and analysis actually apply more exactly for the present case than they did for triodes. It must be kept in mind that the results in both cases are based upon the existence of a linear dynamic curve.

**18-6. Pentodes.** Although the insertion of the screen grid between the control grid and the plate serves to isolate the plate circuit from the grid circuit, nevertheless the folds in the plate characteristic arising from the effects of secondary emission limit the range of operation of the tube. This limitation results from the fact that if the plate-voltage swing is made too large the instantaneous plate potential may extend into the region of rapidly falling plate current. This will cause a marked distortion in the output.

The kinks, or folds, that appear in the plate characteristic curves and that limit the range of operation of the tetrode may be removed or suppressed by inserting a coarse grid structure between the screen grid and the plate of Fig. 18-1. Tubes equipped with this extra suppressor grid are known as "pentodes" and were first introduced commercially in 1929. The suppressor grid must be maintained at a lower potential than the instantaneous potential reached by the plate. It is usually connected directly to the cathode, either internally in the tube or externally. Because the potential of the screen is considerably above that of the suppressor grid, a retarding force prevents the secondary electrons liberated from the screen to flow to the plate. On the other hand, the secondary electrons emitted from the plate are constrained, by the retarding field between the suppressor grid and the plate, to return to the plate. However, the electrons from

the cathode that pass through the screen are not kept from reaching the plate by the presence of the suppressor grid, although their velocities may be affected thereby.

The effect of the introduction of the suppressor grid is shown graphically in the oscillogram of Fig. 18-8. These curves were obtained by using a type 89 triple-grid tube. The tetrode characteristics are obtained by connecting the suppressor grid to the screen grid, the tube being thus converted into the conventional tetrode. The pentode characteristics result when the suppressor is connected to the cathode.
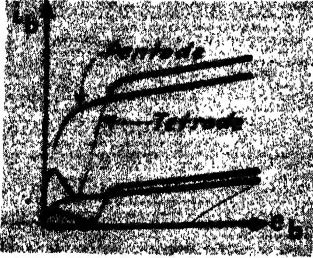


FIG. 18-8. Oscillogram of the plate characteristics of an 89 tube connected both as a tetrode and as a pentode, for two values of grid bias. Note that the pentode connection removes the kinks (due to secondary emission) of the tetrode characteristic.

The pentode has displaced the tetrode in radio-frequency voltage amplifiers because it permits a somewhat higher voltage amplification at moderate values of plate potential than does the tetrode. It has also found extensive application as an audio-frequency power-output tube. When used in the capacity of a power-output tube, the pentode will deliver large amounts of power with relatively small impressed grid signals and with fairly low potentials applied to the other electrodes. These characteristics result from the design of the tube, the plate resistance being of the order of 0.1 megohm and the amplification factor being of the order of 100.
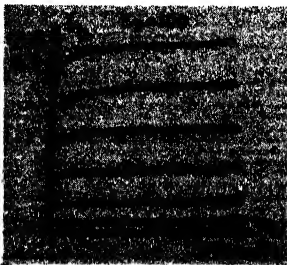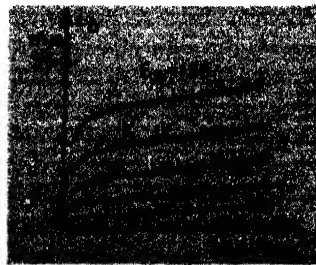


(a)                    (b)

FIG. 18-9. Oscillograms of the plate characteristics (a) of a 6J7 voltage pentode and (b) of an 89 power pentode with grid voltage as a parameter.

For purposes of comparison, the oscillograms of Fig. 18-9 showing the plate characteristics of a typical voltage pentode (type 6J7) and a typical power pentode (type 89) are included. The significant differences to be

noted are the magnitudes of the plate currents and the fact that power-pentode curves show a much more marked increase in plate current with increases of plate voltage. That is, the plate resistance of the type 6J7 tube is much higher than the plate resistance of the type 89 tube. These differences arise from the more complete shielding of the plate by the screen grid, and from the more closely wound grid structures in the 6J7 tube. That is, although the physical dimensions of these tubes are approximately the same, the number of grid and screen turns per inch in the 6J7 is almost twice that in the 89 tube.
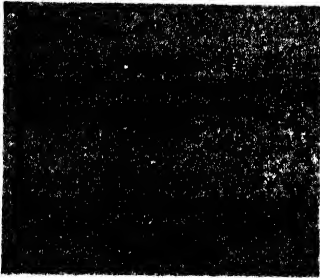


FIG. 18-10. Oscillogram of the total space current, the plate current, and the screen current (inverted for the sake of clarity) of a 6J7 pentode.

FIG. 18-11. Oscillogram of the transfer curves of an 89 tube connected as a pentode for a fixed screen voltage and with the plate voltage as a parameter.

The plate, screen, and total current curves as a function of the plate voltage are shown in Fig. 18-10 for the 6J7 tube. These should be compared with the corresponding tetrode curves of Fig. 18-2. It is noticed that the kinks resulting from the effects of secondary emission are entirely missing in the pentode. Furthermore, the screen current no longer falls asymptotically to zero but approaches a constant value for large plate voltages. This value is determined principally by the amount of space current that is intercepted by the screen-grid wires. An examination of the characteristics of a number of the more important pentodes indicates that the screen current is ordinarily from 0.2 to 0.3, the plate current at the recommended operating point. The total space current is seen to remain practically constant over the entire range of plate voltage, except for the very low values of potential.

The transfer curves of a type 89 tube connected as a pentode for $E_b = 400$ and $E_b = 50$ volts are shown in the oscillogram of Fig. 18-11. It is to be noted that these are much more independent of plate voltage than were the corresponding curves (Fig. 18-5) for the tube as a tetrode. Furthermore, no anomalous low-voltage curves are present, since the secondary emission that gives rise to this behavior has been eliminated.

The large power output with high gain that is possible with the power-output pentode results from the fact that the plate swings may be made very large. For such a tube, the plate voltage may swing lower than the screen voltage without serious distortion.

When used in a circuit as a voltage amplifier, the pentode is connected in the circuit exactly like a tetrode (see Fig. 18-6a), with the addition that the suppressor grid is connected to the cathode. Then, subject to the same considerations as for tetrodes (Secs. 18-4 and 18-5), it follows that the equivalent circuit concepts are also applicable for pentodes. By drawing the complete equivalent circuit of such a unit together with all tube capacitances and using the specified rules of Sec. 17-6 appropriately extended, it is easy to show that the ideal equivalent circuit of a pentode is also given by Fig. 18-7, which is repeated in Fig. 18-12 for convenience. In this diagram



FIG. 18-12. The equivalent circuit of a pentode.

$$C_1 = C_{gk} + C_{gs} \qquad C_2 = C_{pk} + C_{ps} + C_{p3} \qquad (18\text{-}8)$$

where $C_{p3}$ is the capacitance between the plate and grid 3 (the suppressor).

When the input and output capacitances of a tube are listed by the manufacturer, reference is being made to $C_1$ and $C_2$, respectively.

Since the shape and disposition of the control grid and cathode are the same in both the triode and the pentode, both types of tubes have comparable values of $g_m$. However, the values of $r_p$ and $\mu$ may be 100 times as great in the pentode as in the triode.

Since $r_p$ is generally very large compared with $Z_l$, it follows from Fig. 18-12 that the generator current (in the direction from $K$ to $P$ through the load) is $\mu E_g/r_p = g_m E_g$. Under these circumstances, the output-voltage drop is

$$\mathbf{E}_{pk} = -g_m \mathbf{E}_g Z \qquad (18\text{-}9)$$

where $Z$ is the combined parallel impedance in the output circuit. The gain is, as usual,

$$\mathbf{K} = \frac{\mathbf{E}_{pk}}{\mathbf{E}_{gk}} = -g_m Z \qquad (18\text{-}10)$$

If, as is usually the case, the reactance of the output capacitor is large compared with the load impedance, then $Z \doteq Z_l$ and the gain is given by

$$\mathbf{K} \doteq -g_m Z_l \qquad (18\text{-}11)$$

This is a most important result and one well worth remembering.

**18-7. Remote-cutoff Tubes.**[2] If the grid-cathode spacing, the spacing between grid wires, or the diameter of the grid wires is not uniform along the entire length of the control-grid structure, the various portions of the grid will possess different degrees of electrostatic control over the plate current. It is possible for one portion of the grid to cut off the flow of electrons for a given grid voltage, whereas a more open section might allow a considerable number of electrons to pass unimpeded to the plate. The result is that the plate current will decrease rather slowly as the grid voltage is made more and more negative. The asymptotic approach of the transfer character of the type 6K7 tube is illustrated in the oscillogram of Fig. 18-13. This is to be compared with the fairly sharp cutoff of the characteristic of the 6J7 pentode, which is provided with a uniformly wound grid.
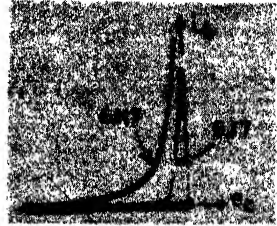


FIG. 18-13. Oscillogram of the transfer curves of a 6J7 sharp-cutoff pentode and a 6K7 remote-cutoff pentode.

Owing to its construction, a given grid-voltage increment results in a plate-current change that is a function of the grid bias. This means, of course, that the mutual conductance is a function of the bias. For this reason, these tubes are called "variable-mu" tubes. They are also known as "remote-cutoff" or "supercontrol" tubes.

**18-8. Beam Power Tubes.** The suppressor grid is introduced into the pentode in order to remove the kinks due to secondary emission in the tetrode. Although the pentode permits a greater range of operation than is possible with the tetrode, nevertheless the curvature of the $i_b$-$e_b$ characteristic limits the extent of operation of the pentode. The broad knee of the plate characteristic in the region of small plate voltage arises from the overeffectiveness of the suppressor grid at these low plate voltages. That is, the suppressor grid may prevent some of the primary electrons that pass through the screen from reaching the plate when the plate voltage is low. This effect arises from the deflections that may be given to some of the primary electrons if they approach close to a screen wire. Thus, if an electron is deflected by this action, its velocity in the plate direction will be reduced and the field of the suppressor grid may repel it back toward the screen.

