McGraw-Hill
ELECTRICAL AND ELECTRONIC
ENGINEERING SERIES

FREDERICK EMMONS TERMAN, *Consulting Editor*

VACUUM TUBES

McGRAW-HILL ELECTRICAL AND ELECTRONIC ENGINEERING SERIES

Frederick Emmons Terman, *Consulting Editor*
W. W. Harman and J. G. Truxal, *Associate Consulting Editors*

Ahrendt and Savant · Servomechanism Practice
Angelo · Electronic Circuits
Aseltine · Transform Method in Linear System Analysis
Bailey and Gault · Alternating-current Machinery
Beranek · Acoustics
Brenner and Javid · Analysis of Electric Circuits
Bruns and Saunders · Analysis of Feedback Control Systems
Cage · Theory and Application of Industrial Electronics
Cauer · Synthesis of Linear Communication Networks
Chirlian and Zemanian · Electronics
Clement and Johnson · Electrical Engineering Science
Cote and Oakes · Linear Vacuum-Tube and Transistor Circuits
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
Frank · Electrical Measurement Analysis
Geppert · Basic Electron Tubes
Glasford · Fundamentals of Television Engineering
Greiner · Semiconductor Devices and Applications
Hammond · Electrical Engineering
Hancock · An Introduction to the Principles of Communication Theory
Happell and Hesselberth · Engineering Electronics
Harman · Fundamentals of Electronic Motion
Harrington · Introduction to Electromagnetic Engineering
Harrington · Time-harmonic Electromagnetic Fields
Hayt · Engineering Electromagnetics
Hill · Electronics in Engineering
Johnson · Transmission Lines and Networks
Koenig and Blackwell · Electromechanical System Theory
Kraus · Antennas
Kraus · Electromagnetics
Kuh and Pederson · Principles of Circuit Synthesis
Ledley · Digital Computer and Control Engineering
LePage · Analysis of Alternating-current Circuits
LePage · Complex Variables and the Laplace Transform for Engineers
LePage and Seely · General Network Analysis

LEY, LUTZ, AND REHBERG · Linear Circuit Analysis
LYNCH AND TRUXAL · Introductory System Analysis
MILLMAN · Vacuum-tube and Semiconductor Electronics
MILLMAN AND SEELY · Electronics
MILLMAN AND TAUB · Pulse and Digital Circuits
MISHKIN AND BRAUN · Adaptive Control Systems
MOORE · Traveling-wave Engineering
PETTIT · Electronic Switching, Timing, and Pulse Circuits
PETTIT AND MCWHORTER · Electronic Amplifier Circuits
PFEIFFER · Linear Systems Analysis
PLONSEY AND COLLIN · Principles and Applications of Electromagnetic Fields
REZA · An Introduction to Information Theory
REZA AND SEELY · Modern Network Analysis
ROGERS · Introduction to Electric Fields
RUDENBERG · Transient Performance of Electric Power Systems
RYDER · Engineering Electronics
SCHWARTZ · Information Transmission, Modulation, and Noise
SEELY · Electron-tube Circuits
SEELY · Electrical Engineering
SEELY · Introduction to Electromagnetic Fields
SEELY · Radio Electronics
SEIFERT AND STEEG · Control Systems Engineering
SISKIND · Direct-current Machinery
SKILLING · Electronic Transmission Lines
SKILLING · Transient Electric Currents
SPANGENBERG · Fundamentals of Electron Devices
SPANGENBERG · Vacuum Tubes
STEVENSON · Elements of Power System Analysis
STEWART · Fundamentals of Signal Theory
STORER · Passive Network Synthesis
STRAUSS · Wave Generation and Shaping
TERMAN · Electronic and Radio Engineering
TERMAN AND PETTIT · Electronic Measurements
THALER · Elements of Servomechanism Theory
THALER AND BROWN · Analysis and Design of Feedback Control Systems
THOMPSON · Alternating-current and Transient Circuit Analysis
TOU · Digital and Sampled-data Control Systems
TRUXAL · Automatic Feedback Control System Synthesis
VALDES · The Physical Theory of Transistors
WILLIAMS AND YOUNG · Electrical Engineering Problems

VACUUM TUBES

KARL R. SPANGENBERG

PROFESSOR OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY

NEW YORK TORONTO LONDON
McGRAW-HILL BOOK COMPANY, INC.
1948

VACUUM TUBES

PRINTED IN THE UNITED STATES OF AMERICA

X

59880

TO MY COLLEAGUES AND STUDENTS

PREFACE

This book is the outgrowth of a course in vacuum-tube design given for many years at Stanford University to senior and graduate students in electrical engineering and physics. It is concerned with the determination of vacuum-tube characteristics in terms of the electron action within the tube. The book attempts to bridge the gap between the physical laws that lie behind the electron behavior and the external characteristics of the tubes themselves.

It is hoped that the point of view taken will be acceptable to both physicists and engineers. The development of the physical laws involved is indicated, after which emphasis is placed upon their description and utilization. Although this book cannot pretend to give much design information, the attempt has been to include enough of the basic relations, physical data, and significant references to make it a useful reference source to vacuum experimenters and tube designers.

Vacuum tubes may seem a rather special subject to which to restrict the material in a book. Actually this is not so. In preparing the book so much material was collected that the contents had to be restricted to first-order effects. It is felt that although engineers and physicists working with vacuum tubes are primarily concerned with the utilization of already developed tubes, the successful application of these tubes is greatly enhanced by a knowledge of their limitations and an understanding of the origin of their characteristics. This is particularly true since there are many occasions when it is desired to use tubes under conditions different from those specified by the manufacturer. Under these conditions it is imperative to know how far one may depart from recommended operating conditions without exceeding some design limitation of the tube. This, in turn, requires a knowledge of how the tube operates.

Circuits and tube applications are so completely covered in the textbook and periodical literature that no effort has been made to include information on these subjects. Only in the case of ultra-high-frequency tubes where the tube cannot be completely separated from the circuit have circuit considerations been included.

The author is indebted to many people for assistance rendered in the preparation of this book. He is particularly indebted to Dr. F. E. Terman, dean of the Stanford School of Engineering, who was a constant source of inspiration and encouragement, and who made many valuable suggestions and gave much direct assistance in checking the work. The author is also indebted to Prof. Paul Kirkpatrick, head of the Physics Department at Stanford, for suggestions on the material of Chaps. 3 to 6 and 9; to Prof. L. Marton for suggestions on the material of Chaps. 13 to 15 and 20; and to C. V. Litton for much information and suggestions relative to Chap. 21. He is indebted to Evelyn G. Sarson, who typed a large part of the manuscript in its final form. O. O. Pardee and Will Harman assisted in the correction of the entire work. Lastly, the author is more than a little indebted to his wife, who personally typed much of the manuscript and was a source of constant assistance.

KARL R. SPANGENBERG

PALO ALTO, CALIF.
January, 1948

CONTENTS

CHAPTER 1

INTRODUCTION

Vacuum tubes are found as basic or auxiliary elements in numerous technical devices now in use. They are indispensable in communication systems and industrial control. Their development has facilitated advances in the fields of power and transportation. Without the vacuum tube we should be back in the days of the gravity-cell telegraph and the ringer telephone.

In the United States the number of vacuum tubes in use is several times the number of human beings and household pets. The 50,000,000 radio sets manufactured in the United States in the year 1947, alone contained more vacuum tubes than the adult population of the country. Associated with the 25,000,000 telephones and 120,000,000 miles of telephone and telegraph wire in the United States are many more vacuum tubes. Various industrial devices include almost as many more. The United States uses nearly half the world's total of vacuum tubes.

One may conclude that there are many vacuum tubes in use. They must be of some importance. They are.

1.1. Devices Using Vacuum Tubes. This book is more concerned with the properties and functions of vacuum tubes than with the systems utilizing these properties. However, it is well to be reminded of the extent of vacuum-tube applications and the degree to which we are dependent upon them. The following devices are totally dependent upon vaccum tubes.

Radio Receivers. These are too well known to require much description. They range from portable receivers the size of a brick and capable of receiving local broadcast stations to large-size all-wave receivers capable of picking up a signal stronger than the noise level from any point on the globe. Even the smallest receivers use 4 or 5 vacuum tubes. The average home receiver has about 7 tubes. An all-wave receiver may have 20 or more tubes.

Radio Transmitters. Transmitters range from portable walkie-talkie sets to large power-broadcast and short-wave stations. In output power they vary from 0.1 watt to hundreds of kilowatts. In frequency they may range from 100 kc to 60,000 mc. The short-wave transmitters are capable of producing an audible signal at any point on the earth's surface.

Transmitters may use voice or code. They may incorporate static-elimination or secrecy features in their operation. A small transmitter may use only a few vacuum tubes. The largest transmitters may use 50 or more tubes.

Long-distance Wire Telephones. The connections between telephone stations on the same continent are effected by wire transmission lines rather than by radio. When the distance between telephone stations is large, it is necessary to amplify the speech energy about every 16 miles for cables and every 50 miles for open-wire lines. Each speech amplifier contains several vacuum tubes and amplifies the speech power from about 10 microwatts to about 1 milliwatt, a power amplification of 100. Thus a telephone call from San Francisco to New York passes through 30 or more speech amplifiers.

Television Systems. Television systems achieve the modern miracle of reproducing a visual scene at a point remote from the original. This is done entirely with vacuum tubes and electrical-circuit elements. No mechanical devices are needed. In its present stage of development the reproduced picture as viewed from 6 ft on an 8-in. cathode-ray-tube screen is as good as a motion picture seen from the first row of the balcony. Each television transmitter contains hundreds of vacuum tubes, including a special camera tube. Every television receiver contains 20 or more tubes, including a special viewing tube.

Measurement Devices. Electronic measurement devices are too numerous to mention. Quantities that can be measured, besides all the electrical quantities, are color, weight, light intensity, odor, time interval, and many others. In fact, it can be said that any quantity which can be measured at all can probably be measured by electronic means.

Industrial Control. The number of electronic industrial-control devices is legion. They include counting circuits, sorting systems, illumination-control systems, welding-control devices, and liquid- and gaseous-flow regulators. Typical devices are those which automatically regulate temperature or humidity. All these devices have their primary dependence upon the vacuum tubes in them.

In addition to the above devices, which are totally dependent upon vacuum tubes, there are many others that have acquired a strong dependence upon electronic devices. Thus all commercial flying makes constant use of radio communications to keep posted on the weather and on terminal traffic and to keep ground stations posted on plane positions as well as to guide the planes directly. The invasion of other fields by electronics has already been considerable and is bound to be greater in time to come.

1.2. Functions of Vacuum Tubes. Although the applications of vacuum tubes are almost infinite, the specific functions that vacuum tubes can perform by virtue of their own properties are relatively few. It is these few fundamental functions and their combinations that give rise to the numerous applications.

A list of the functions of vacuum tubes is bound to be an arbitrary one since the tube cannot function by itself without an associated circuit. However, some of the jobs that vacuum tubes can perform are so fundamental that they may be considered properties of the tube itself, independent of the associated circuits.

The principal functions that may be performed by vacuum tubes are listed below.

Rectification. Vacuum tubes are able to convert alternating currents to direct currents. This is known as "rectification." Rectification is an inherent property of vacuum tubes because current can flow in only one direction from a source of electrons.

If a sinusoidal wave of voltage is applied to a vacuum tube of the right type, current will flow in only one direction, giving rise to a succession of half-wave pulses all of the same polarity. It is possible to connect another like tube to insert half-wave pulses of the same polarity between the pulses of the first tube. The average of these pulses constitutes a direct current; the other frequency components are rejected by a filter circuit.

Rectification is important because electronic devices operate best on direct current, while power is usually generated and transmitted in alternating form. It is thus necessary to convert, or rectify, the a-c power to d-c power.

Amplification. The amplification of voltage or power is the outstanding function that vacuum tubes are able to perform. With the exception of the mechanical torque amplifier, no other device can do anything like it. Strictly speaking, the vacuum tube does not amplify power but rather controls the flow of a relatively large amount of power from one source with a small amount of power from another source. The British use the expression "electric valve" for certain types of electron tubes. This term is really better than ours, for it indicates the nature of the amplifying action.

Oscillation. The generation of high-frequency alternating currents, or oscillation, is another remarkable function that vacuum tubes can perform. Oscillation is obtained by causing part of the output of an amplifier to excite the amplifier and thus make the device self-excited and self-sustaining. Tubes can be built that will produce oscillations at frequencies as low as 1 cycle per sec, while other tubes can be built that will oscillate at frequencies as high as 60,000 mc per sec.

Frequency Conversion. Vacuum tubes are able to shift the frequency of a wave. This they are able to do by an electrical "beat" action.