Because of this, the shape of the suppressor grid in some modern pentodes has been so dimensioned [3] that the effects of secondary emission are just suppressed or are only admitted to a slight extent at the low anode voltages. This results in an improved plate characteristic and is manifest by a sharper break in the plate characteristic.

If the complete overeffectiveness of the suppressor grid could be avoided, the ideal power-tube characteristic, *viz.*, constant current for all plate

voltages, would be achieved. This is approached in the beam power tube,[4] a sketch of which is given in Fig. 18-14.

The general character of the plate characteristics from such a tube is compared with that of a power pentode in Fig. 18-15. One feature of the design of this tube is that each spiral turn of the screen is aligned with a spiral turn of the control grid. This serves to keep the screen current small and so leaves the plate current (the difference between the total space current and the screen current) virtually unchanged. The screen current in such tubes ranges from 0.05 to 0.08 of the plate current, which is considerably below the range 0.2 to 0.3 for pentodes. Other features are the flattened cathode, the beam-forming side plates (maintained at zero potential), the shape of the plate, the curvature of the grids, and the spacing of the various elements. As a result of these design characteristics, the electrons flow between the grid wires toward the plate in dense sheets or beams.



FIG. 18-14. Schematic view of the shapes and arrangements of the electrodes in a beam power tube. (*Courtesy of RCA.*)

The region between the screen and the plate possesses features which are somewhat analogous to those which exist in the space-charge limited diode. That is, a flow of electrons exists between two electrodes between which a difference of potential exists. There is one significant difference,
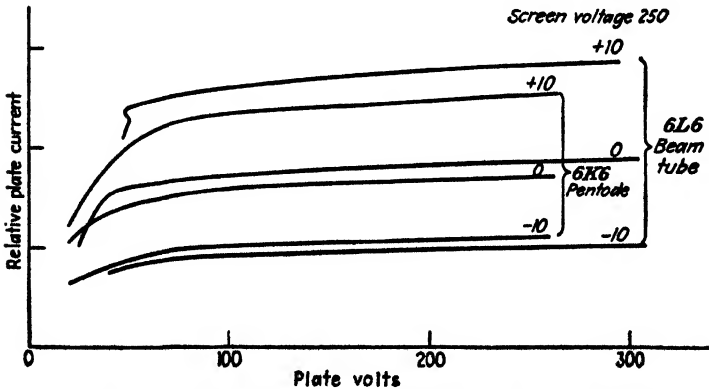


FIG. 18-15. The plate characteristics of a beam power tube and a power pentode. The beam power tube current is approximately twice that of the pentode.

however.   Whereas the electrons leave the cathode of a diode with almost zero initial velocities, the electrons that pass through the screen wires in the beam tube do so with a velocity corresponding essentially to the screen potential.   As described in Sec. 7-4 in connection with the effects of initial velocities on the space-charge equation, the effect of the initial velocities of the electrons in the screen-plate region will appear as a potential minimum in this region (see Fig. 7-4).   This is shown in the approximate potential
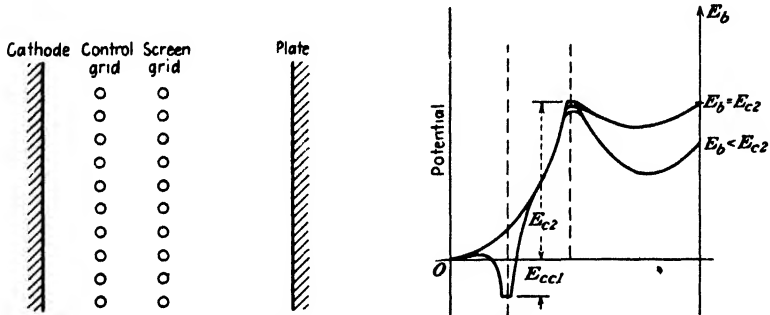


FIG. 18-16.   Approximate potential profile in an idealized beam power tube for two values of plate voltage.   Two curves are shown for each plate voltage.   One is for a path between grid and screen wires, and the other is for a path through the wires.   Note the potential minimum in the region between the screen grid and the plate.

profile in Fig. 18-16 and is to be compared with the corresponding figure for the tetrode (see Fig. 18-3).

The production of the potential minimum also receives contributions from a suppressor grid, if one is present.   Since the predominant effect in the beam tube is the space charge, the suppressor grid is sometimes reduced to a pair of plates (the beam-forming side plates of Fig. 18-14).[5] However, other beam tubes are provided with a mechanical suppressor grid.   The potential minimum that is produced acts as a virtual suppressor grid, since any secondary electrons emitted from either the plate or the screen will encounter a potential-energy barrier and so will be compelled to return to the electrode (which is at a positive potential with respect to the potential minimum) from which they originate.

The actual potential distribution [6] in the screen-plate region will depend upon the instantaneous plate potential and the plate current (a constant screen potential being assumed) and so is not constant.   This results in variable suppressor action.   This is quite different from the action that arises in a simple pentode provided only with a mechanical grid structure for supplying the retarding field.

Thus, because of the beam formation, which serves to keep the screen current small, and because of the variable suppressor action, which serves

to suppress secondary emission from the screen and from the plate, the ideal power-tube characteristic is closely approximated. A family of plate characteristics for the 6L6 is shown in the oscillograms of Fig. 18-17. It should be noted that this tube is a tetrode when considered in terms of the number of active electrodes. At low currents, where the suppressor action of the beam is too small, the characteristic "kinks" of a tetrode are noticeable.
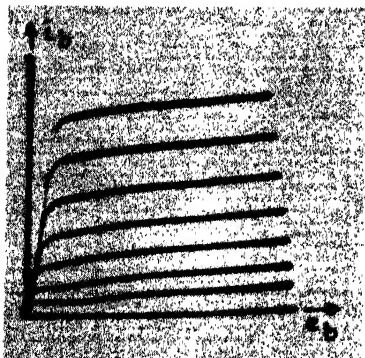


Fig. 18-17. Oscillograms of the volt-ampere characteristics of a 6L6 beam power tube.
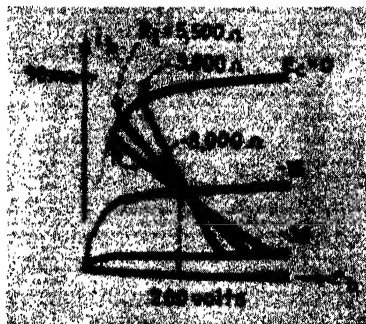


Fig. 18-18. Oscillograms of the plate characteristics of a 6F6 pentode. Three load lines corresponding to re-sistance values of 3,000, 5,500 and 8,000 ohms are shown passing through the quiescent point $E_b = 200$ volts, $E_c = -15$ volts.

**18-9. Distortion in Multielectrode Tubes.** The discussion (Sec. 17-12) of harmonic generation in a vacuum tube resulting from the nonlinear dynamic characteristic is applicable to multielectrode tubes as well as to triodes. Under ordinary conditions of operation of a triode, the second harmonic is the principal component of distortion. It will be shown below that it is possible to eliminate completely, by the proper choice of the load resistance, the second harmonic distortion component in a multielectrode tube. The third-order distortion term is then the most important.

Consider a 6F6 pentode operating under the following conditions: $E_f = 6.3$ volts, $E_{c2} = 250$ volts, $E_b = 200$ volts, and $E_c = -15$ volts. Three load lines corresponding to load resistances of 3,000, 5,500, and 8,000 ohms pass through the quiescent point, as shown in the oscillograms of Fig. 18-18. The corresponding three dynamic curves are shown in the oscillograms of Fig. 18-19. Suppose that a signal voltage having a peak value of 15 volts is impressed on the grid. The grid will then swing from 0 to $-30$ volts. The corresponding output currents for the three load resistances are shown

in the oscillogram of Fig. 18-20. It is seen that the output closely resembles a sine wave for the case $R_l = 5,500$ ohms. For $R_l = 8,000$ ohms, the upper half of the wave is decidedly smaller than the lower loop, whereas



Fig. 18-19. Oscillograms of the dynamic curves corresponding to the three load lines of Fig. 18-18.

for $R_l = 3,000$ ohms the converse is true. The lower half of the current wave is identical for all three resistances.

An inspection of either the static or the dynamic characteristics of the 6F6 reveals the reason for these results. The bottom half of the wave is obtained as the grid swings from $-15$ to $-30$ volts. At the quiescent
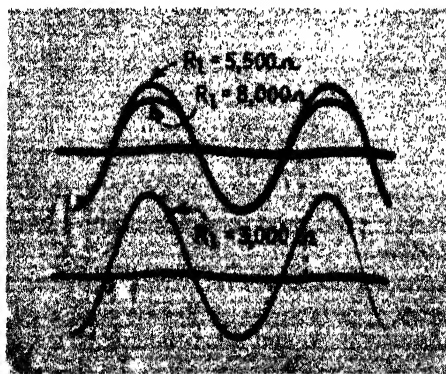


Fig. 18-20. Oscillograms of the wave forms of the output current of a 6F6 pentode when operated at the quiescent point and with the load resistances indicated in Figs. 18-18 and 18-19. The 5,500-ohm and the 8,000-ohm curves are superposed in order to demonstrate that the lower halves of these curves are identical.

point ($-15$ volts) the current is the same for all three loads. At $-30$ volts the current is still practically the same for all loads, for the $i_b$-$e_b$ characteristic is practically a horizontal line, as seen in Fig. 18-18. Consequently, the lower halves of the current waves are almost identical for the three loads.

The upper loop of the output-current wave is obtained as the grid swings from $-15$ volts to zero. Since the $i_b$-$e_b$ characteristic for $e_c = 0$ is a rapidly varying one in the region of low values of $e_b$, the peak current depends critically upon the point of intersection of the load line with this characteristic. For low resistances the point of intersection will be on the upper portion of the $e_c = 0$ curve. This results in a peaked output wave. For high resistances, the point of intersection will be on the rapidly falling portion of this curve and gives rise to a flat-topped curve. Had the load resistors been higher than 8,000 and lower than 3,000 ohms, the differences between the upper and lower loops of the output wave would have been much more pronounced.

The choice of $R_l = 5,500$ ohms was determined by the requirement that the second harmonic distortion be zero. This value was chosen to make the value of $B_2$, given by Eq. (17-49), equal to zero. That is, since

$$B_2 = \tfrac{1}{4}(I_{\max} - 2I_b + I_{\min})$$

$I_{\max}$ must equal $2I_b - I_{\min}$ in order that $B_2 = 0$. It is found from Fig. A9-4 that $I_b$, under the prescribed conditions of operation, is equal to 37 ma. Also, $I_{\min}$ (corresponding to $E_c = -30$ volts) is substantially independent of the load line and is found to be 7 ma. Hence, $I_{\max}$ must equal

$$2 \times 37 - 7 = 67 \text{ ma}$$

The corresponding value of $e_b$ (for $e_c = 0$, $i_b = 67$ ma) is 35 volts. It follows that

$$R_l = \frac{200 - 35}{67 - 37} = 5.5 \text{ kilohms}$$

The other two values of load resistance were arbitrarily chosen in order to illustrate the distortion.

It follows from the discussion above, which is entirely general, that the load required in order to reduce the second harmonic term to zero will depend upon the quiescent point and upon the grid swing.

*Example.* · Calculate the distortion in the output of the amplifier discussed above, if it is working into an 8,000-ohm load.

*Solution.* A load line corresponding to 8,000 ohms is drawn through the quiescent point, $E_c = -15$, $E_b = 200$, on the static curves of Fig. A9-4. The following points of intersection of this line with the plate characteristics are obtained:

| $e_c$ | 0 | $-5$ | $-10$ | $-15$ | $-20$ | $-25$ | $-30$ |
|---|---|---|---|---|---|---|---|
| $i_b$ | 59 | 56 | 48 | 37 | 26 | 15 | 7 |

The values of $I_{\frac{1}{2}}$ and $I_{-\frac{1}{2}}$ introduced in Sec. 17-12 are the current values when the grid swings to one-half its peak values (positively and negatively, respectively). The peak swing in this case is 15 volts, so that $I_{\frac{1}{2}}$ is the current for $e_c = -7.5$ volts. $I_{-\frac{1}{2}}$ represents the current for $e_c = -22.5$ volts. These values do not appear in the table and must be obtained either by interpolation or from the plotted dynamic curve ($i_b$ vs. $e_c$). The latter method yields $I_{\frac{1}{2}} = 52$ ma and $I_{-\frac{1}{2}} = 21$ ma. From the table, $I_{max} = 59$ ma, $I_b = 37$ ma, and $I_{min} = 7$ ma. These are the five values required for the five-point schedule. Using the formulas of Eq. (17-49), there result

$$B_0 = -1.7 \qquad B_1 = 28 \qquad B_2 = -2.0 \text{ ma}$$

$$B_3 = -1.7 \qquad B_4 = -0.3 \text{ ma}$$

Hence, the second, third, and fourth harmonic distortion components are

$$D_2 = \% \, 2d = 100 \frac{|B_2|}{|B_1|} = \frac{200}{28} = 7\%$$

$$D_3 = \% \, 3d = 100 \frac{|B_3|}{|B_1|} = \frac{170}{28} = 6\%$$

$$D_4 = \% \, 4th = 100 \frac{|B_4|}{|B_1|} = \frac{30}{28} = 1\%$$

The total distortion is

$$\text{Distortion} = \sqrt{7^2 + 6^2 + 1^2} = 9.3\%$$

A similar analysis for the 3,000-ohm load yields the values

$$B_0 = 2.5 \qquad B_1 = 37 \qquad B_2 = 2.8 \text{ ma}$$

$$B_3 = -0.5 \qquad B_4 = 0.25 \text{ ma}$$

$$D_2 = 7.6 \qquad D_3 = 1 \qquad D_4 = 0.7\%$$

For the 5,500-ohm load, the corresponding quantities are

$$B_0 = 0 \qquad B_1 = 31 \qquad B_2 = 0 \text{ ma}$$

$$B_3 = -1.3 \qquad B_4 = 0 \text{ ma}$$

$$D_2 = 0 \qquad D_3 = 4\% \qquad D_4 = 0$$

**18-10. Hexodes, Heptodes, Multipurpose Tubes.** Some special-purpose tubes contain two or three conventional tubes in a common envelope but otherwise are independent of each other. For example, the 6SN7 is a twin triode, each section of which has the characteristics of the 6J5 triode.

A number of special-purpose tubes, some of which contain more grid elements than in the pentode, are used extensively. These tubes possess a wide variety of characteristics, depending on the grids to which fixed potentials are applied and those to which signals are applied. These tubes are used in applications which are outside the scope of this book and so will not be considered further.

## PROBLEMS

**18-1.** Draw the equivalent circuit of a pentode including *all* interelectrode capacitances. Show that to a very good approximation the input capacitance is equal to $C_{gk}$ and $C_{gs}$ in parallel and that the output capacitance is equal to $C_{pk}$, $C_{ps}$, and $C_{p3}$ in parallel. The symbols are defined in Sec. 18-6.

**18-2.** *a.* From the plate characteristics of a 6F6 power pentode (see Appendix IX), draw the transfer characteristics for a screen voltage of 250 volts and the following values of plate voltage: $E_b = 80$, 160, 240, and 320 volts.

*b.* Repeat part *a* for a 6SJ7 voltage pentode.

**18-3.** *a.* A 6F6 pentode (see Appendix IX) is operated at the quiescent point $E_c = -15$ volts, $E_b = 200$ volts. Draw the dynamic curves for the following values of load resistance: 1, 5.5, and 10 kilohms.

*b.* The grid signal voltage is sinusoidal with a peak value of 15 volts. Plot the output current as a function of time for each of the load resistances in part *a*.

*c.* Calculate the second, third, and fourth harmonic distortion for the 1- and the 10-kilohm loads.

**18-4.** *a.* Calculate the distortion components in the output of a 6F6 tube connected as a pentode that supplies power to an effective 2-kilohm load resistor. The screen is maintained at 250 volts, the grid bias is $-20$ volts, and the plate voltage at the quiescent point is 300 volts. The grid signal is sinusoidal and has a peak value of 20 volts.

*b.* The tube is converted into a triode by connecting the screen to the plate. All other conditions are the same as in part *a*. Recalculate the distortion components.

The plate characteristics of the 6F6 when connected as either a pentode or triode are given in Appendix IX.

**18-5.** *a.* A 6F6 pentode is operated with a screen voltage of 250 volts and a grid bias of $-15$ volts. The grid signal is sinusoidal and has a peak value of 15 volts. Calculate and plot as a function of the quiescent-plate voltage $E_b$ the load resistance that must be used to give zero second harmonic distortion.

*b.* Calculate and plot as a function of $E_b$ the third harmonic distortion for the conditions in part *a*.

**18-6.** A 6L6 beam power tube (see Appendix IX for the plate characteristics) operates at the quiescent point $E_b = 250$ volts, $E_{c2} = 250$ volts, $E_{c1} = -14$ volts. The grid signal is sinusoidal and has a peak value of 14 volts. The load resistor is adjustable from 1 to 5 kilohms. Calculate and plot the d-c current, the second harmonic distortion, and the third harmonic distortion as a function of the load resistance.

**18-7.** A 6L6 is operated at the quiescent point $E_{c1} = -14$, $E_b = 250$, and $E_{c2} = 250$. The peak grid swing is 14 volts. Use Fig. A9-6.

*a.* What must the load resistance be in order to eliminate second harmonic distortion?

*b.* For a load resistance of 2.5 kilohms, calculate the per cent third harmonic distortion.

**18-8.** The circuit of Fig. 11-12 on page 307 is modified by replacing the resistor $R$ by a 6SJ7 pentode (see Appendix IX for the plate characteristics).

*a.* Explain clearly why this will result in an approximately linear sweep. How can the amplitude of the sweep be controlled? How can the frequency be adjusted without changing the amplitude?

*b.* The screen voltage on the 6SJ7 is 100 volts. The plate supply voltage is 300 volts. The capacitor is 0.1 μf. A 60-cycle 150-volt linear sweep is desired. Specify the grid voltages of the 6SJ7 and the 884 tubes. Assume that the critical grid characteristic of the 884 is given in Fig. 11-10 (page 305).

*c.* The screen voltage on the 6SJ7 is 100 volts. The plate supply voltage is 300 volts.

A 200-volt 5,000-cycle sweep is desired. Specify the grid voltages of the 6SJ7 and the 884 tubes and also the value of the capacitance $C$.

**18-9.** *a.* Why would it be impossible to design an 80-volt 100-cycle linear sweep using a d-c power supply voltage of 250 volts, a 0.05-$\mu$f capacitor, an 884 thyratron (see Fig. 11-10) and a 929 phototube (see Fig. 15-17)?

*b.* Replace the phototube with a suitable pentode (selected from Appendix IX), and complete the design. Sketch the complete circuit, specify the pentode used, and give *all* battery voltages.

*c.* If the peak current rating of the 884 is 0.5 amp, what is a suitable value for the plate-current limiting resistance?

**18-10.** Design a 150-volt 10,000-cycle linear sweep using a d-c power supply voltage of 250 volts. Sketch the circuit and specify all tubes, all tube voltages, all resistors, capacitors, etc. Explain your analysis, and state clearly any assumptions you may make.

**18-11.** *a.* The circuit shown can be used as a d-c voltmeter with a very high input impedance. The two tubes are identical pentodes although shown as triodes for convenience. The indicating instrument has a resistance $R_m$. Draw the equivalent circuit, and show that the meter current will be

$$\frac{\mu RE}{(R_k + \mu R_k + R + r_p)(2R + R_m) - 2R^2}$$

where $E$ is the d-c voltage to be measured.

*b.* The meter is a 1-ma milliammeter whose resistance is 100 ohms. $R = 10$ kilohms. The tubes are 6SJ7's for which $g_m = 1,650$ micromhos and $r_p = 1.0$ megohms, approximately. If $R_k \gg r_p/\mu = 1/g_m$, show that $I_m = E/2R_k$ approximately. For example, if $R_k = 50$ kilohms, full-scale reading will correspond to 100 volts. Note that the meter reading depends only upon $R_k$ and hence the calibration will be stable, *i.e.*, independent of the aging or the replacement of the tubes or voltage supply. Show that, in general, $I_m = E/2R_k$ provided that the following inequalities are true:

$$r_p \gg R \gg R_m, \qquad \mu \gg 1, \qquad \text{and} \qquad R_k \gg \frac{1}{g_m}$$
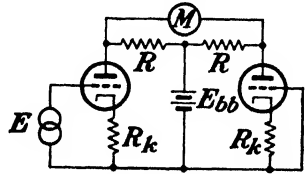
PROB. 18-11.