Thus a wave of a given frequency can be mixed with a wave of another frequency in a vacuum tube, and among the products of the interaction is found the difference of the two frequencies. If one of the original waves had certain effects associated with it, these same effects are associated with the difference frequency. The beat action results from the nonlinear characteristics of the vacuum tube.

Modulation. The transmission of intelligence by radio waves or by certain types of wire telephony requires the use of frequencies higher than those audible. It is necessary to superimpose the audible frequencies upon the higher transmitted frequency. This superimposition is known as "modulation." Modulation is best performed by vacuum tubes.

Basically, modulation takes the form of varying some property of the r-f wave at the audible rate. The commonest form of modulation varies the amplitude of the r-f wave in accordance with the intelligence to be transmitted. This is known as "amplitude modulation." Frequency modulation is also common.

Detection. Detection is the inverse of modulation and is sometimes known as "demodulation." It is the process of extracting the intelligence from the modulated wave. In the case of amplitude modulation the detection may be effected by rectifying the r-f wave and then utilizing the average value of the rectified wave, since it follows the amplitude variations in magnitude. Detection of modulated radio signals is best performed by vacuum tubes over most of the range of radio frequencies.

Light-image Production. It is possible for vacuum tubes to convert part of their output energy into visible light. This is done in cathode-ray tubes in which a stream of electrons is caused to hit a fluorescent screen, causing light to be emitted. The cathode-ray tube can be used for viewing wave forms and for doing many other wonderful things, including the reproduction of visual scenes. The fundamental property involved here is the conversion of electrical energy into visual energy.

Photoelectric Action. Vacuum tubes can be made that will convert light energy into electrical energy. This is possible by virtue of the photoelectric effect, which is the emission of electrons from certain surfaces when illuminated with visual energy. The liberated electrons constitute an electric current whose measure is related to the frequency and intensity of the exciting light. Tubes making use of this principle are known as "photoelectric tubes." The photoelectric tube is one of the tubes most extensively used in industrial-control systems.

The above paragraphs have given a bird's-eye view of the functions of vacuum tubes. The reader is probably familiar with all the above functions, which are now commonly encountered in everyday life. The rest of the book is devoted to the description and explanation of the characteristics of the vacuum tubes themselves.

CHAPTER 2

BASIC TUBE TYPES

THE electronic engineer has about a dozen *types* of vacuum tube he can call upon for his high-frequency and industrial-control circuits. This is a surprisingly small number of distinct tube types. The small number of types is balanced, however, by the large number of forms in which each type may appear, as determined by the required power capacity and frequency range.

The purpose of this chapter is to list the basic types and their fundamental characteristics as a prelude to a detailed study of their characteristics and the physical laws from which these are derived.

2.1. Vacuum Diode. The vacuum diode is a two-electrode vacuum tube. One electrode acts as an emitter of electrons and is called the "cathode." The other electrode acts as a collector of electrons and is called the "anode" or "plate." The emitter may be either directly or indirectly heated. In physical form the vacuum diode may vary from a small metal tube to a large glass rectifier tube.

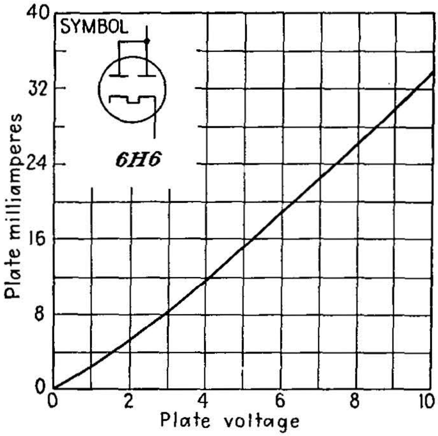


FIG. 2.1.—Plate-current—plate-voltage characteristics of a diode.

The current-voltage characteristics of a typical diode are shown in Fig. 2.1. The current follows a three-halves-power law of voltage over the normal range of operation. At high values of plate voltage or at low values of heater current the plate current tends to be limited by the cathode emission and to increase only very slowly with plate voltage.

The most useful property of the diode is that it passes current only in one direction. This property makes the diode useful as a detector and as a rectifier for d-c power supplies.

2.2. Vacuum Triode. A vacuum triode is a three-electrode tube containing an emitting electrode called the "cathode," a control electrode called the "grid," and a current-collecting electrode called the "anode" or "plate."

The emitting electrode may be an indirectly heated oxide cathode, an oxide-coated filament, or a filament of tungsten or thoriated tungsten.

The control electrode, usually in the form of a grid of fine wire, surrounds the emitter and is in turn surrounded by the plate in the commonest form of triode. By virtue of its proximity to the cathode the grid is able to influence the electrostatic field at the cathode to a greater extent than can the plate, and thus it is able to control the flow of current from the cathode. The grid is usually operated on a slight negative potential so that the electrons will pass between the grid wires without hitting the wires themselves.

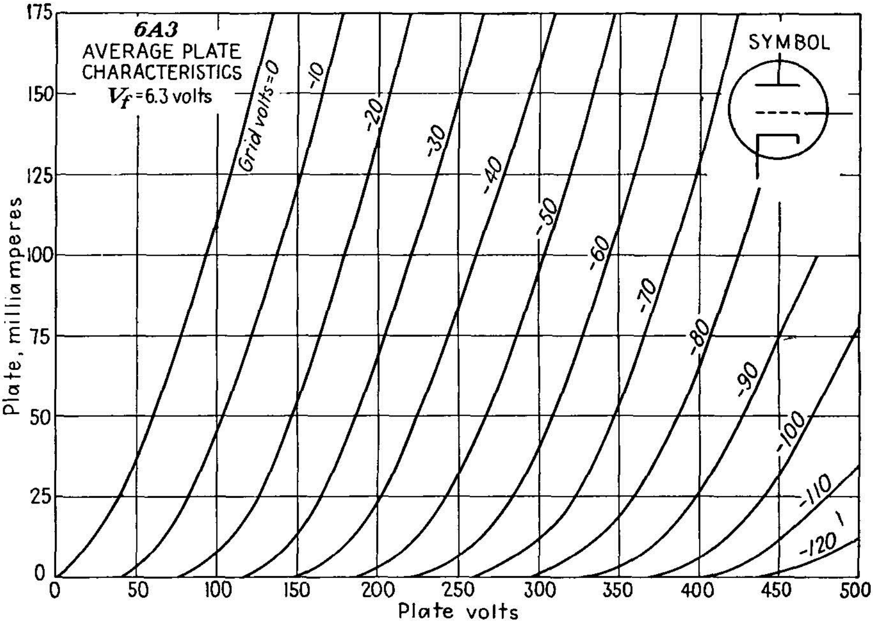


FIG. 2.2.—Plate-current—plate-voltage characteristics of a triode.

Some typical characteristics of a triode illustrating the variation of plate current with plate voltage for various fixed values of grid voltage are shown in Fig. 2.2. The plate current increases if either grid or plate voltage is increased. The increase in plate current for a given increase in grid voltage is always much larger than the increase in plate current for the same increase in plate voltage.

The relative effectiveness of the plate and grid potentials in controlling the plate current is known as the amplification factor of the tube (mu; symbol μ). The amplification factor is the *maximum* amplification that can be obtained by using the tube as an amplifier. With triodes the useful amplification is about two-thirds of the amplification factor.

Study of the family of curves of Fig. 2.2 shows that all the curves are alike in shape and further are somewhat similar to the characteristic of a diode. This is true in that the plate current of a triode is found to vary nearly as the three-halves power of an equivalent voltage which is the sum of the plate voltage divided by mu and the grid voltage.

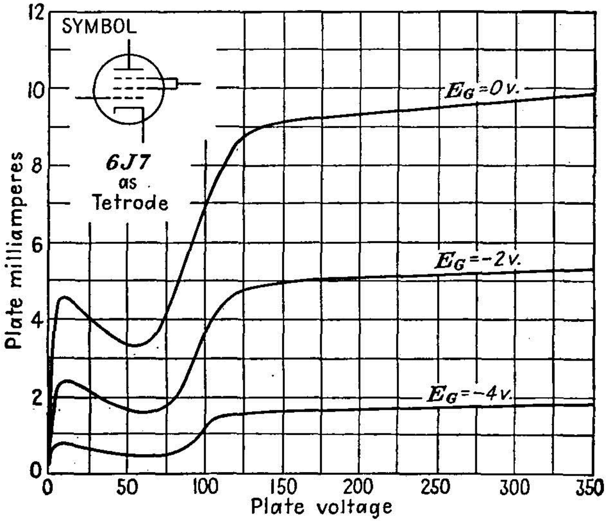


FIG. 2.3.—Plate-current—plate-voltage characteristics of a screen-grid tube. V_f = 6.3 volts, V_{sg} = 100 volts.

Triodes have their greatest use as power amplifiers. They are also used extensively in control applications wherever a small voltage is wanted to control an appreciable amount of current.

2.3. Screen-grid Tube. The screen-grid tube is a four-element vacuum tube. The four elements are cathode, control grid, screen grid, and plate. The electrode construction is similar to that of the triode except that an extra grid of mesh a little coarser than that of the control grid is inserted between the control grid and the plate.

The screen-grid tube is the historical predecessor of the pentode. Its invention was the result of an effort to overcome a limitation of the triode. Triodes do not work well as amplifiers of high frequencies, for the high interelectrode capacity between plate and grid causes the tube to regenerate and oscillate. In the screen-grid tube the capacity between the control grid and plate is reduced by inserting the extra grid, known as the "screen grid," between these elements. The insertion of the screen grid and its operation at a constant potential succeeded in producing the low control-grid—plate capacity desired but caused distortions in the plate-current—plate-voltage characteristics, for the new electrode arrangement permitted secondary electron flow between plate and screen grid. This detrimental effect was overcome in the pentode by the addition of a coarse-mesh suppressor grid between screen grid and plate.

The screen-grid tube is usually operated with cathode near ground potential, control grid at a small negative potential, and screen grid and plate at a medium and high positive potential, respectively. Some typical screen-grid-tube plate-current—plate-voltage characteristics are shown in Fig. 2.3. The dips in the low-voltage portion of the curves are the result of secondary electron current flowing from plate to screen. The low slope of the high-voltage portion of the curves results from the fact that the cathode is screened from the plate by the screen grid as well as by the control grid, and hence the magnitude of the plate current is increased only a little by an increase in plate voltage. Screen-grid tubes have been rendered virtually obsolete by the development of the pentode and some special tetrodes not subject to the tremendous distortions of current characteristics by secondary emission. Screen-grid tubes may be used as a-f and r-f amplifiers. They are also occasionally used in laboratory apparatus in which it is desirable to utilize the negative resistance characteristic which is available at the points on the current characteristics where the slope is negative.

2.4. Pentode. The pentode is a five-element high-vacuum tube. The five electrodes, in the order in which they occur in the tube, are cathode, control grid, screen grid, suppressor grid, and plate. In normal operation the cathode is operated near ground potential, the control grid at a small negative potential, the screen grid at a relatively large positive potential, the suppressor grid at cathode potential, and the plate at the screen potential or a more positive potential.

Some typical plate-current—plate-voltage curves of a pentode are shown in Fig. 2.4. In these it is seen that the insertion of the suppressor grid at cathode potential between screen grid and plate has eliminated the distortions in the characteristic observed in the case of the screen-grid tube. This it does by causing a negative potential gradient at both the screen grid and plate, which suppresses the secondary electrons from these electrodes. The slope of the plate-current characteristic for high plate voltages is even less than in the screen-grid tube, for there is another screening grid between plate and cathode in the pentode. The result of this high screening action is to make the amplification factor of the pentode extremely high, of the order of 1,000 or more, and to give the tube a high effective resistance in the plate circuit. The pentode is, in fact, very nearly a constant-current device. The variation of plate current with grid voltage, which is measured by a factor known as the "grid-plate transconductance" or, more commonly, the "mutual conductance" of the tube, is about the same as in the triode. Only about one-tenth of the high amplification factor of the pentode can be realized in amplifier operation. However, the reduced plate—control-grid capacity makes the pentode a better tube in voltage-amplifier applications.