## REFERENCES

1. PIDGEON, H. A., *Bell System Tech. J.*, **14**, 44, 1935.
2. BALLANTINE, S., and H. A. SNOW, *Proc. IRE*, **18**, 2102, 1930.
3. VAN DER VEN, A. J. H., *Wireless Eng.*, **16**, 444, 1939.
4. SCHADE, O. H., *Proc. IRE*, **26**, 137, 1938.
5. JONKER, J. L. H., *Wireless Eng.*, **16**, 274, 1939.
6. FAY, C. E., A. L. SAMUEL, and W. SCHOCKLEY, *Bell System Tech. J.*, **17**, 49, 1938.
   SALZBERG, B., and A. V. HAEFF, *RCA Rev.*, **2**, 336, 1938.

### General References

For general references, see end of Chap. 16.

# APPENDIX I

## PROBABLE VALUES OF GENERAL PHYSICAL CONSTANTS *

| Constant | Symbol | Value |
|---|---|---|
| Electronic charge......... .......... | $e$ | $4.802 \times 10^{-10}$ statcoulomb |
| | | $1.602 \times 10^{-19}$ coulomb |
| Electronic mass...................... | $m$ | $9.107 \times 10^{-31}$ kg |
| Ratio of charge to mass of an electron.... | $e/m$ | $1.759 \times 10^{11}$ coulombs/kg |
| Mass of atom of unit atomic weight (hypo- | | |
| thetical)........................... | .... | $1.660 \times 10^{-27}$ kg |
| Mass of proton....................... | $M_1$ | $1.672 \times 10^{-27}$ kg |
| Ratio of proton to electron mass......... | $M_1/m$ | 1,836.5 |
| Planck's constant..................... | $h$ | $6.624 \times 10^{-34}$ joule-sec |
| Boltzmann constant................... | $k$ | $1.381 \times 10^{-23}$ joule/°K |
| Stefan-Boltzmann constant............. | $\sigma$ | $5.672 \times 10^{-8}$ watt/$(m^2)(°K^4)$ |
| Avogadro's number.................... | $N_0$ | $6.023 \times 10^{23}$ molecules/mole |
| Gas constant................... ....... | $R$ | 8.314 joules/(deg)(mole) |
| | | 1.986 cal/(deg)(mole) |
| Velocity of light...................... | $c$ | $2.998 \times 10^8$ m/sec |
| Faraday's constant.................... | $F_0$ | 96,488 coulombs |
| Volume per mole...................... | .... | $2.2415 \times 10^{-2}$ m³ |
| Mechanical equivalent of heat.......... | .... | 4.185 joules/cal |
| Acceleration of gravity................. | $g$ | 9.807 m/sec² |

* R. T. BIRGE, *Rev. Modern Phys.*, **13**, 233, 1941.
  R. T. BIRGE, *Am. J. Phys.*, **13**, 63, 1945.

# APPENDIX II

# CONVERSION FACTORS

| | |
|---|---|
| 1 ampere | $= 3 \times 10^9$ statamp |
| | $= \frac{1}{10}$ abamp |
| 1 coulomb | $= 3 \times 10^9$ statcoulombs |
| | $= \frac{1}{10}$ abcoulomb |
| 1 volt | $= \frac{1}{300}$ statvolt |
| | $= 10^8$ abvolts |
| 1 electron volt | $= 1.60 \times 10^{-19}$ joule |
| 1 weber | $= 10^8$ maxwells |
| 1 weber per square meter | $= 10^4$ gauss |
| 1 milliweber per square meter | $= 10$ gauss |
| 1 farad | $= 9 \times 10^{11}$ statfarads |
| | $= 10^{-9}$ abfarad |
| 1 angstrom unit | $= 10^{-8}$ cm |
| 1 micron | $= 10^{-4}$ cm |
| 1 atmosphere pressure | $= 76.0$ cm Hg |
| | $= 29.92$ in. Hg |
| | $= 14.70$ psi |
| 1 inch | $= 2.54$ cm |
| | $= 10^3$ mils |
| 1 foot | $= 0.305$ m |
| 1 gram force | $= 980.6$ dynes |
| 1 gram-calorie | $= 4.185$ joules |
| 1 joule | $= 10^7$ ergs |
| | $= 1$ watt-sec |
| 1 kilogram | $= 10^3$ g |
| | $= 2.205$ lb |
| 1 liter | $= 10^3$ cm$^3$ |
| 1 lumen per square foot | $= 1$ ft-c |
| 1 meter | $= 100$ cm |
| | $= 39.37$ in. |
| 1 mile | $= 5,280$ ft |
| | $= 1.609$ km |
| 1 mile per hour | $= 0.447$ m/sec |
| 1 pound | $= 453.6$ g |
| 1 radian | $= 57.3°$ |

# APPENDIX III *

## PERIODIC TABLE OF THE ELEMENTS

| | I | II | III | IV | V | VI | VII | VIII |
|---|---|---|---|---|---|---|---|---|
| 1 | H 1 1.0081 | | | | | | | He 2 4.002 |
| 2 | Li 3 6.940 | Be 4 9.02 | B 5 10.82 | C 6 12.01 | N 7 14.008 | O 8 16.000 | F 9 19.00 | Ne 10 20.183 |
| 3 | Na 11 22.997 | Mg 12 24.32 | Al 13 26.97 | Si 14 28.06 | P 15 31.02 | S 16 32.06 | Cl 17 35.457 | A 18 39.944 |
| 4 | K 19 39.096 | Ca 20 40.08 | Sc 21 45.10 | Ti 22 47.90 | V 23 50.95 | Cr 24 52.01 | Mn 25 54.93 | Fe 26 55.84   Co 27 58.94   Ni 28 58.69 |
| 4 | Cu 29 63.57 | Zn 30 65.38 | Ga 31 69.72 | Ge 32 72.6 | As 33 74.91 | Se 34 78.96 | Br 35 79.916 | Kr 36 83.7 |
| 5 | Rb 37 85.48 | Sr 38 87.63 | Y 39 88.92 | Zr 40 91.22 | Cb 41 92.91 | Mo 42 96.0 | Tc 43 | Ru 44 101.7   Iu 45 102.91   Pd 46 106.7 |
| 5 | Ag 47 107.880 | Cd 48 112.41 | In 49 114.76 | Sn 50 118.70 | Sb 51 121.76 | Te 52 127.61 | I 53 126.92 | Xe 54 131.3 |
| 6 | Cs 55 132.91 | Ba 56 137.36 | La 57 138.92 | Hf 72 178.6 | Ta 73 180.88 | W 74 184.0 | Re 75 186.31 | Os 76 191.5   Ir 77 193.1   Pt 78 195.23 |
| 6 | Au 79 197.2 | Hg 80 200.61 | Tl 81 204.39 | Pb 82 207.21 | Bi 83 209.00 | Po 84 | At 85 | Rn 86 222 |
| 7 | Fr 87 | Ra 88 226.05 | Ac 89 | Th 90 232.12 | Pa 91 231 | U 92 238.07 | | |

### THE RARE EARTHS

[To go between La (57) and Hf (72)]

| Ce 58 140.13 | Pr 59 140.92 | Nd 60 144.27 | Pm 61 | Sm 62 150.43 | Eu 63 152.0 | Gd 64 156.9 |
|---|---|---|---|---|---|---|
| Tb 65 159.2 | Dy 66 162.46 | Ho 67 163.5 | Er 68 167.64 | Tm 69 169.4 | Yb 70 173.04 | Lu 71 175.0 |

* The number to the right of the symbol for the element gives the atomic number.
The number below the symbol for the element gives the atomic weight.
This table does not include Np 93, Pu 94, Am 95, and Cm 96, the synthetically produced elements above 92.

# APPENDIX IV

# THE MKS SYSTEM

| Quantity | Symbol | Unit |
|----------|--------|------|
| Displacement . . . . . . . . . . . . | $x, y, z$ | meters |
| Mass . . . . . . . . . . . . . . . . . . . . . | $m$ | kilograms |
| Time . . . . . . . . . . . . . . . . . . . . . . | $t$ | seconds |
| Force . . . . . . . . . . . . . . . . . . . . . | $\mathbf{f}$ | newtons |
| Velocity . . . . . . . . . . . . . . . . . . | $\mathbf{v}$ | meters/second |
| Power . . . . . . . . . . . . . . . . . . . . | $P$ | watts |
| Energy . . . . . . . . . . . . . . . . . . | $W$ | joules |
| Electric charge . . . . . . . . . . . . | $q$ | coulombs |
| Displacement flux density . . . . . | $D$ | coulombs/square meter |
| Potential . . . . . . . . . . . . . . . . . | $V$ | volts |
| Current . . . . . . . . . . . . . . . . . . | $I$ | amperes |
| Electric-field intensity . . . . . . . | $\boldsymbol{\mathcal{E}}$ | volts/meter |
| Charge density . . . . . . . . . . . . . | $\rho$ | coulombs/cubic meter |
| Current density . . . . . . . . . . . . | $J$ | amperes/square meter |
| Magnetomotive force . . . . . . . . . | | amperes |
| Magnetic-field intensity . . . . . . | $\mathbf{H}$ | amperes/meter |
| Magnetic flux . . . . . . . . . . . . . . | | webers |
| Magnetic-flux density . . . . . . . | $\mathbf{B}$ | webers/square meter |
| Resistance . . . . . . . . . . . . . . . . | $R$ | ohms |
| Capacitance . . . . . . . . . . . . . . . | $C$ | farads |
| Inductance . . . . . . . . . . . . . . . . | $L$ | henrys |
| Permeability of free space . . . . . | $\mu_0$ | $4\pi \times 10^{-7}$ henry/meter |
| Permittivity of free space . . . . . | $\epsilon_0$ | $\dfrac{1}{36\pi \times 10^9}$ farad/meter |

# APPENDIX V

# DEFINITE INTEGRALS OF THE FORM $\xi_n = \int_0^\lambda x^n \epsilon^{-\lambda x^2} dx$

<div align="center"><i>n</i> Even</div>

$$n = 0 \qquad \xi_0 = \frac{1}{2}\sqrt{\frac{\pi}{\lambda}}$$

$$n = 2 \qquad \xi_2 = \frac{1}{4}\sqrt{\frac{\pi}{\lambda^3}}$$

$$n = 4 \qquad \xi_4 = \frac{3}{8}\sqrt{\frac{\pi}{\lambda^5}}$$

$$\vdots \qquad \vdots$$

$$n = 2m \qquad \xi_{2m} = \frac{1 \times 3 \times (2m - 1)}{2^{m+1}}\sqrt{\frac{\pi}{\lambda^{2m+1}}}$$

<div align="center"><i>n</i> Odd</div>

$$n = 1 \qquad \xi_1 = \frac{1}{2\lambda}$$

$$n = 3 \qquad \xi_3 = \frac{1}{2\lambda^2}$$

$$n = 5 \qquad \xi_5 = \frac{1}{\lambda^3}$$

$$\vdots \qquad \vdots$$

$$n = 2m + 1 \qquad \xi_{2m+1} = \frac{m!}{2\lambda^{m+1}}$$

# APPENDIX VI

# POISSON'S EQUATION

The ease with which certain problems in electrostatics can be solved depends upon the method of analysis. For example, if it is desired to find the potential at any point in space resulting from a given configuration of discrete point charges, one would use Coulomb's law. In this case the electrostatic potential is given by *

$$V = \sum_{\text{all charges}} \frac{q}{4\pi\epsilon r}$$

where $q$ is the magnitude of each charge (in coulombs), $r$ is the corresponding distance (in meters) from the charge to the point at which the potential is desired, and $\epsilon$ is the dielectric constant of the medium (in farads per meter). If the number of charges is so large that they may be considered to form a practically continuous distribution of charge, it is more convenient to discuss the problem in terms of Gauss's law, which relates the total number of lines of electric flux emanating from a given volume of charge with the total charge contained within the volume. It is assumed that the reader is familiar with this law from elementary electrostatic theory.

In some cases a slightly different mathematical form of Gauss's law is found useful. This is Poisson's equation, which is not so well known as either Coulomb's law or Gauss's law. Actually, however, Poisson's equation is nothing more than an analytical expression relating the potential as a function of the distance in the region of free charges. It is, in fact, simply Gauss's law stated in differential form, as will be evident from the following derivation.

Consider the region between a plane infinite cathode and a plane-parallel anode, the coordinate axes being chosen as illustrated in Fig. A6-1. The $X$ axis is normal to the plates. Assume that a *positive* space-charge density $\rho$ exists in the region between the plates. Since the plates are assumed to be infinite in extent, the density of free charge in the region between the plates will vary only along the $X$ direction. The number of lines of electric

* F. W. Sears, "Principles of Physics II," Addison-Wesley Press, Inc., Cambridge, Mass., 1947.

displacement flux that emanate from within the element of volume $dx\,dy\,dz$, situated at the point $P(x, y, z)$ will be $\rho\,dx\,dy\,dz$ in accordance with Gauss's law. Further, the lines of electric flux must be parallel to the $X$ axis, for the electrodes have been chosen as infinite in extent in the $Y$ and $Z$ directions.

The electric-displacement flux density $D$ is expressed as the number of lines of flux per square meter. Thus, the amount of flux entering the vol-
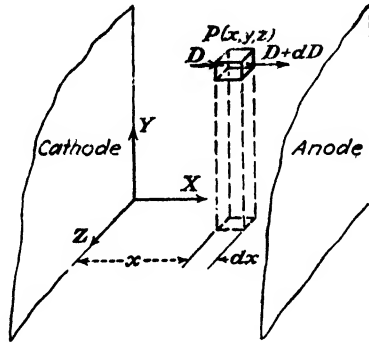


FIG. A6-1.

ume element through the left face is $D\,dy\,dz$. Because of the presence of charge in this volume element, the electric-field intensity at the opposite face will be slightly different from that at the left face. By denoting this value as $D + dD$, then the flux leaving the parallelepiped is $(D + dD)\,dy\,dz$. Owing to the assumed geometry of the system, the flux through the other pairs of bounding faces is zero. Hence, the net number of lines of force arising in this volume is

$$(D + dD)dy\,dz - D\,dy\,dz = dD\,dy\,dz$$

Equating this expression for the net flux that arises within the volume element to the value dictated by Gauss's law, then

$$dD\,dy\,dz = \rho\,dx\,dy\,dz$$

which leads to the expression

$$\frac{dD}{dx} = \rho \tag{A6-1}$$

Since, by definition, the electric-field intensity $\mathcal{E}$ is the negative gradient of the potential, $V$, and since $D = \epsilon\mathcal{E}$, then

$$D = -\epsilon\frac{dV}{dx}$$

Then Eq. (A6-1) may be expressed in the form

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon} \qquad (A6\text{-}2)$$

This expression is known as *Poisson's equation*.

Of course, if the variation of potential does not occur in a single direction only, Eq. (A6-2) must be extended in order to take due account of the variations in both the $Y$ and the $Z$ directions. The general form of this extended expression, in practical units, can be shown, by a simple extension of the foregoing method of development, to be

$$\frac{\partial^2V}{\partial x^2} + \frac{\partial^2V}{\partial y^2} + \frac{\partial^2V}{\partial z^2} = -\frac{\rho}{\epsilon} \qquad (A6\text{-}3)$$

It should be kept strictly in mind that $\rho$, the volume density of charge at any point $P(x, y, z)$, may be either positive or negative, depending upon whether the free charge density arises from positive ions or from electrons. If both positive ions and electrons are present, then $\rho$ will be the difference between the positive-ion and electron densities.

Inside of a vacuum tube $\epsilon$ must be replaced by $\epsilon_0[=1/(36\pi \times 10^9)$ farads/m], the permittivity of a vacuum in mks rationalized units.

For a system possessing cylindrical symmetry, in which the potential is a function only of the radial distance $r$ outward from an axis, Poisson's equation attains the form

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dE}{dr}\right) = -\frac{\rho}{\epsilon} \qquad (A6\text{-}4)$$

# APPENDIX VII

# GRAPHICAL SYMBOLS FOR TUBE ELEMENTS

High-vacuum tube envelope............................

Gas-tube envelope (dot placed where convenient)..........

Directly heated cathode or heater........................

Cold cathode.........................................

Indirectly heated cathode...............................

Photoelectric cathode...................................

Pool cathode.........................................

Pool cathode with immersion ignitor......................

Grid.................................................

Plate or anode........................................

Deflecting, reflecting, or repelling electrode (electrostatic type)

Rectifier (arrow points in direction of low resistance or forward
 current flow).........................................

# APPENDIX VIII

# CIRCUIT NOTATION CONVENTIONS

Many voltage-notation conventions are found in texts on a-c circuit analysis. The one recommended by Reed and Lewis * is adopted in this text. This system of notation has a twofold purpose: (1) that of relating the notation on a circuit diagram with the instantaneous values of the potentials and currents in the circuit, and (2) that of relating the sinusoidal varying quantities with the equivalent sinor † representations and sinor diagrams. The essentials of the system are tabulated below.

1. Arrows are used to designate the reference direction of current on a circuit diagram. Positive instantaneous *numerical* value of a current indicates that the current is in the reference direction at this instant.

2. If a voltage symbol on a network diagram has no letter subscript, then a reference polarity is indicated by a plus sign near the positive reference terminal, *independent* of the reference current direction. Positive instantaneous numerical value of a voltage indicates that the reference and actual polarities are the same at this instant.

3. If a voltage symbol on a network diagram has a single letter subscript, such as $e_g$, then this represents the voltage of the point $g$ with respect to some reference point which has previously been designated. A positive instantaneous numerical value of this voltage indicates that the point $g$ is positive with respect to the reference point at this instant. For example, if the reference point is the cathode, and if the symbol $g$ denotes the grid, then $e_g$ represents the voltage of the grid with respect to the cathode. This also means that the symbol $e_g$ denotes the voltage drop from the grid to the cathode.

4. If a voltage symbol on a network diagram has double subscripts, such as $e_{gk}$, then this designates the reference polarity of that voltage with a plus sign at the first, or left-hand, subscript, regardless of whether the voltage is that of a generator or is that across a passive element. Thus the symbol

---

* M. B. REED and W. A. LEWIS, *Elec. Eng.*, **67**, 41, 1948.

† A number of different terms have been used to describe the rotating line segment, the projection of which generates a sinusoid. One will find in the literature, in addition to the term "vector," such terms as "rotating vector," "complex vector," "complexor," "phasor," and "sinor." These terms all serve to emphasize the fact that the line segments under consideration are not to represent vectors in the normal space of ordinary vector analysis. The term sinor (see W. R. LePage, *Elec. Eng.*, **68**, 1949) is used in this text as being the most descriptive.

$e_{gk}$ designates the voltage of the point $g$ with respect to the point $k$, or $e_{gk}$ is the voltage drop from point $g$ to point $k$.

5. No separate symbol is introduced for a voltage rise.

6. A current or voltage which varies sinusoidally with time is represented as a directed line segment which rotates *counterclockwise* with the angular frequency of the sinusoidally varying quantity. The letter symbols for these rotating line segments, which are called "sinors," are indicated in boldfaced type. The projection of the tip of the sinor on a refer-
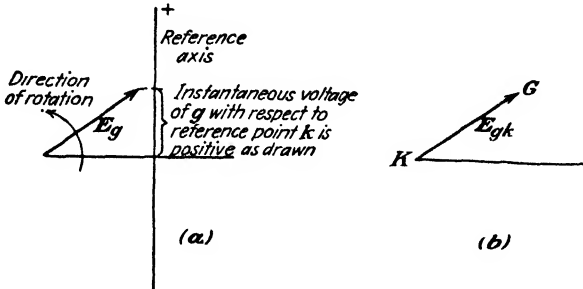


FIG. A8-1.

ence axis is proportional to the instantaneous magnitude of the voltage. As drawn in Fig. A8-1a, the sinor $\mathbf{E}_g$ represents a positive voltage of $g$ with respect to $k$. The symbol $\mathbf{E}_{gk}$ has the same meaning as $\mathbf{E}_g$, except that it states explicitly that $k$ is the reference point. Clearly, if a letter is used at each end of the sinor, then the tip, or head, must be marked $g$, and the tail must be marked $k$, although there is no real need for the arrowhead in this case. This is indicated in Fig. A8-1b, where $\mathbf{E}_{gk}$ represents the drop in voltage from the letter $g$ at the head of the arrow to the letter $k$ at the tail of the arrow.