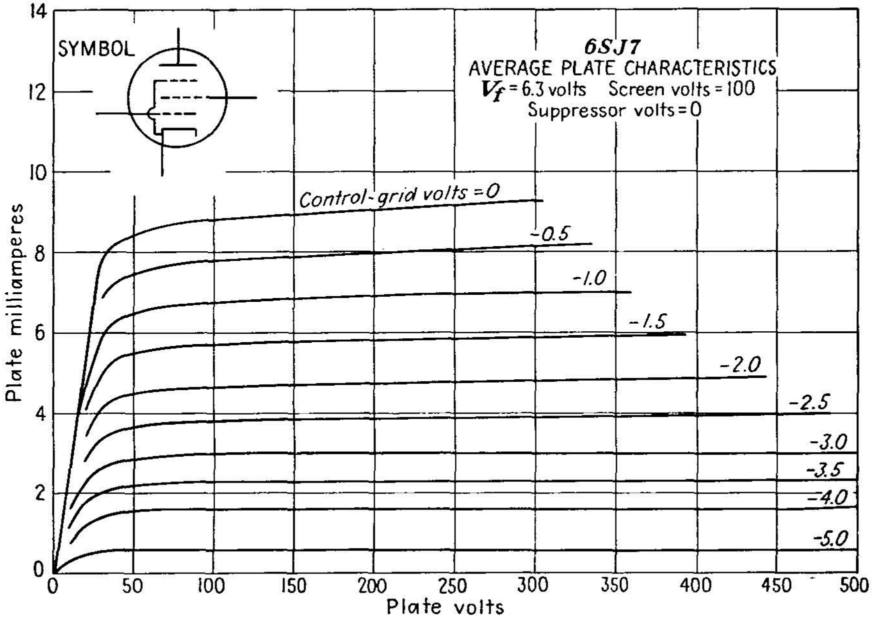


FIG. 2.4.—Plate-current—plate-voltage characteristics of a pentode.

The pentode is a versatile tube. It can be connected to give diode, triode, and screen-grid as well as pentode action. It is available in constant- and variable-mu forms. It is probably the most extensively used tube in low-power applications. There are probably more pentodes in use today than any other type of electron tube. A cutaway drawing of a pentode showing the electrode structure is given in Fig. 2.5.

2.5. Beam-power Tube. The beam-power tube is a special type of tetrode. It is designed so that the electrons move from cathode to plate in dense sheets. This effect is achieved by making the control grid and screen grid of the same pitch and aligning the grid wires. The electrode structure of the tube is shown in Fig. 2.6.

The effect of the dense current sheets between the screen grid and plate is to depress the potential between these two electrodes within the tube because of the high concentration of negative charge. The potential between screen grid and plate is depressed enough so that secondary electron flow from plate to grid is suppressed without the aid of a sup-

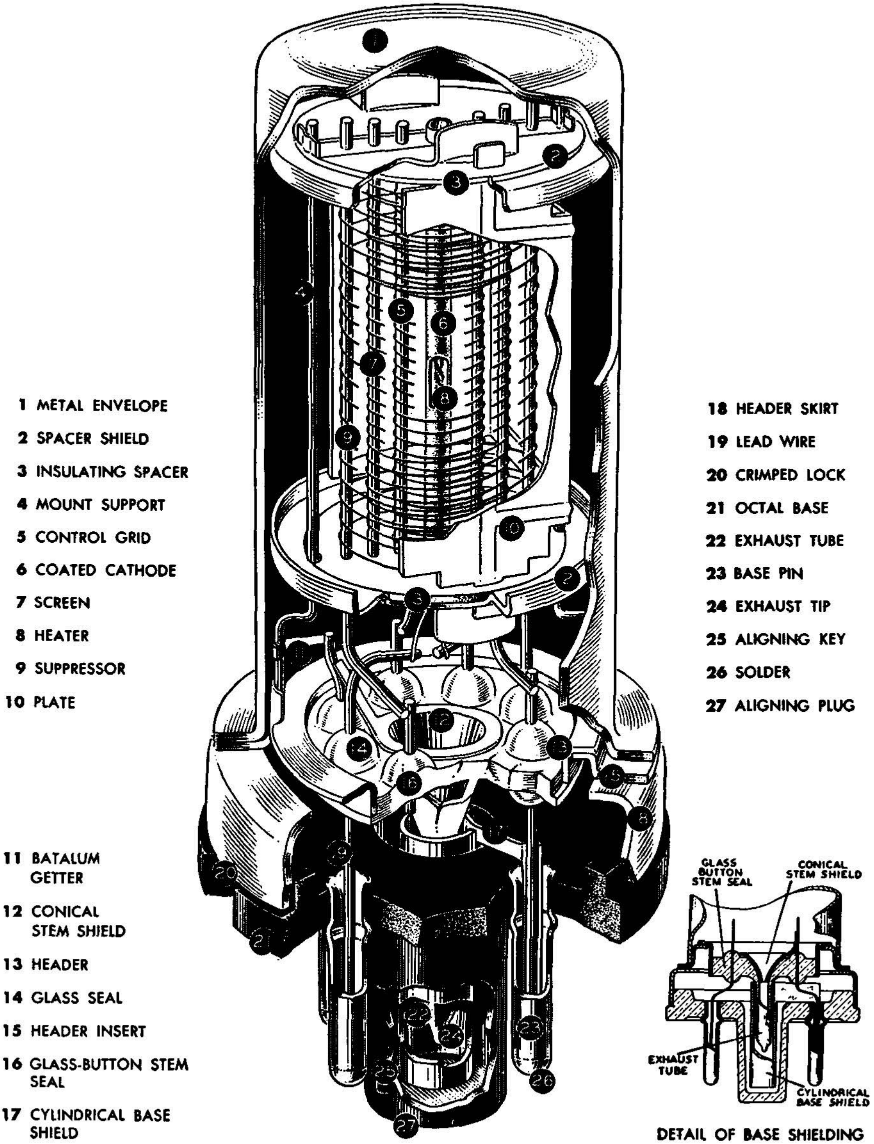


FIG. 2.5.—Cutaway picture of a single-ended metal-envelope pentode.

pressor grid. Thus the tube represents another solution to the problem of overcoming the distortions in the current characteristics of the ordinary screen-grid tube.

The plate-current—plate-voltage characteristics of a beam-power tube are shown in Fig. 2.7. It is seen that these characteristics are free of the dip in the shoulder due to secondary electron flow. The distinctive features of the beam-power tube's characteristics as contrasted with the pentode characteristics are that the plate current rises much more rapidly at low plate potentials and the condition of complete transmission of current to the plate is reached at a lower plate potential. The plate current rises rapidly because the high space-charge density blocks the flow of electrons to the plate at low plate potentials, and this blocking action stops quite abruptly as the plate potential is increased. In the beam-power tube, complete transmission of current passed by the screen grid to the plate occurs when the plate potential has risen to about 20 per cent of the screen-grid potential, whereas in the pentode the transmission is not complete until the plate potential has risen to about 50 per cent of the screen-grid potential. This results from the behavior of the individual electrons, which, in the beam-power tube, are more uniform in direction and velocity than in the pentode, in which the electrons are strongly deflected by the suppressor grid.

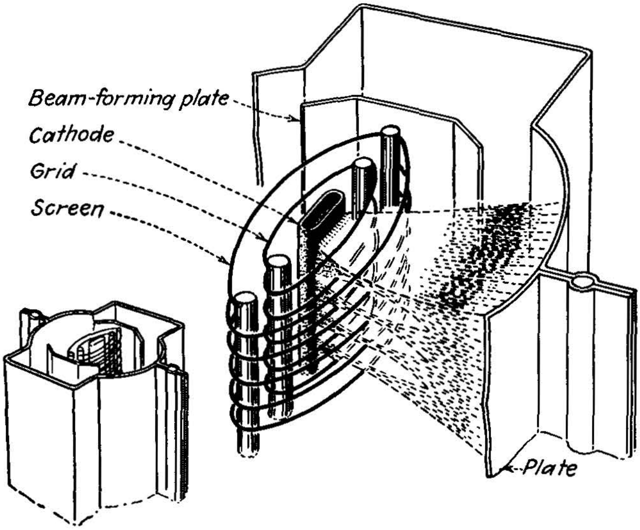


FIG. 2.6.—Cutaway view of the electrode arrangement in a beam-power tube. (*Courtesy of RCA.*)

The beam-power tube is made in small and medium-size metal tubes and in a medium-size glass tube. The beam-power tube is used in many ways. It is extensively employed as an audio-power amplifier tube and also as a r-f amplifier and oscillator tube.

2.6. Cathode-ray Tubes. The cathode-ray tube is in a class by itself among the vacuum tubes. It makes use of the geometrical form rather than the intensity of its electron stream and converts the energy of its electron stream into a visual indication. In its commonest application the cathode-ray tube uses its electron beam to show the shape of an applied voltage wave as a light trace upon a fluorescent screen. The cathode-ray tube is an electronic oscilloscope that produces on a screen a light spot that can be deflected in two dimensions.

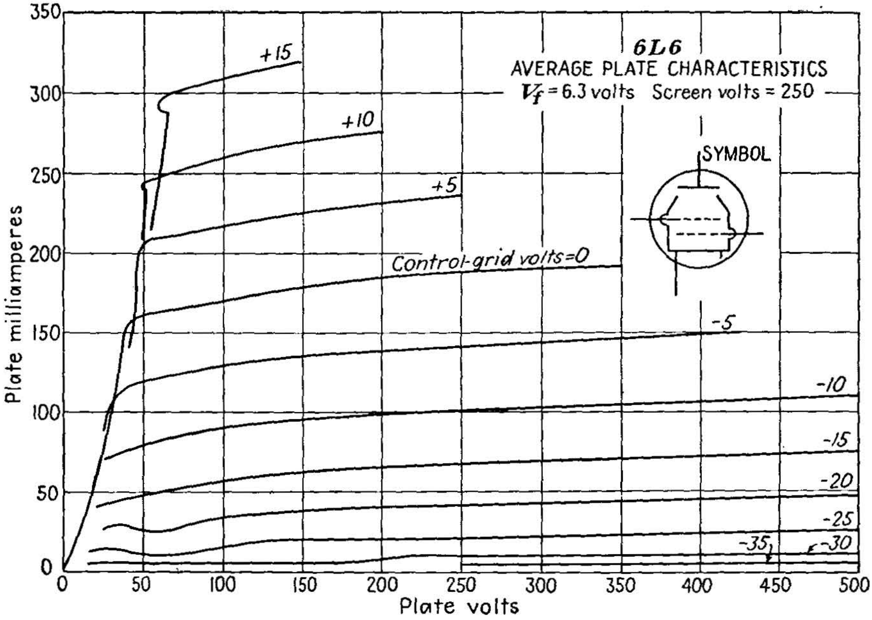


FIG. 2.7.—Plate-current—plate-voltage characteristics of a beam-power tube.

The cathode-ray tube is generally housed in a large glass envelope shaped like an Erlenmeyer flask. In the neck of the glass envelope is located a set of electrodes known as the "electron gun." This gun serves to produce a circular beam of electrons that is fired at the large end of the envelope, which is covered with a fluorescent material. Also housed in the neck of the envelope are deflecting devices that serve to bend the beam in horizontal and vertical directions. The fluorescent screen on the inside of the large end of the envelope gives off light at the point at which the electron beam strikes.

In physical size the ordinary cathode-ray tubes range from 10 to 20 in. in length and have fluorescent screens from 3 to 5 in. in diameter. The tubes operate with a beam-accelerating potential between 800 and 10,000 volts. The electrode arrangement in a typical cathode-ray tube is shown in Fig. 2.8.

Cathode-ray tubes are principally used to observe electrical wave forms. They may also be used to compare frequencies, plot the *B-H* curves of iron, and plot the current-voltage characteristics of vacuum tubes. They are extensively employed as indicators of elapsed-time intervals in ionosphere height-measuring devices and radar sets. They are built in a special form known as the "kinescope" for use as television viewing tubes. They have so many uses as measuring and testing devices that no radio or electronic laboratory worthy of the name is without one.

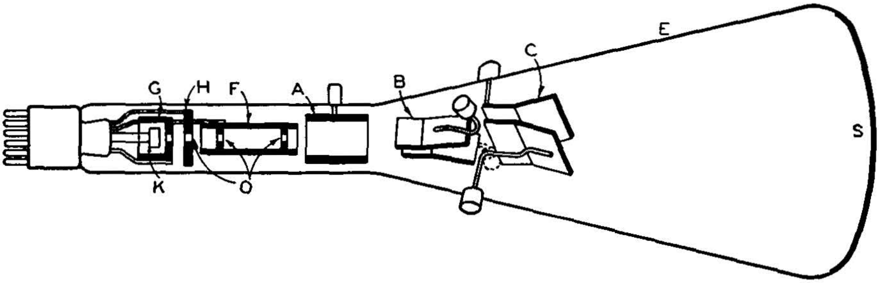


FIG. 2.8.—Typical electrode arrangement in a cathode-ray tube. K, cathode; G, control grid; H, accelerating electrode; F, focusing electrode; A, final accelerating electrode; O, limiting apertures; B, vertical deflecting plates; C, horizontal deflecting plates.

2.7. Klystron. The klystron is a newcomer to the group of vacuum tubes in use today. It is a special ultra-high-frequency tube that is capable of generating, detecting, and amplifying radio waves ranging in frequency from 600 to 30,000 mc (50 to 1 cm).