7. The following are the consequence of elementary a-c circuit theory: The voltage drop in the direction of the current through a resistor is in phase with the current. The voltage drop in the direction of the current through an inductor leads the current by 90 deg. The voltage drop in the direction of the current through a capacitor lags the current by 90 deg.

The relationships for a simple series circuit containing resistance $R$ and inductance $L$ are indicated in Fig. A8-2. The voltage $\mathbf{E}$ is usually taken
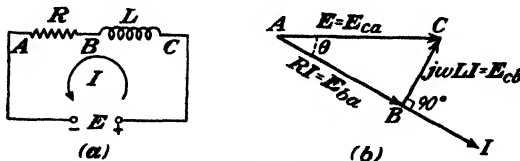


FIG. A8-2.

along the horizontal axis.   The current $\mathbf{I}$ must lag this voltage by some angle $\theta$, since the circuit is inductive.   The voltage drop from $C$ to $B$ leads the current by 90 deg.   The voltage drop from $B$ to $A$ is in phase with the current.   In the usual notation of complex algebra,

$$\mathbf{E}_{cb} = j\omega L\mathbf{I} \qquad \text{and} \qquad \mathbf{E}_{ba} = R\mathbf{I}$$

where $j = \sqrt{-1}$ and $\omega$ is the angular frequency.   Kirchhoff's voltage law yields

$$\mathbf{E}_{ca} = \mathbf{E}_{cb} + \mathbf{E}_{ba}$$

or

$$\mathbf{E} = j\omega L\mathbf{I} + R\mathbf{I}$$

# APPENDIX IX

# PLATE CHARACTERISTICS OF RECEIVING TYPE TUBES

(All curves through the courtesy of the RCA Manufacturing Co. and the General Electric Co.)

AVERAGE PLATE CHARACTERISTIC

TYPE 5U4-G
$E_f$ = 5.0 VOLTS A.C.

ONE PLATE ONLY

PLATE MILLIAMPERES

PLATE VOLTS D.C.