The principle of operation of the klystron amplifier differs from that of other vacuum tubes. It makes use of a velocity-modulation principle that causes a stream of electrons, which initially has a uniform current density, to form in bunches. It is the periodic bunch impact that excites the output resonator, from which energy is extracted. This use of a beam passing through gaps in closed cavity resonators built into the tube made it possible for the klystron to overcome the transit-time limitations that the conventional negative-grid tubes encounter at high frequencies.

A cutaway drawing of an early type of klystron is shown in Fig. 2.9. The beam of electrons used in the tube is generated in a cathode at one end of the tube. The electrons liberated from this cathode are accelerated toward the main body of the tube, where they pass through a tube and then through a set of grids in a cavity resonator. In passing through this first resonator some of the electrons are speeded up and some slowed down by an alternating axial electric field. This action, called "velocity modulation," causes the electrons to form in bunches by the time they pass through the grids of the second resonator, and it is the bunch impact here that converts the kinetic energy of the electrons into high-frequency electromagnetic energy of the second, or catching, resonator.

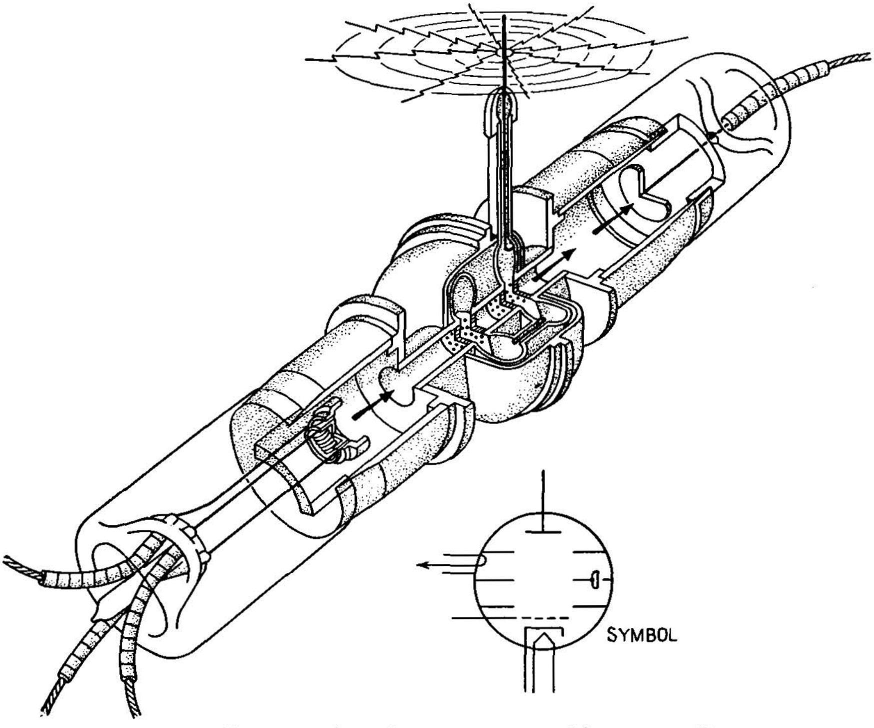


FIG. 2.9.—Cutaway view of a two-resonator klystron oscillator.

A klystron tube may be used as an oscillator by feeding part of the output from the output resonator back to the input resonator. The tube will oscillate when the total phase shift around the circuit composed of the input resonator, the electron beam, the output resonator, and the coupling line back to the input resonator is some integral multiple of 360 deg. Because of this phase requirement it is found that the oscillating action is voltage selective; *i.e.*, the tube will oscillate at certain voltages but not at others since the phase-angle equivalent of the transit time of the electrons along the beam is involved. In Fig. 2.10 is shown a curve of klystron output versus beam voltage. This shows how the tube oscillates at certain select bands of voltage. The maximum power output that can be obtained from a klystron is nearly inversely proportional to the frequency for which the tube is designed, being about 200 watts at 40 cm.

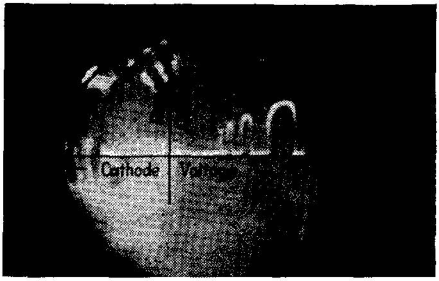


FIG. 2.10.—Power output—beam voltage characteristics of a two-resonator klystron oscillator. This is a picture of an oscilloscope trace, which shows that oscillations are selective with beam voltage.

2.8. Magnetron. The magnetron is a vacuum tube whose current may be influenced by a magnetic field. In certain special forms it is useful as an ultra-high-frequency oscillator. As such it may oscillate at wave lengths from 100 to 1 cm. It is capable of a continuous power output of several hundred watts and instantaneous powers of several thousand kilowatts.

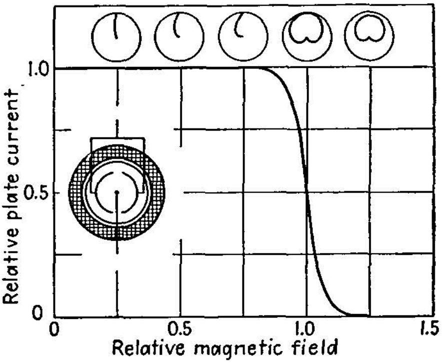


FIG. 2.11.—Cutoff characteristic of a split-anode magnetron. The curve shows that as the axial magnetic field is increased the plate current is at first constant and then suddenly drops rapidly to zero. This results from the electrons becoming progressively more curved in their paths until they finally are unable to reach the plate.

Early forms of the tube were of the split-anode type. The important parts of this type of magnetron are the cathode, frequently in the form of a straight wire filament, and the anode, in the form of a circular plate concentric with the cathode and split into an even number of similar segments. The segments of the plate are operated at the same positive d-c potential relative to the cathode, and a magnetic field is applied parallel to the tube axis. This combination of electric and magnetic fields causes the electrons emitted from the cathode to move in nearly circular paths in the region between cathode and anode.

The radii of the nearly circular electron paths in a magnetron depend upon the strength of the radial electric field and the axial magnetic field. The radii of the paths decrease as the electric field is decreased or as the magnetic field is increased. Thus if a magnetron has its circular plate segments maintained at a constant d-c potential and the strength of the axial magnetic field is increased from zero strength to a large value, the electrons in the tube will at first move radially from the cathode to the plate and then move in paths which are more and more strongly curved until finally the magnetic-field strength is reached at which the electrons miss the plate entirely. This action is shown in Fig. 2.11, in which there is given a plate-current–magnetic-field characteristic and sketches of the associated electron paths. It is seen that the magnetic field is capable of entirely cutting off the current from the plate.

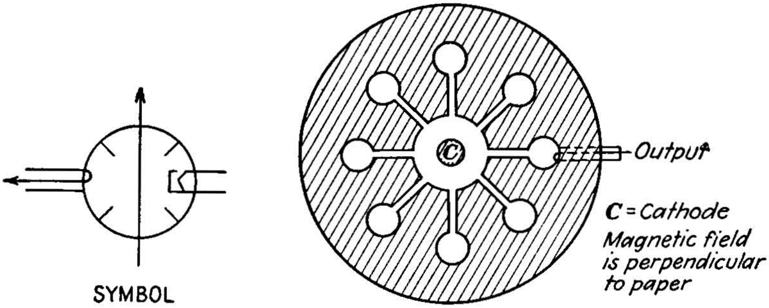


FIG. 2.12.—Electrode structure of a multianode cavity-resonator magnetron. The outer electrode serves as the anode. Each of the hole-and-slot combinations acts like a parallel resonant *L-C* circuit.

For operation as a high-frequency oscillator the plate segments are made part of resonant circuits, and the magnetic field is adjusted to approximately the value that causes the electrons just to graze the plate. If any small disturbance occurs, a complex electronic action results in which the damped oscillation of the resonant circuit affects the electron paths so that some electrons extract energy from the system while others give up part of their kinetic energy to the oscillating system. The tube can be adjusted so that energy is extracted from the majority of the electrons as they graze the plates, and thus powerful oscillations are maintained.

Modern super-high-frequency magnetrons are made in the form of a multianode cavity. The basic structure of such magnetrons is shown in Fig. 2.12. The cathode is in the form of a cylinder of appreciable diameter located in the center of the structure. The anodes are cut out of one piece of metal and have the form of a large circular hole in a block with radial slots leading out to smaller circular holes. Electrically, each slot and terminating hole are equivalent to a tuned resonant circuit, the slot having a predominantly capacitive action and the terminating hole having a predominantly inductive action. One of the resonant conditions possible in this equivalent circuit is one in which alternate segments of the anode exhibit the same electrical polarity and thus give the same action as a split multisegment anode, with the advantage that the fields associated with this resonance are confined. Under proper conditions of voltage and magnetic-field strength parallel to the long axis of the cathode, energy will be transferred from the swarm of gyrating electrons around the anode to the resonant circuit and powerful oscillations will be sustained.

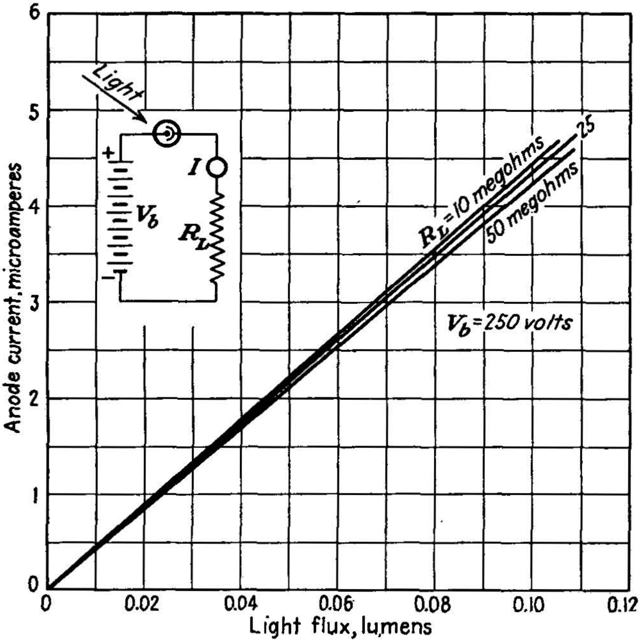


FIG. 2.13.—Current-illumination curves of a typical vacuum phototube. The current is linear with the illumination. Although actual currents are quite small, the voltage developed across the large series resistors used is ample for operating vacuum-tube devices.

2.9. Phototubes. The phototube is a vacuum tube that permits current to pass through it when light falls upon one of its electrodes. The tubes are generally small and contain an electrode in the form of a half cylinder coated with some photosensitive material such as caesium oxide. Various other light-sensitive materials enable the phototube to respond to light of different colors or even to invisible radiations.

Phototubes are extensively used for counting and sorting devices. They may be used to operate doors and drinking fountains, to turn on lights, and to provide safety devices for machine operators. They may be employed anywhere where the interruption or detection of a beam of light is to be correlated with some operation.

Phototubes are able to operate by virtue of an effect known as "photoemission." Certain materials exhibit the property of emitting electrons when exposed to light. The number of electrons emitted is directly proportional to the intensity of the illumination so that a variable light intensity may be translated into a variable electric current or potential. Use is made of this linear property in the recording and reproduction of sound on film. The sensitivity of a phototube in conjunction with a voltage amplifier is so great it may be used to study the light from stars. A typical set of current-illumination curves of a vacuum phototube is given in Fig. 2.13.

CHAPTER 3

ELECTRONS AND IONS

3.1. The Electron. It is the electron that makes vacuum tubes possible and endows them with their remarkable properties. The electron is one of the fundamental particles of matter. It is the lightest particle known. It cannot be subdivided into anything smaller than itself. It is so small that it cannot be observed directly; all observations of its properties must be made in terms of the effects associated with it, such effects as the heat generated upon impact of an electron with a stationary object or the magnetic field surrounding an electron in motion.

For most of the purposes of electronics the electron may be considered to be a small, dense particle carrying a negative charge of electricity. However, it should be borne in mind that this picture of the electron is far from adequate. There are some applications in which the electron displays more of a "wave" aspect than a "particle" aspect. This is the case with the electron microscope, where a high-velocity beam of electrons acts as though it were a light ray of very short wave length.

In the majority of applications the particle aspect of the electron predominates, with the following characteristics:

Mass	9.1066×10^{-31} kg
Negative charge	1.6020×10^{-19} coulomb
Apparent radius	1.9×10^{-13} cm

It is seen that the electron is very dense and is highly charged. It has an apparent density of 0.50×10^{11} g per cm^3, which is millions of times greater than that of our heaviest metals (the density of iron is 7.86 g per cm^3). Further, if the classic concepts of electrostatics be applied to the electron, it may be thought of as being charged to a potential of about 750 kv.