92C-6109
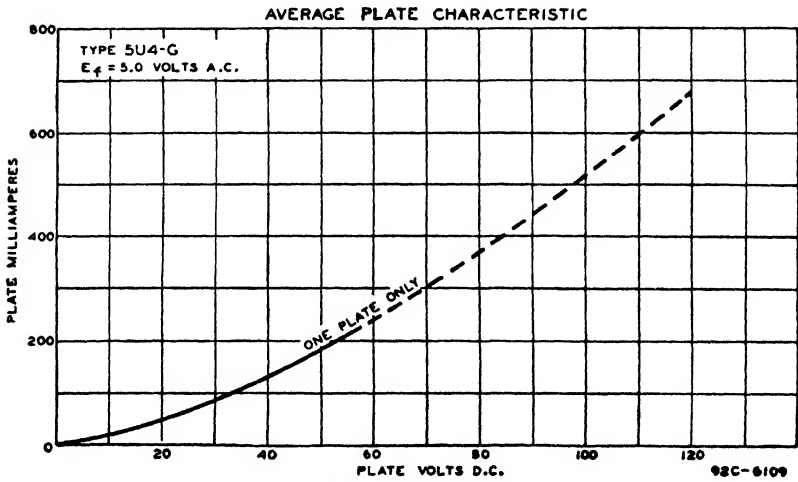
Fig. A9-1.  5U4-G diode.
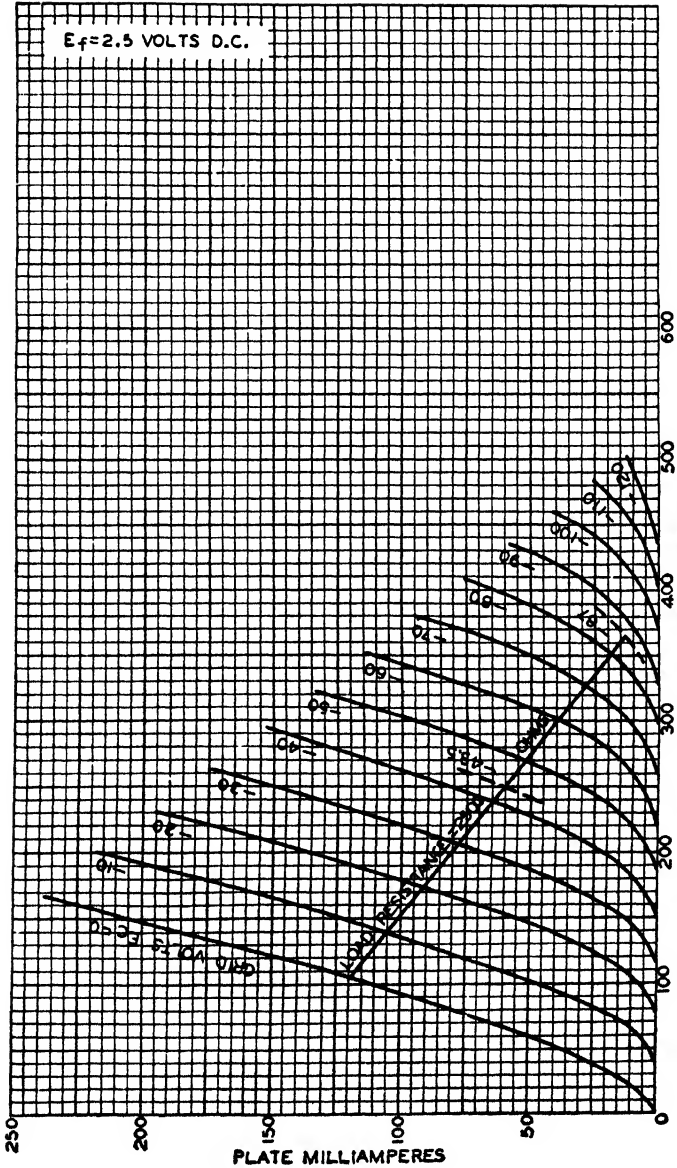
FIG. A9-2. 6A3 triode. Class A amplifier: $E_b = 250$ volts, $E_c = -45$ volts, $I_b = 60$ ma, $\mu = 4.2$, $r_p = 800$ ohms, $g_m = 5,250$ $\mu$mhos. Interelectrode capacitances: $C_1 = 7$ $\mu\mu$f, $C_2 = 5$ $\mu\mu$f, $C_3 = 16$ $\mu\mu$f.
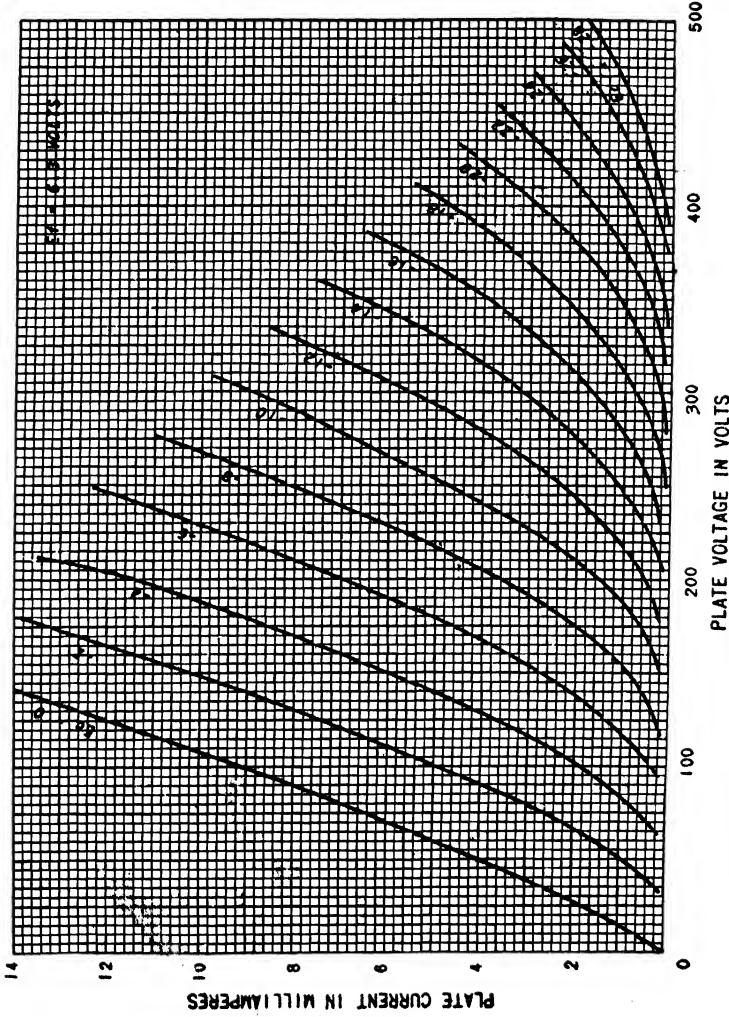
Fig. A9-3. 6C5 triode. Class A amplifier: $E_b = 250$ volts, $E_c = -8$ volts, $I_b = 8$ ma, $\mu = 20$, $r_p = 10{,}000$ ohms, $g_m = 2{,}000$ $\mu$mhos. Interelectrode capacitances: $C_1 = 3$ $\mu\mu$f, $C_2 = 11$ $\mu\mu$f, $C_3 = 2.0$ $\mu\mu$f.
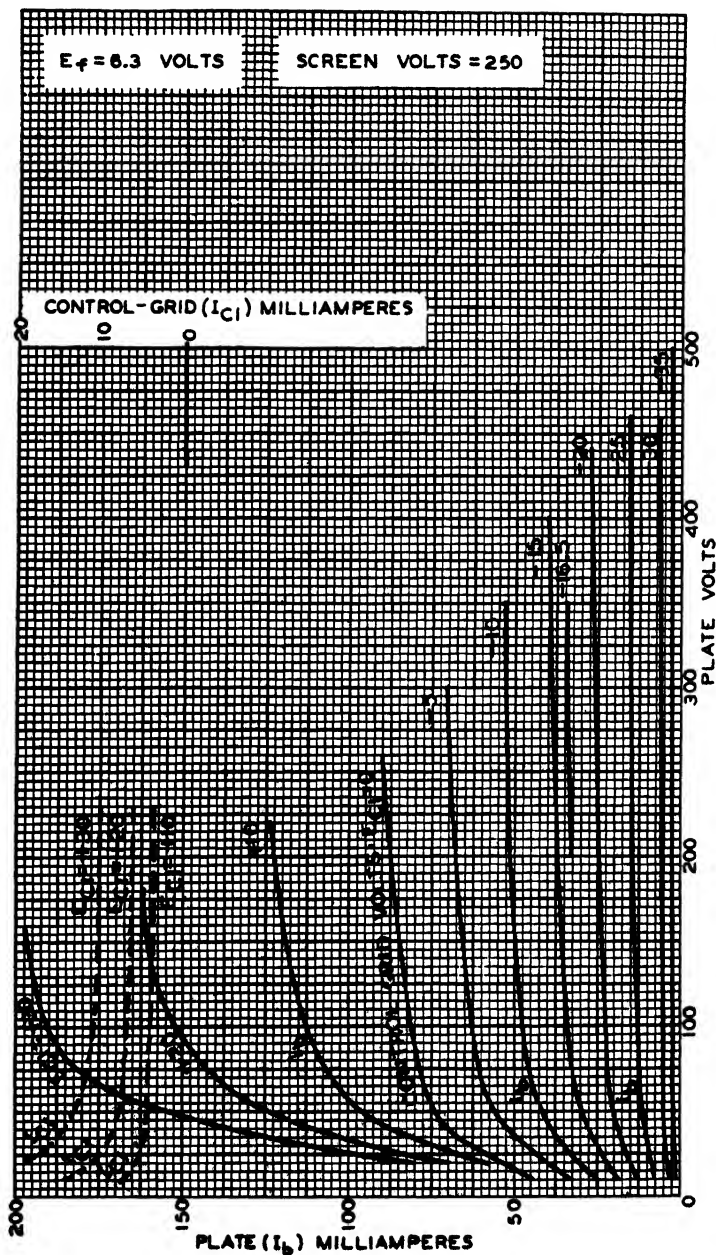
FIG. A9-4.  6F6 pentode.   Class A amplifier: $E_b$ = 250 volts, $E_{c2}$ = 250 volts, $E_c$ = −16.5 volts, $I_b$ = 34 ma, $r_p$ = 80,000 ohms approximately, $g_m$ = 2,500 μmhos.   Interelectrode capacitances: $C_1$ = 6.5 μf, $C_2$ = 13 μf, $C_3$ = 0.2 μf.
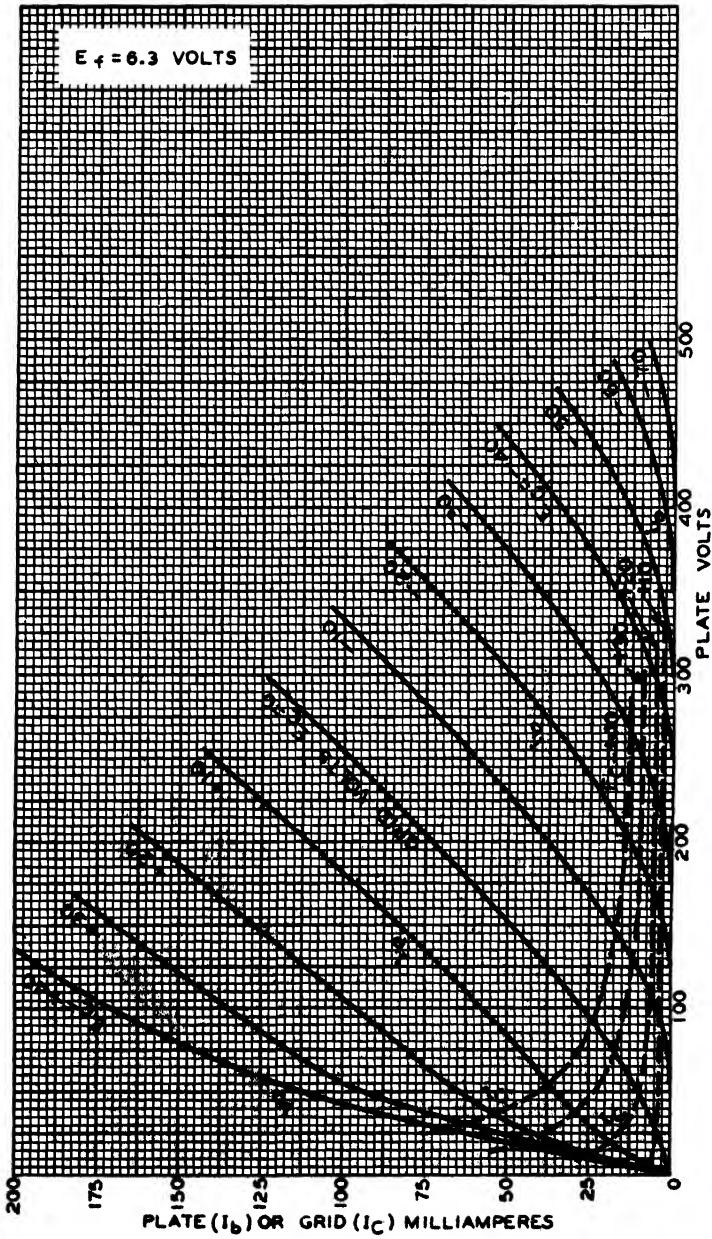
FIG. A9-5.  6F6 triode (screen connected to plate).  Class A amplifier: $E_b = 250$ volts, $E_c = -20$ volts, $I_b = 31$ ma, $\mu = 6.8$, $r_p = 2,600$ ohms, $g_m = 2,600$ $\mu$mhos.
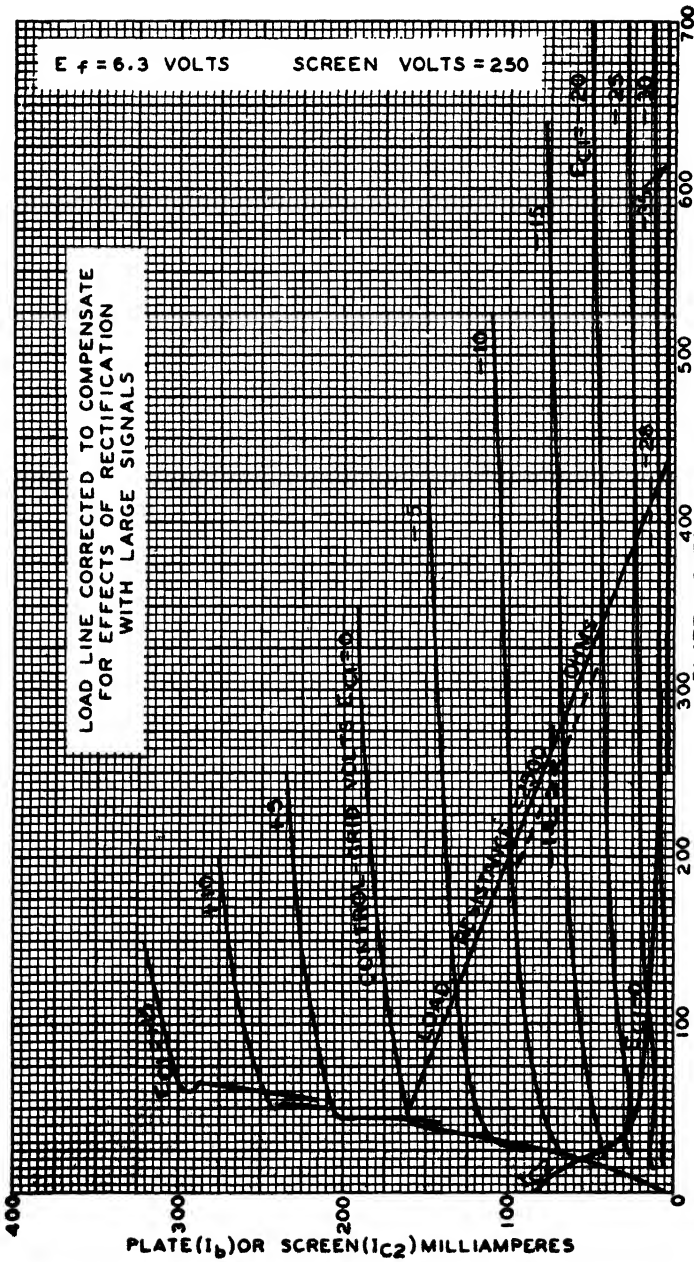
FIG. A9-6. 6L6 beam tube. Class A amplifier: $E_b = 250$ volts, $E_{c2} = 250$ volts, $E_c = -14$ volts, $I_b = 72$ ma, $I_{c2} = 5$ ma, $r_p = 22,500$ ohms approximately, $g_m = 6,000$ μmhos. Interelectrode capacitances: $C_1 = 10$ μμf, $C_2 = 12$ μμf, $C_3 = 0.4$ μμf.