Electrons are a basic constituent of all matter, being the planetary unit of all atoms. No matter can exist without electrons, but electrons may exist by themselves. It is the free electrons that are responsible for most electrical phenomena. They are the units that carry the current in vacuum tubes. They constitute currents in conductors when in motion. Their motion in special conductors such as antennas gives rise to electromagnetic radiations. They constitute cathode rays and beta rays and are emitted from hot bodies.

3.2. The Proton. The proton is the companion piece to the electron. It is the fundamental particle carrying a positive charge of electricity. It, too, is a constituent of all matter, existing as it does in the nucleus of all atoms. The vital statistics of the proton are that its charge is the same in magnitude as that of the electron but with a positive sign, that its mass is 1,845 times that of the electron, and that its apparent diameter is a little less than 10^{-13} cm. The proton is not nearly so much in evidence as is the electron in vacuum tubes. It rarely exists as an isolated particle. Because of its great mass it has a smaller effect than does the electron in determining the characteristics of materials and in constituting a current flow.

3.3. Other Fundamental Particles. Until 1932 the electron and the proton were the only fundamental particles known. Then there were found a number of other fundamental particles whose rarity and short life had hitherto precluded their discovery.

Among these new particles is the *neutron*, which is basically a proton with no charge. There is also a *positron*, which is an electron with a positive charge. There is some evidence of a neutrino, which is a particle of small mass and with no charge. Strangest of all is the *mesotron*, often abbreviated as "meson," which is a particle with about one-tenth the mass of the proton and carrying either a positive or a negative charge. These particles, however, are of no concern to the electronic engineer since they seldom make their appearance in ordinary vacuum tubes.

Another "particle" that has been known for some time is the *photon*. The photon, though classed as a particle, exhibits a wave nature most of the time and is the one particle whose dual nature is most evident. It is a packet of electromagnetic energy whose apparent mass is directly proportional to the frequency of its wave aspect. It carries no charge.

3.4. Atoms and Molecules. Electrically neutral combinations of electrons and protons constitute atoms according to the atomic theory of Rutherford, Bohr, and subsequent workers. The word "atom" is derived from the Greek word meaning "indivisible." Atoms are indivisible in the sense that they are the smallest bits of matter which maintain the properties of the several elements of materials of which they are part. There are 92 types of atoms, corresponding to 92 materials known as "elements." Combinations of the different atoms form molecules, which are the smallest constituent parts of all other materials composing the physical world.

The basic structure of the atom is believed to be a kind of planetary system consisting of a nucleus, which is a group of neutrons and protons, and having a group of planetary electrons equal to the number of protons in the nucleus.

The simplest atom is that of hydrogen. It has 1 nuclear proton and 1 planetary electron. To get an idea of the size of such an atom, if the proton were 1 cm in radius, the electron would normally be spaced a distance of 5 km. The helium atom is the next simplest atom. Its nucleus consists of 2 protons and 2 neutrons. It has 2 planetary electrons. Other atoms are relatively more complicated. The oxygen atom has 8 nuclear protons and 8 nuclear neutrons, whose charge is balanced by that of 8 planetary electrons.

The weight of an atom is determined almost entirely by the sum of the number of protons and neutrons in its nucleus. The physical properties of the atom are determined by the number and arrangement of its planetary electrons. The number of the planetary electrons of an atom of an element is known as the "atomic number" of that element. The order of the elements when listed according to their atomic number is very nearly but not exactly the same as the order according to the atomic weights. If the elements are arranged in a periodic table according to their atomic weights and chemical affinity (valence), as was done by Mendelyeev, it is found that elements with similar characteristics are grouped in columns of equal valence (see Appendix I for a periodic table of the elements).

The planetary electrons of an atom were shown by Bohr to lie in restricted orbits. They were further found to lie in shells about the nucleus, each shell having a maximum capacity for electrons. The maximum capacity of the successive shells from the nucleus out is 2, 8, 18, 32, 18, 18, 2. Thus the atom of neon, whose atomic number is 10 and whose atomic weight is 20.183, has 10 planetary electrons arranged with 2 electrons in the first shell and 8 in the second. These 10 electrons balance the electrical charge of the nucleus, which consists of 10 protons and 12 neutrons.

The number of electrons in the outermost shell of an atom determines its valence and is the principal factor in determining the physical properties of the atom. Atoms with an outer shell filled to its capacity are relatively inactive, while atoms with only 1 electron in their outer shell are most active.

The atomic weights of the elements are taken as relative to that of oxygen, which is chosen to be 16. The fact that the atomic weights are not integers is due in most cases to the fact that there exist atoms of the same element with different numbers of neutrons in the nucleus. The atomic weight of a sample of an element is then determined by the relative number of these different atoms. Atoms with the same number of planetary electrons but with different numbers of nuclear neutrons are known as "isotopes" of the same element. Neon has isotopes with 20, 21, and 22 nuclear particles mixed in such a way that the atomic weight is 20.183. Hydrogen has isotopes with 0, 1, and 2 nuclear neutrons. The first isotope is the common one. The others are the relatively rare "heavy hydrogen" isotopes.

Combinations of the atoms of the elements form molecules. The molecule is the smallest particle of a compound which can exist without losing the characteristics of that compound. Molecules range in size from those of atomic size to a few large enough to be seen with an electron microscope. The molecules of some elements are not just single atoms but groups of two identical atoms.

3.5. Ions. An ion is a molecule or atom with a charge of electricity acquired by the loss or gain of one or more electrons. Electrons in the outer shell of an atom are rather loosely bound to the atom and so may be dislodged by impact of a particle or by exposure to X rays. Ionization of an atom of an element does not change it from one element to another. This is because the nucleus of the atom is unchanged and the form of the nucleus determines the arrangement of the electrons in neutral form.

Ions are important in vacuum tubes because they constitute a current when in motion and thus affect the characteristics of tubes, if they exist in sufficient number. Since even the most completely evacuated tubes contain billions of molecules per cubic centimeter, ions are always created by the impact of electrons and depending upon the type of tube may be a large factor in determining the tube characteristics.

Ions are of most importance in certain special tubes that contain considerable amounts of a definite gaseous element deliberately introduced in great quantities and are an important factor in the tube operation.

CHAPTER 4

ELECTRONIC EMISSION

EVERY vacuum tube depends for its action upon a stream of electrons that acts as a carrier of current. As necessary as the stream of electrons is the electrode that emits them. Whatever the nature of the tube and the arrangement of electrodes, an emitting electrode cannot be dispensed with. Even in cold-cathode tubes, one of the electrodes is treated with a low-work-function material to facilitate the production of some electrons that will initiate the action.

In general, the excellence of performance of a given tube depends upon the efficiency with which free electrons are produced. When the emission fails, the tube is useless. We infer correctly then, that the subject of electron emission is worthy of considerable study.

The types of electronic emission may be listed as follows:

1. Thermionic, or primary, emission.
2. Secondary emission.
3. Photoelectric emission.
4. Field emission.

The common feature of all types of emission is that energy is imparted to the free electrons in a solid in an amount sufficient to enable them to overcome the restraining forces at its surface and thus escape from the solid.

The types of emission differ only in the way in which the escape energy is imparted to the free electrons. *Thermionic emission* occurs when a material is heated to incandescence in a vacuum. In this case the escape energy is imparted by heating the material. *Secondary emission* occurs when a high-velocity electron or ion strikes a material in a vacuum and knocks out one or more electrons. In this case the energy that enables the free electrons to escape comes from the striking particle. *Photoelectric emission* occurs when energy in the form of light falls upon a surface. *Field emission* occurs at cold surfaces under the influence of extremely strong fields.

All types of emission are most effective in vacuum. If the emission did occur in air, the emitted electrons would not get very far through the relatively dense surrounding atmosphere. Most metals would burn up in air at the temperatures to which they must be raised to emit satisfactorily. Only primary and secondary emission will be discussed in this chapter. Photoelectric emission will be discussed in a separate chapter. Field emission is not yet of much practical importance.

4.1. Theory of Thermionic Emission. Every metal has a crystalline structure of its atoms, *i.e.*, the atoms have an orderly arrangement in some sort of lattice pattern. The atoms in this lattice structure have certain of their outer electrons loosely bound. These loosely bound electrons may move from atom to atom in a relatively unrestricted fashion. Such electrons are known as the "free electrons" in the metal in that they are not bound to any one atom. The free electrons in a metal act much like the molecules in a gas. An increase in temperature increases their activity and average velocity. A potential gradient in the metal causes them to move progressively in one direction, giving rise to a conduction-current flow.

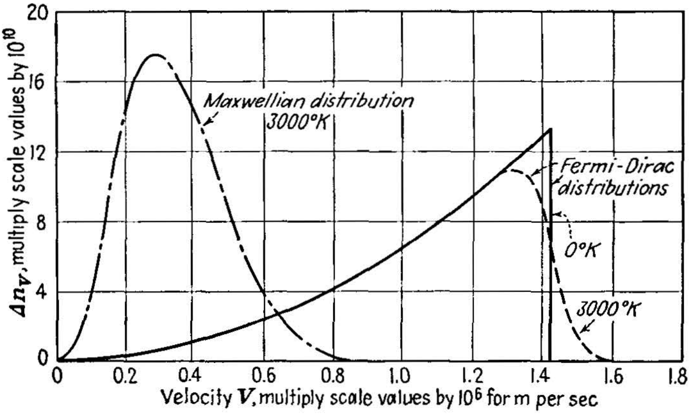


FIG. 4.1.—Maxwellian and Fermi-Dirac distribution of velocities in 1 cu mm of solid tungsten. The abscissa gives the relative number of electrons in a velocity increment of 10^{-3} meters per sec in units of 10^{10} electrons.

Because of the atomic restraints it is not expected that the velocity distribution in a metal is Maxwellian, as is almost exactly the case for gases. The true distribution was found by Fermi and Dirac from quantum-mechanical statistical considerations. For comparison there are shown in Fig. 4.1 the Maxwellian and Fermi-Dirac distribution of velocities. The distinctive feature of the Fermi-Dirac distribution of velocities is that at zero temperature only a small fraction of the electrons have zero velocity. As temperature increases, the velocity and corresponding energy distribution change so that more electrons have higher velocities.

The high-velocity electrons that escape from the metal constitute the emitted current.

The Maxwellian distribution of velocities referred to above and shown in Fig. 4.1 is one given by the equation

$$y = \frac{4x^2}{\sqrt{\pi}} \epsilon^{-x^2} \tag{4.1}$$

This is the general form of the probability y that a particle will have a velocity x times the most probable velocity. It applies perfectly for most gases but does not give the true picture for electrons in metals. For large velocities, however, the Maxwellian and Fermi-Dirac distributions differ only by a constant. Thus the electrons emitted from an incandescent surface do have a Maxwellian distribution, but the energies of the electrons are (at 3000°K) about 1,000 times those predicted from the simple Maxwellian theory. Upon converting Eq. (4.1) to a form involving energy instead of velocity and taking the derivative properly, the fraction $\frac{n}{n_0}$ of the emitted electrons that can move against a retarding field of V volts is given by

$$\frac{n}{n_0} = \epsilon^{-\frac{Ve}{kT}} \tag{4.2}$$

where $-e$ is charge of the electron, 1.602×10^{-19} coulomb
k is Boltzmann's constant, 1.380×10^{-23} watt-sec per °K
T is absolute temperature, 273+ °C

A nomographic chart of Eq. (4.2) is given in Fig. 4.2. From this it is seen that about 50 per cent of the electrons emitted from a cathode at 1500°K, typical oxide operating temperature, have velocities greater than 0.09 volt.

Work Function. The surface restraints that prevent the majority of the free electrons in a metal from leaving it are the electrostatic forces produced by the charges in the atoms. These come not only from the residual positive charges but also from a rearrangement of the negative charges. A free electron must have a certain minimum kinetic energy before it can tear itself free from these forces. The work per unit charge required to free an electron from the influence of the charges in the metal and thus to escape from it is known as the *work function* of the metal. The work function is usually expressed in volts.

The electrostatic forces within a metal are rather complex and not completely understood. Indications are that the forces are small within the metal, reach a maximum several atomic diameters outside the metal, and then decrease according to an inverse-square law at greater dis-

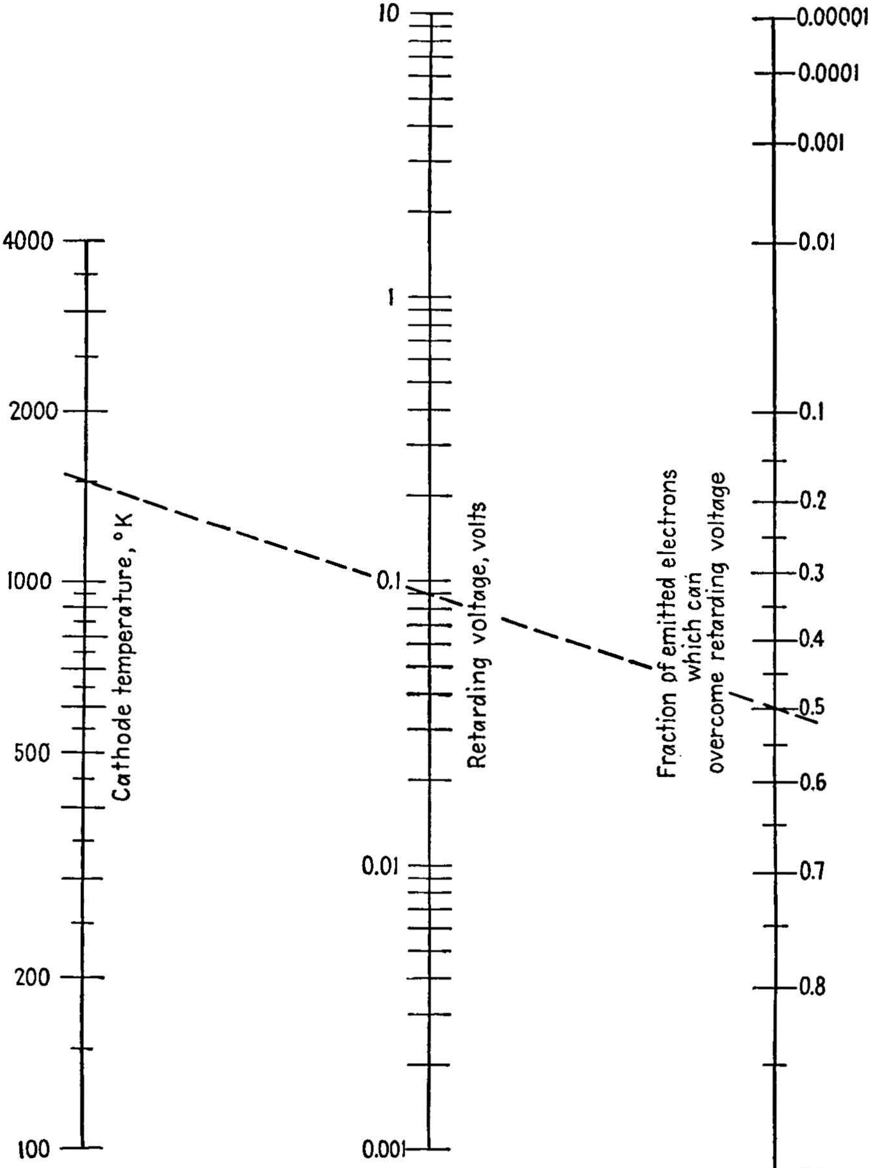


FIG. 4.2.—The velocity distribution of electrons resulting from thermionic emission as given by Eq. (4.2). The nomographic chart gives the fraction of the emitted electrons associated with a given cathode temperature that can overcome a retarding potential of a given number of volts.

tances, where an image action manifests itself. It would be expected that the work function would decrease as the distance between the atoms in the crystalline structure increased. This turns out to be the case, and experimentally, a curve of the work function of the alkali metals of the first column of the periodic table plotted against their atomic spacing is a smooth one, nearly inversely proportional to the square root of the atomic spacing, as may be seen in Fig. 4.3. Conclusions for other metals can hardly be drawn, for there are so few having the same valence and crystalline structure. Since the atomic spacing is a periodic function of the atomic number, the work function is also a periodic function of the atomic number.

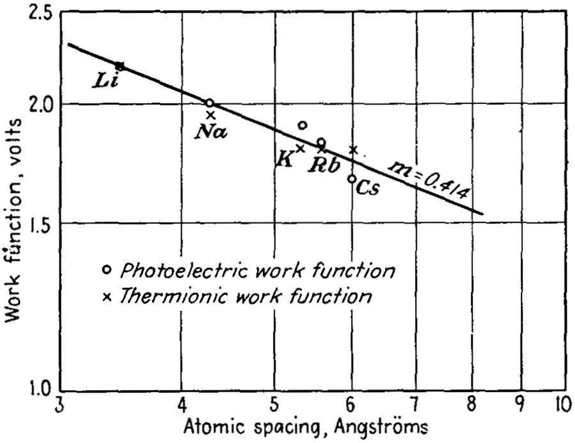


FIG. 4.3.—Work function of the alkali metals as a function of atomic spacing. The curve shows that for a given crystal structure, the further the atoms are apart the lower is the work function.

No completely successful theoretical determination of the work function has apparently as yet been made. The general nature of the restraining forces is probably very much like that shown in Fig. 4.4. Within the metal the force has an average value of zero. Near the surface there are the attractive forces of atoms that have lost an electron by emission and forces due to rearrangement of residual charges. The forces are undoubtedly greatest near the surface, where the force-producing charges are closest and yet not symmetrically disposed with respect to the surface. Well outside the surface the force is probably one that varies with the inverse square of the distance from the metal, for in this region the charges in the metal arrange themselves so as to give the effect of an image charge of the electron escaping from the metal. The force cannot be inverse-square law all the way out from the surface, for then an infinite energy would be needed for escape. It may be concluded that the work function depends in some complex way upon the atomic spacing, crystal structure, and valence of the metal.

The work function of materials is most accurately determined experimentally from observations of the photoelectric emission of the material, but it may also be deduced from the thermionic-emission characteristics. A list of the work functions of the metal emitters most often used is given in Table I.[1,2]

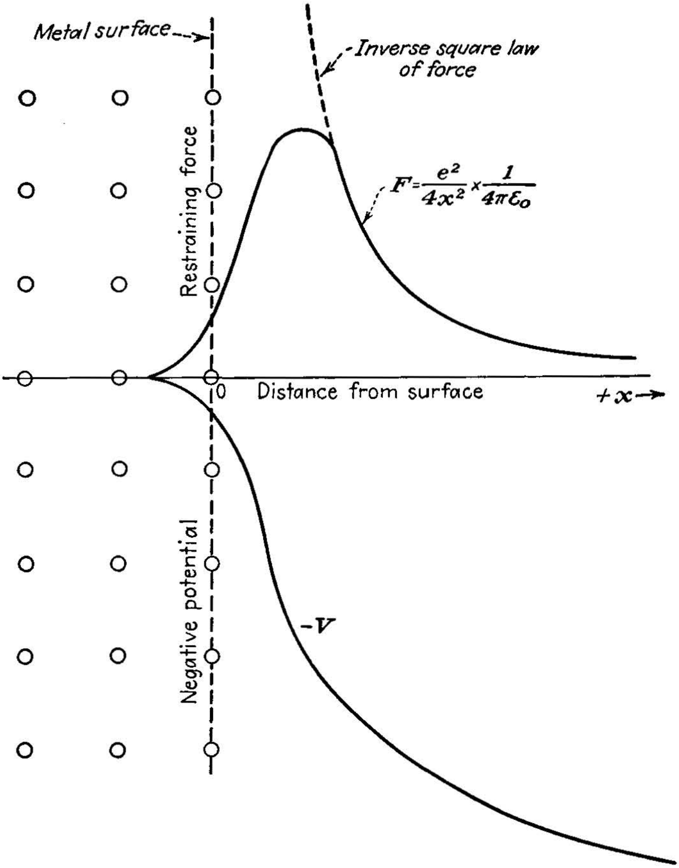


FIG. 4.4.—Restraining force on an electron near a metal surface. At large distances from the metal the force is that due to an image charge located in the metal.

---

[1] HUGHES, A. L., and L. A. DUBRIDGE, "Photoelectric Emission Phenomena," McGraw-Hill, New York, 1932.

[2] BECKER, J. A., Thermionic Emission and Adsorption, *Rev. Modern Phys.*, vol 7, pp. 95–128, April, 1935.

TABLE I
EMISSION CONSTANTS OF THE METALS

	Probable ϕ_T,* volts	Average ϕ_P,† volts	Melting temp., °C	Lattice const, angstrom units
Ag	4.7	4.6	960.5	4.08
Al	3.0	3.0	659.7	4.04
Au	4.8	4.78	1063	4.07
Ba	2.52		850	5.015
Bi	4.1	4.2	271.3	4.75
C	4.7	4.77	>3500	2.455
Ca	3.2	3.0	810	5.56
Cd	4.1	4.0	320.9	2.97
Cs	1.8	1.67	28.5	6.05
Cu	4.1	4.3	1083	3.61
Fe	4.7	4.74	1535	2.90
Hg	4.5	4.53	−38.87	
K	1.8	1.90	62.3	5.33
Li	2.2	2.21	186.0	3.46
Mg	2.4	2.43	651.0	3.20
Mo	4.3	4.15	2620	3.14
Na	1.9	2.0	97.5	4.24
Ni	5.0	5.01	1455	2.66
Pb	4.0	3.9	327.4	4.94
Pt	6.0	6.3	1773.5	3.91
Rb	1.8	1.82	38.5	5.62
Sr	2.1	2.06	800	6.05
Ta	4.1	4.13	3269	3.28
Th	3.4	3.50	1845	5.07
W	4.52	4.61	3370	3.16
Zn	8.3	3.44	419.47	2.66
Zr	4.1	3.73	1900	3.22

* Work function as determined by thermionic measurements.

† Work function as determined by photoelectric measurements.

The Emission Equation. In view of the foregoing discussion it would be expected that the emission from a metal would depend upon its temperature and upon the work function. Richardson[1] and Dushman[2] have

---

[1] RICHARDSON, O. W., The Distribution of the Molecules of a Gas in a Field of Force, *Phil. Mag.*, vol. 28 (No. 5), pp. 633–647, 1914.

[2] DUSHMAN, S., Electron Emission from Metals as a Function of Temperature, *Phys. Rev.*, vol. 21 (No. 6), pp. 623–636, 1923. See also the summarizing source article, S. Dushman, Thermionic Emission, *Rev. Modern Phys.*, vol. 2, pp. 381–476, October, 1930, which gives a comprehensive survey of the subject as developed to that date. See also the book, A. L. Riemann, "Thermionic Emission," Wiley, New York, 1934.

shown this to be the case and have shown specifically that the thermionic emission from a metal is given by

$$J = AT^2\epsilon^{-\frac{b_0}{T}} \tag{4.3}$$

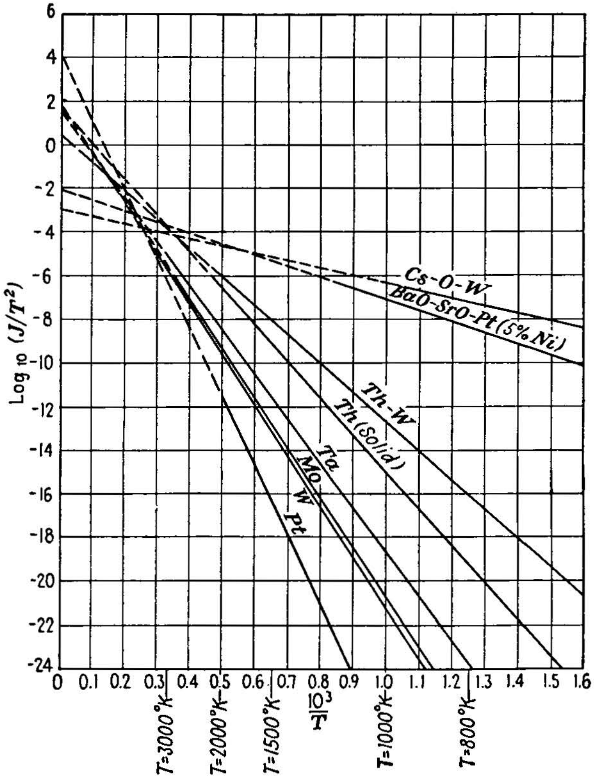


FIG. 4.5.—Characteristics of the common emitters shown as a curve of log J/T^2 against $1/T$. This type of plot demonstrates the validity of the Richardson-Dushman equation (4.3). The y-axis intercepts give the emission constant A. The slope of the lines is proportional to the work function of the emitter.

where J is current density, amperes per cm²

A is 120.4 amperes per cm² per deg², a universal theoretical constant

T is absolute temperature, °K (273 + °C)

b_0 is temperature equivalent of the work function, 11,600ϕ_0,°K

ϕ_0 is work function of the metal, volts

Equation (4.3) may be derived from either thermodynamic or quantum-mechanical considerations. The resulting equation is the same in either case.