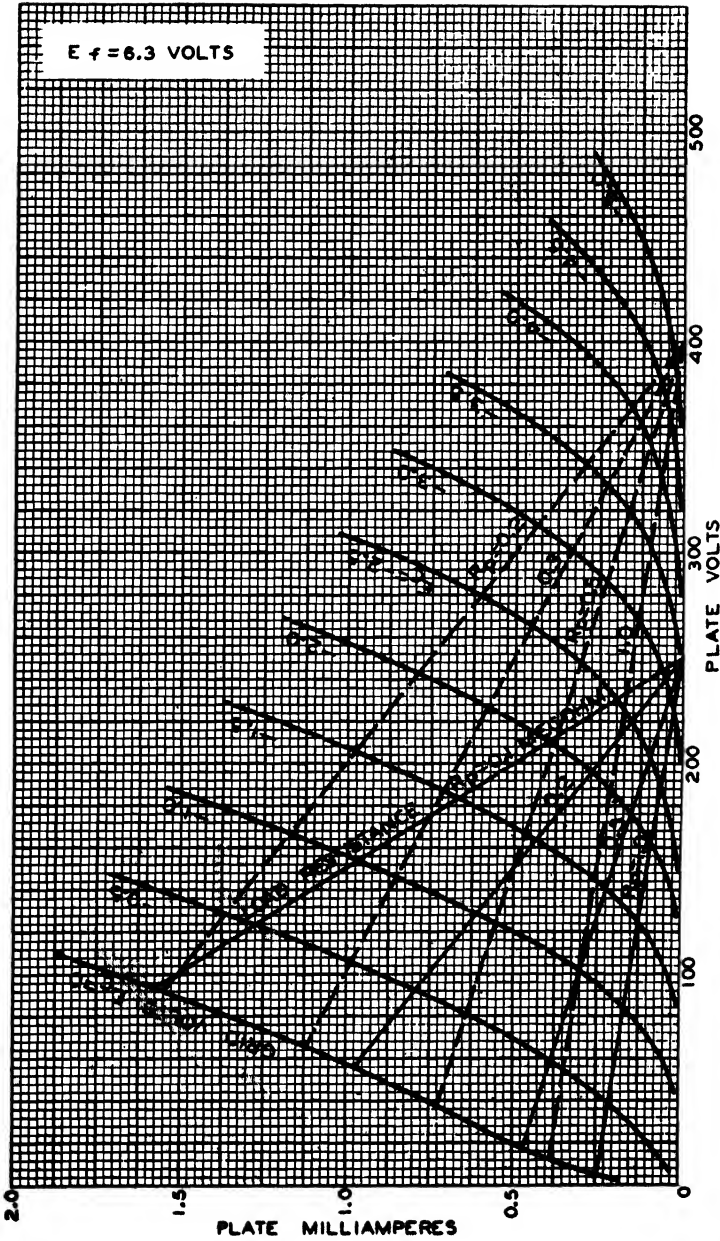
FIG. A9-7. 6SF5 triode. Class A amplifier: $E_b = 250$ volts, $E_c = -2$ volts, $I_b = 0.9$ ma, $\mu = 100$, $r_p = 66,000$ ohms, $g_m = 1,300$ $\mu$mhos. Interelectrode capacitances: $C_1 = 4.0$ $\mu$f, $C_2 = 3.6$ $\mu$f, $C_3 = 2.4$ $\mu$f.
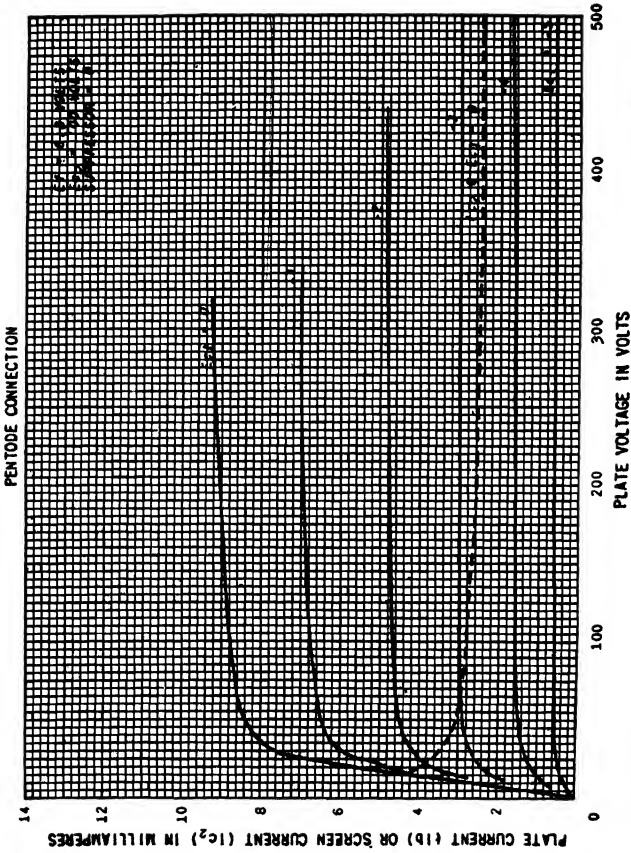
Fig. A9-8.  6SJ7 pentode.  Class A amplifier: $E_b = 250$ volts, $E_{c2} = 100$ volts, $E_c = -3$ volts, $I_b = 3$ ma, $I_{c2} = 0.8$ ma, $r_p = 1.0$ megohm approximately, $g_m = 1,650$ μmhos.  Interelectrode capacitances: $C_1 = 6.0$ μμf, $C_2 = 7.0$ μμf, $C_3 = .005$ μμf.
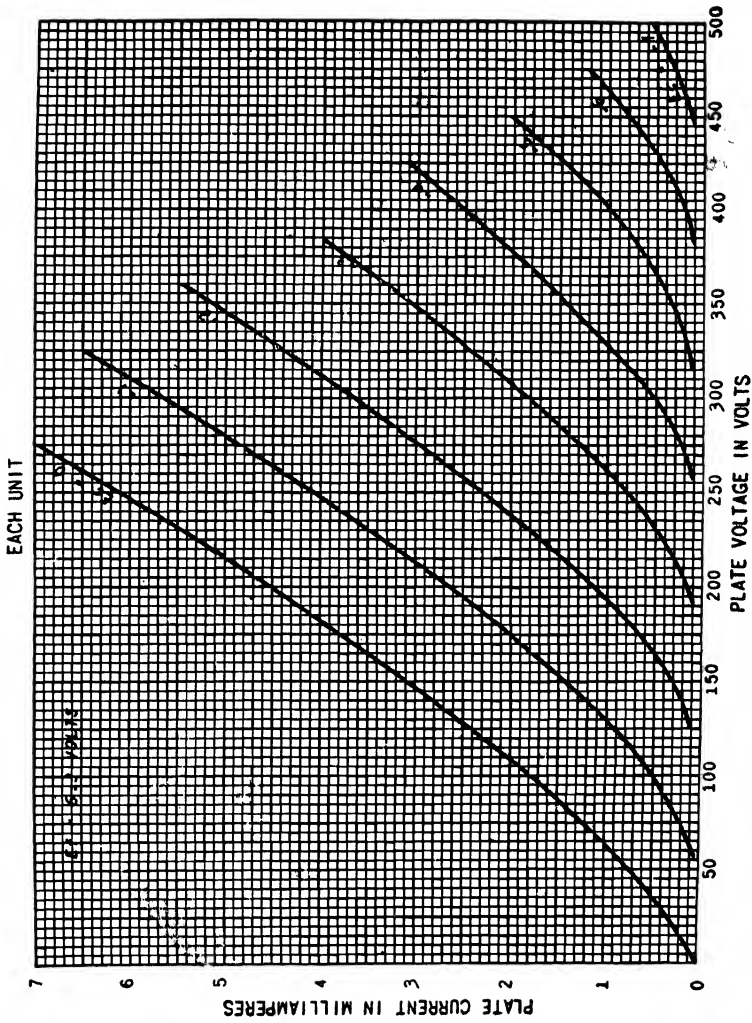
FIG. A9-9. 6SL7 dual triode. Class A amplifier: $E_b = 250$ volts, $E_c = -2$ volts, $I_b = 2.3$ ma, $\mu = 70$, $r_p = 44{,}000$ ohms, $g_m = 1{,}600$ μmhos. Interelectrode capacitances: Unit 1: $C_1 = 3.0$ μμf, $C_2 = 3.8$ μμf, $C_3 = 2.8$ μμf. Unit 2: $C_1 = 3.4$ μμf, $C_2 = 3.2$ μμf. $C_3 = 2.8$ μμf.
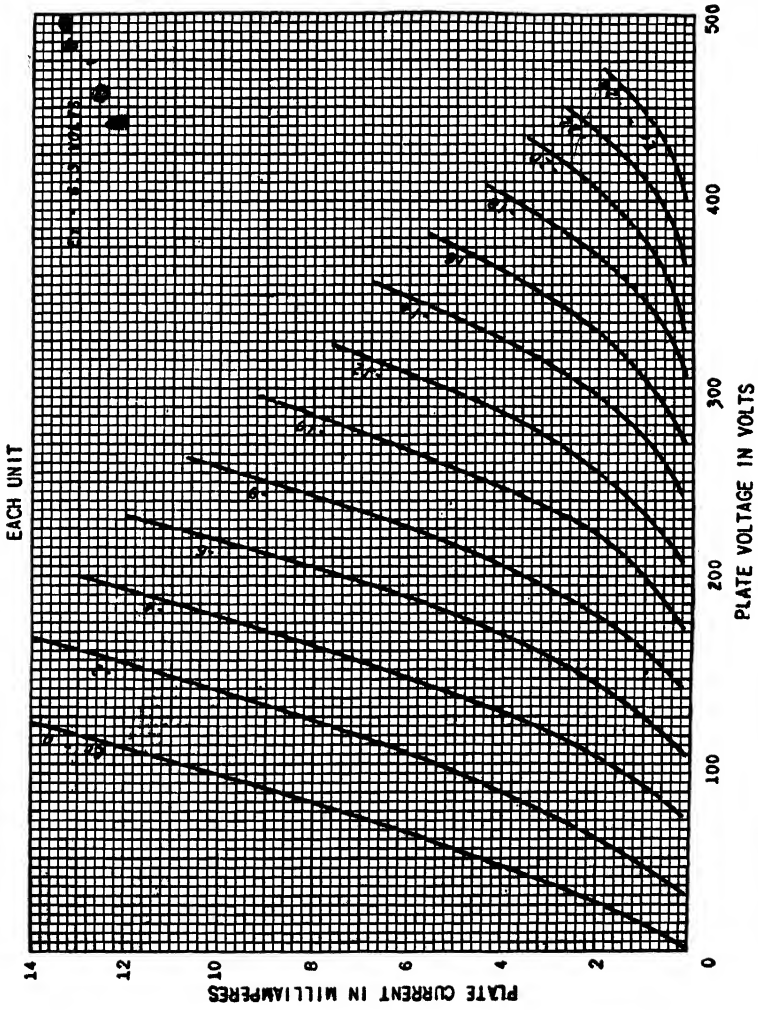
Fig. A9-10. 6SN7 dual triode. Class A amplifier: $E_b = 250$ volts, $E_c = -8$ volts, $I_b = 9$ ma, $\mu = 20$, $r_p = 7,700$ ohms, $g_m = 2,600$ $\mu$mhos. Interelectrode capacitances: Unit 1: $C_1 = 3.2$ $\mu\mu f$, $C_2 = 3.4$ $\mu\mu f$, $C_3 = 4$ $\mu\mu f$. Unit 2: $C_1 = 3.8$ $\mu\mu f$, $C_2 = 2.6$ $\mu\mu f$, $C_3 = 4$ $\mu\mu f$.

# INDEX