From the form of the emission equation (4.3) it is seen that if the logarithm of $\frac{J}{T^2}$ be plotted against the reciprocal of T there will result a straight line whose slope is $-b_0$ and whose y-axis intercept is $\ln_e A$. The correctness of the emission equation has been verified by so plotting experimentally determined results. It is found in all cases that the results produce a straight line. A group of such curves for common emitters is given in Fig. 4.5. In this figure those lines with the lowest slope correspond to metals with the lowest work function. Theoretically, the intercept should be 2.08, corresponding to the $\log_{10} 120.4$. Actually, it is about 1.78, corresponding to a value of A of 60 instead of 120.4 for most of the pure metals. Values of A are found higher as well as lower than the theoretical values so that the theory is not discredited by this discrepancy. There is some evidence that the work function is not entirely independent of temperature as has been assumed in the derivation of the emission equation. The differences in the value of the work function as determined by thermionic and photoelectric methods may possibly be due to temperature. A decrease in the work function of 6 parts per 100,000 per degree would cause the observed discrepancy in the constant A.

The exponential term in the emission equation accounts for most of the variation of emission with temperature. The variation with the T^2 term is so small that the correctness of the exponent 2 can hardly be verified experimentally. In the case of tungsten at 2500°K a 1 per cent change in temperature changes the T^2 term by 2 per cent but changes the exponential term by 20 per cent. This causes the emission-temperature function to be one of the most rapidly varying functions found in nature. Doubling the temperature may increase the emission by a factor of 10^7. Halving the work function will have nearly the same effect as doubling.

The quantities of the curves of Fig. 4.5 are not in very convenient form for ordinary use, and therefore a better method of representing the emission characteristics of materials is sought. It is possible to plot emission current against temperature directly as in Fig. 4.6, but the variation of current with temperature is so rapid that such a curve is not very satisfactory. It would also be possible to plot emission against heating power by making use of the fact that at the high temperatures required for emission most of the power is lost through radiation according to the Stefan-Boltzmann law.

$$P = Ke_R T^4 \tag{4.4}$$

where P is radiated power, watts per cm^2

K is 5.73×10^{-12} watt per cm^2 per deg^4, a universal constant known as the "Stefan-Boltzmann constant"

e_R is radiation efficiency as fractional radiation of a black body or perfect radiator

Such a plot gives curves that are nearly but not quite straight lines because of the two temperature factors in the thermionic-emission equation. It is possible, however, to warp the lines of the emission scale to take account of the nonuniform temperature variation and get a straight-line plot as shown in Fig. 4.7. The coordinate paper used in Fig. 4.7 is known as "power-emission paper." On it curves of emission against heating power are straight lines to the extent that the radiation efficiency of the emitter remains constant with temperature. Contours of emission efficiency in milliamperes per watt are also readily drawn. Since heat-

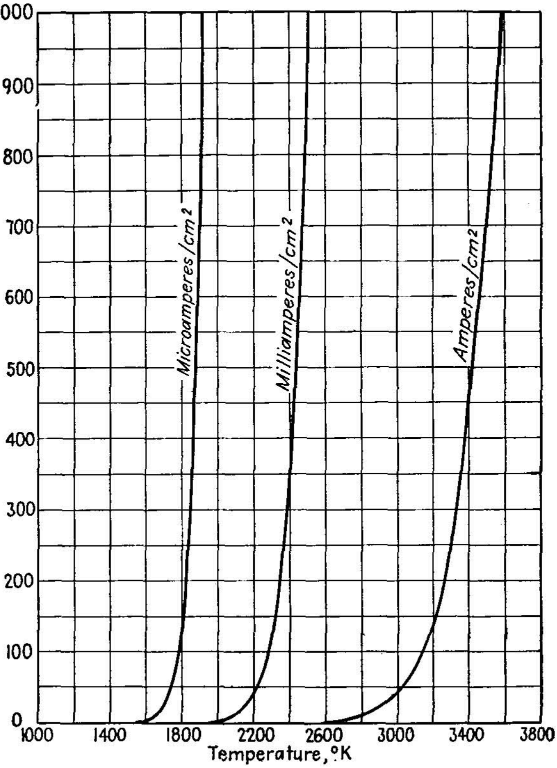


FIG. 4.6.—The emission-current density of a tungsten emitter as a function of temperature.

radiation efficiency varies rather slowly with temperature, as shown in Fig. 4.8, the emission-power curves can be extrapolated as straight lines with considerable assurance. Radiation efficiency is defined as the per

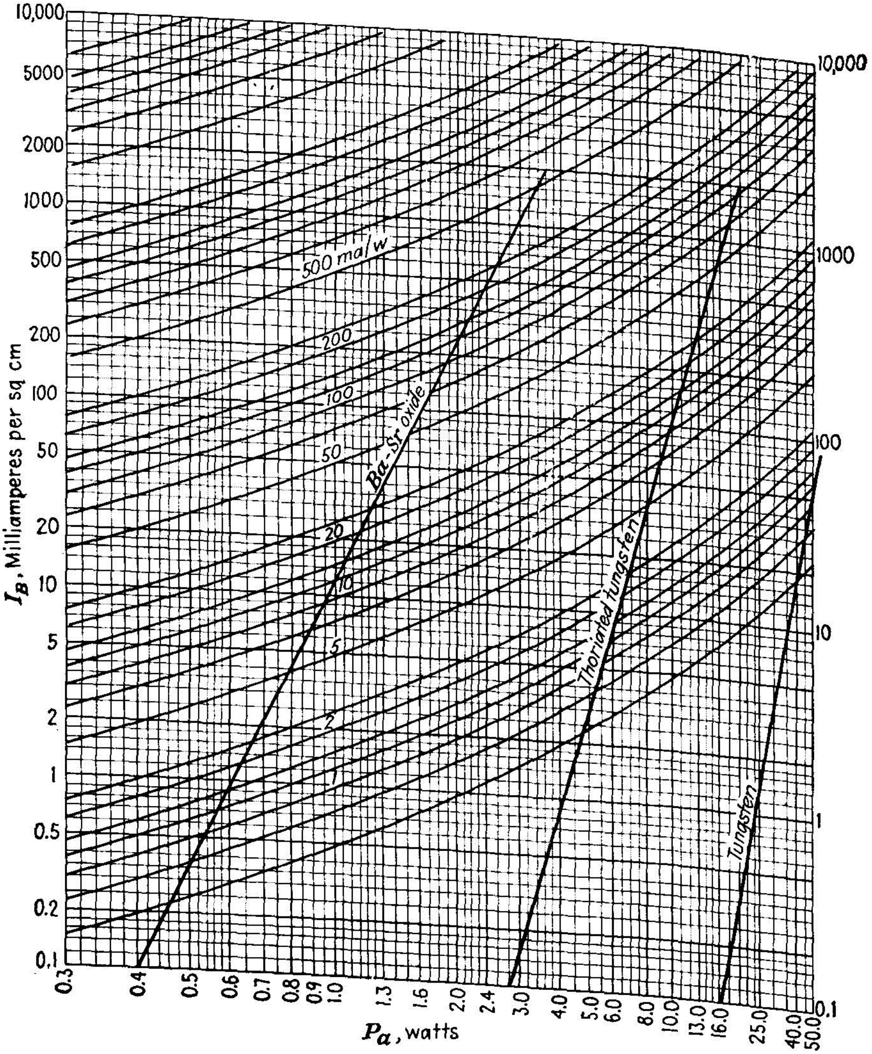


FIG. 4.7.—Emission-current density as a function of heating power for typical examples of various emitter types. The curves show the relative emission efficiencies of the different emitters.

cent of black-body, or perfect, radiation. Black-body radiation as given by Eq. (4.4) is shown in Fig. 4.9. Power-emission paper is manufactured and sold by the Keuffel and Esser Company.

Types of Emitter.—Because of the dependence of emission upon temperature and the work function it is not necessarily true that the metal with the lowest work function is the best emitter. This is shown by the case of caesium, which has the lowest work function of all the metals, 1.8 volts. It cannot be made to give much thermionic emission because it can be raised only to 300°K, slightly over average room temperature, before it melts. On the other hand, tungsten, which has a rather high work function, 4.52 volts, has the highest melting temperature of all the metals, 3655°K, and as a result gives the highest emission of all the pure metals just below its melting temperature. Caesium, however, is preferred for photoelectric emission and secondary emission where temperature is not a factor.

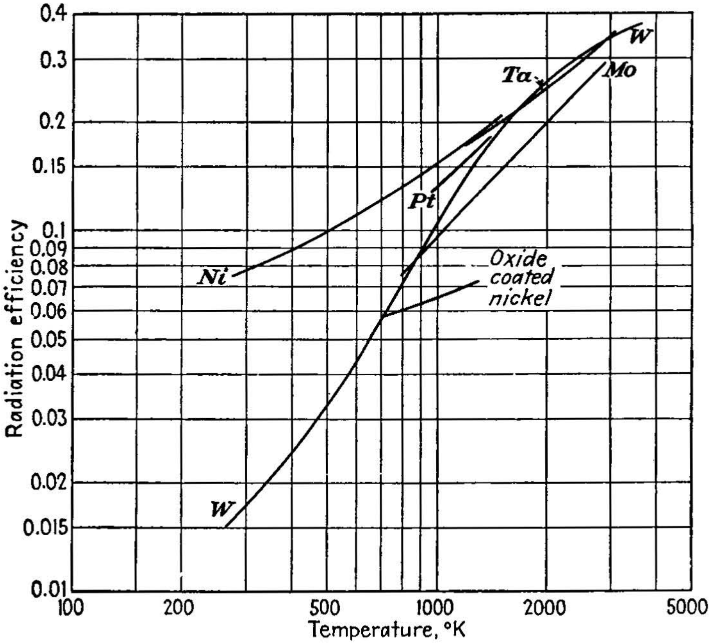


FIG. 4.8.—The radiation efficiency of various metals used in vacuum-tube construction as a function of temperature. Efficiency is given as a fraction of black-body radiation, which is shown in Fig. 4.9.

It has been found that it is possible to raise some metals to temperatures higher than their melting temperatures in the pure state by using them in various chemical and physical combinations. Thus a monatomic layer of thorium on tungsten can be operated at or above the melting temperature of thorium itself. Also, it has been found that small bits of the pure metal can be made to diffuse out of an oxide in the case of the rare-earth metals so that advantage can be taken of the low work function of these metals, which would otherwise melt at low temperatures.

From the above remarks it is seen that three classes of emitters exist. They are

1. Pure metals.
2. Atomic-film emitters.
3. Oxide emitters.

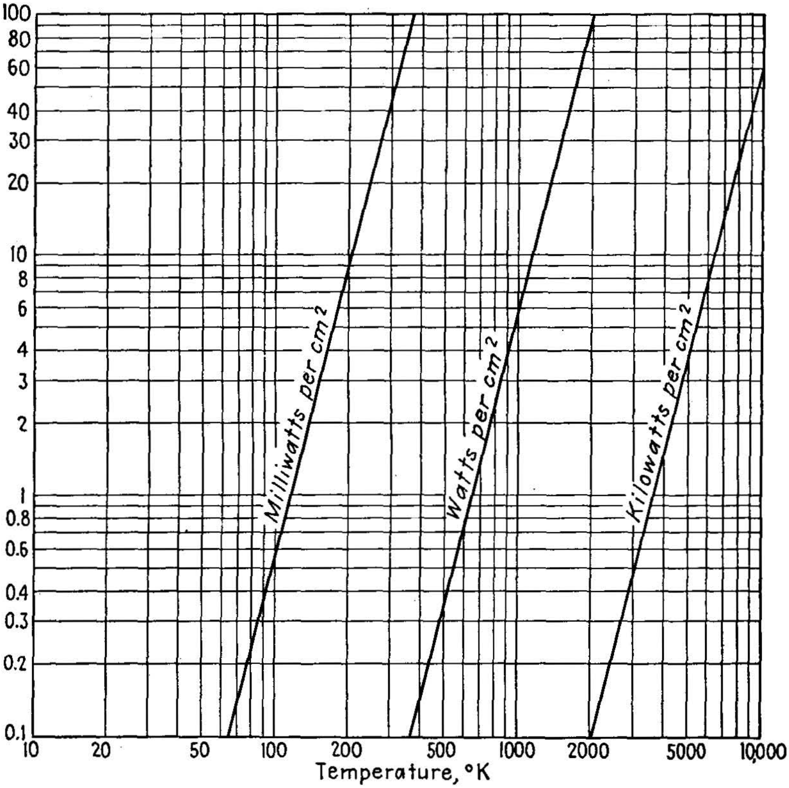


FIG. 4.9.—The thermal radiation of a black body or ideal radiator as a function of temperature as given by the Stefan-Boltzmann law of Eq. (4.4).

These different types of thermionic emitters will now be discussed separately.

4.2. Emission of Pure Metals. *Tungsten.* The pure metals follow the Richardson-Dushman emission equation as closely as can be determined experimentally. In general, the metals with suitable physical characteristics for emission have a relatively high work function and so even at best are not very good emitters. Of all the metals *tungsten* is the most extensively used because it can be raised to a higher temperature without melting than any other metal. Although tungsten has a desirable high melting temperature, its other physical characteristics are less desirable. It is a hard metal to work because of its crystalline structure. It was not until 1908, when Coolidge discovered that tungsten becomes ductile when extensively worked, that it became practical to use the metal at all. Tungsten cannot be drawn into wire form as can most metals but must be hammered into shape, a process known as "swaging."

The emission characteristics of tungsten have been extensively studied, and more is known of its thermionic behavior than is known of any other metal.[1,2]

The principal characteristics of tungsten as given by Jones and Langmuir are recorded in Table II. The data in this table are for a wire of unit length and unit diameter. The characteristics for any other diameter and length are readily determined by the dimensional equations given. The principal features of tungsten emission are given in the curves of Fig. 4.10. An example of the use of Table II is given in Prob. 4.3.

Because of its relatively low emission, tungsten is not used as an emitter unless the application is such that other emitters cannot be used. Tungsten is used almost exclusively for filaments of tubes with plate potentials higher than 4,000 volts. This is because other emitters cannot stand the positive-ion bombardment at energies corresponding to this high potential. The positive ions referred to have their origin in residual gases in the tube. All other emitters have their emission impaired when subjected to bombardment by these high-energy particles. Except for the brittleness caused by crystallization at high temperatures, tungsten filaments are more rugged than any other. Like all emitters, tungsten is subject to reduction of emission from contamination by various gases. Tungsten cleans up more readily by heating or bombardment than any other material.

Tantalum. The only other pure-metal emitter of any importance is tantalum. Tantalum cannot be heated to as high a temperature as tungsten because its melting temperature is 3300°K. However, the work function of tantalum is relatively low, being 4.1 volts against 4.53 volts for tungsten, so that its emission is at least ten times that from

---

[1] JONES, H. A., and I. LANGMUIR, The Characteristics of Tungsten Filaments as Functions of Temperature, *Gen. Elec. Rev.*, vol. 30, Part I, pp. 310–319, June; Part II, pp. 354–361, July; Part III, pp. 408–412, August, 1927.

[2] FORSYTHE, W. E., and A. G. WORTHING, The Properties of Tungsten and the Characteristics of Tungsten Lamps, *Astrophys. Jour.*, vol. 61, pp. 146–185.

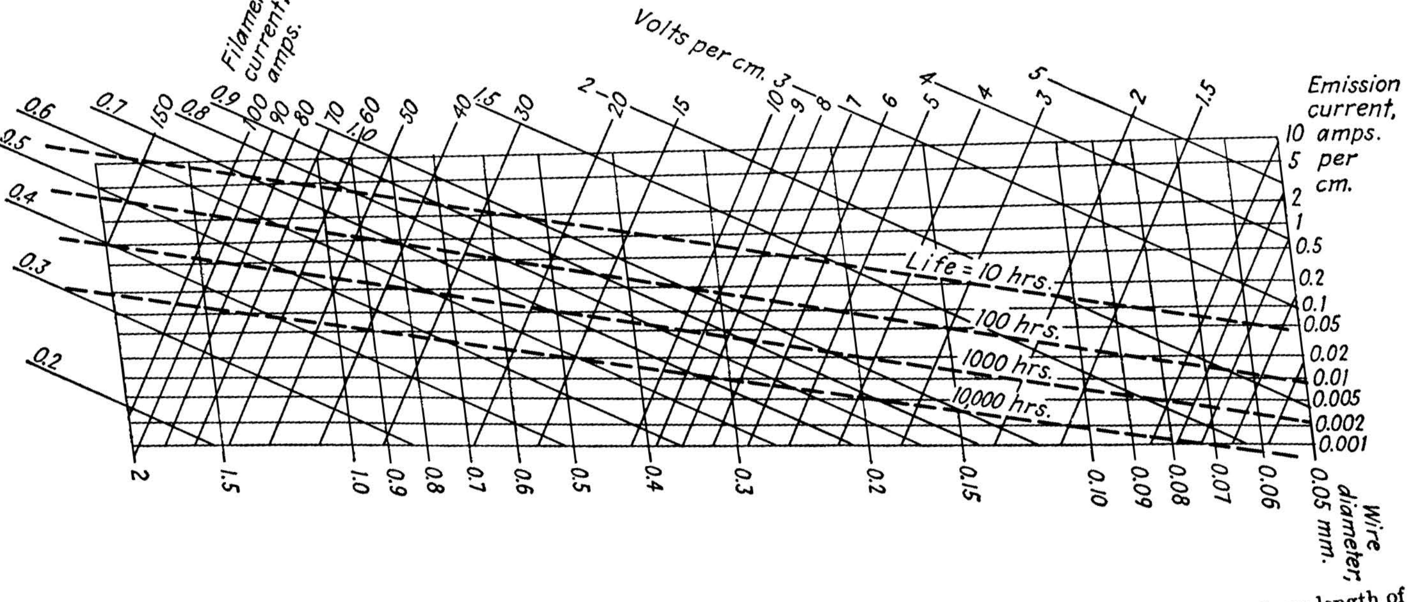
TABLE II
SPECIFIC CHARACTERISTICS OF IDEAL TUNGSTEN FILAMENTS*

(For a wire 1 cm in length and 1 cm in diameter)

T, °K	W', $\frac{W}{ld}$ watts per cm^2	$R' \times 10^6$, $\frac{Rd^2}{l} \times 10^6$, ohm-cm	I'_f $\frac{I_f}{d^{3/2}}$, amps per $cm^{3/2}$	$V' \times 10^3$, $\frac{V\sqrt{d}}{l} \times 10^3$, volts per $cm^{1/2}$	I'_e, $\frac{I_e}{ld}$, amp per cm^2	M', $\frac{M}{ld}$, g per cm^2 per sec, evaporation	$\frac{R'_T}{R'_{293°}}$, $\frac{R_T}{R_{293°}}$
273	....	6.37	....	....	....	....	0.911
293	0.0	6.99	0.0	0.0	....	....	1.00
300	0.000100	7.20	3.727	0.02683	....	....	1.03
400	0.00624	10.26	24.67	0.2530	....	....	1.467
500	0.0305	13.45	47.62	0.6404	....	....	1.924
600	0.0954	16.85	75.25	1.268	....	....	2.41
700	0.240	20.49	108.2	2.218	....	....	4.93
800	0.530	24.19	148.0	3.581	....	....	3.46
900	1.041	27.94	193.1	5.393	....	....	4.00
1,000	1.891	31.74	244.1	7.749	3.36×10^{-15}	1.16×10^{-33}	4.54
1,100	3.223	35.58	301.0	10.71	4.77×10^{-13}	6.81×10^{-30}	5.08
1,200	5.210	39.46	363.4	14.34	3.06×10^{-11}	1.01×10^{-26}	5.65
1,300	8.060	43.40	430.9	18.70	1.01×10^{-9}	4.22×10^{-24}	6.22
1,400	12.01	47.37	503.5	23.85	2.08×10^{-8}	7.88×10^{-22}	6.78
1,500	17.33	51.40	580.6	29.85	2.87×10^{-7}	7.42×10^{-20}	7.36
1,600	24.32	55.46	662.2	36.73	2.91×10^{-6}	3.92×10^{-18}	7.93
1,700	33.28	59.58	747.3	44.52	2.22×10^{-5}	1.31×10^{-16}	8.52
1,800	44.54	63.74	836.0	53.28	1.40×10^{-4}	2.97×10^{-15}	9.12
1,900	58.45	67.94	927.4	63.02	7.15×10^{-4}	4.62×10^{-14}	9.72
2,000	75.37	72.19	1,022	73.75	3.15×10^{-3}	5.51×10^{-13}	10.33
2,100	95.69	76.49	1,119	85.57	1.23×10^{-2}	4.95×10^{-12}	10.93
2,200	119.8	80.83	1,217	98.40	4.17×10^{-2}	3.92×10^{-11}	11.57
2,300	148.2	85.22	1,319	112.4	1.28×10^{-1}	2.45×10^{-10}	12.19
2,400	181.2	89.65	1,422	127.5	0.364	1.37×10^{-9}	12.83
2,500	219.3	94.13	1,526	143.6	0.935	6.36×10^{-9}	13.47
2,600	263.0	98.66	1,632	161.1	2.25	2.76×10^{-8}	14.12
2,700	312.7	103.22	1,741	179.7	5.12	9.95×10^{-7}	14.76
2,800	368.9	107.85	1,849	199.5	11.11	3.51×10^{-7}	15.43
2,900	432.4	112.51	1,961	220.6	22.95	1.08×10^{-6}	16.10
3,000	503.5	117.21	2,072	243.0	44.40	3.04×10^{-6}	16.77
3,100	583.0	121.95	2,187	266.7	83.0	8.35×10^{-6}	17.46
3,200	671.5	126.76	2,301	291.7	150.2	2.09×10^{-5}	18.15
3,300	769.7	131.60	2,418	318.3	265.2	5.02×10^{-5}	18.83
3,400	878.3	136.49	2,537	346.2	446.0	1.12×10^{-4}	19.53
3,500	998.0	141.42	2,657	375.7	732.0	2.38×10^{-4}	20.24
3,600	1,130	146.40	2,777	406.7	1,173	4.86×10^{-4}	20.95
3,655	1,202	149.15	2,838	423.4	1,505	7.15×10^{-4}	21.34

---

* The values given are taken from H. A. Jones and I. Langmuir. The Characteristics of Tungsten Filaments, *Gen. Elec. Rev.*, vol. 30, pp. 312–313, 1927, Table I. The notation of Jones and Langmuir is retained in this table.



Characteristics of ideal tungsten filaments as given by the data of Langmuir and Jones. Curves are for a 1-cm length of wire. for a 10 per cent reduction in mass through evaporation.

tungsten at any temperature less than 2500°K. Tantalum has the advantage over tungsten that it can be worked in sheet form to produce specially shaped cathodes, and the like. A disadvantage is that it is easily contaminated by residual gases, which form oxides that greatly reduce the emission.

4.3. Atomic-film Emitters. It is possible to get emission higher than that from pure metals from an atomic film of one metal on another. Of the various combinations that are possible, the most extensively used is that of *thorium on tungsten.* It was discovered by Langmuir and Rogers that the small amount of thorium put into tungsten to reduce the crystallization gave rise to very high emission under certain conditions. What apparently happens is that a certain amount of thorium in the metal diffuses to the surface, where it emits much as thorium would, with the advantage that the thorium can be heated above its own melting temperature and that the work function is reduced by the redistribution of charges in the tungsten and surface layer of thorium.

Thorium was originally added to tungsten to reduce crystallization. As now added to increase the emission, the amount is about 1½ per cent, and this amount is quite critical. If more than this amount is added, the tungsten wire is too hard to work. If less is added, there may not be enough to produce high emission. The thorium is added in the form of thoria (thorium oxide, ThO_2).

A rather intricate schedule of operations is required to produce and activate a film of thorium on tungsten. The process includes the following steps:

1. *Reduction of Thoria to Metallic Thorium.* This is achieved by heating the filament to 2800°K for 1 or 2 min. During this time, most of the thorium oxide is reduced to thorium, and such thorium as reaches the surface evaporates. If the emission is measured at this point, it will be found to be very nearly the emission of pure tungsten.

2. *Diffusion of Metallic Thorium to the Surface.* This takes place as the filament is held at a temperature of 2100°K for a period of 15 to 30 min. During this time the emission increases by a factor of about 1,000. The explanation of this behavior is that metallic thorium diffuses to the surface, where it builds up a monatomic layer of thorium. Studies with the electron microscope[1] show that the thorium arrives at the surface both through pores in the tungsten and at the grain boundaries, from which places it spreads over the surface. At this reduced temperature the evaporation is not very large. In the range of temperatures between 2100 and 2300°K the thorium diffuses to the surface faster than

---

[1] Brüche, E., and H. Mahl, Ueber das Emissions bild von thorierten Volfram und thoriertem Molybdän, *Zeit. für Tech. Phys.*, vol. 16, pp. 623–627, December, 1935.

it evaporates, so that this is a suitable range for activation. In this range of temperature the percentage of the surface covered varies from 20 to 85 per cent, decreasing as temperature increases as shown in Fig. 4.11. The final layer of thorium that forms is believed to be monatomic.

3. *Operation.* After the above treatment the filament temperature is reduced to 1900°K, where it may be operated for long periods of time in a very stable fashion. At this temperature, both the diffusion and evaporation are low, but there is a sufficient preponderance of diffusion to maintain a good emitting surface. Any temperature below 1900°K is suitable for operation. At this temperature, the tungsten surface is about 85 per cent covered by thorium, and the life of the coating is several thousand hours. If the temperature is reduced, the effective work function is decreased, the life is increased, the percentage surface coverage is increased, but the emission is decreased.

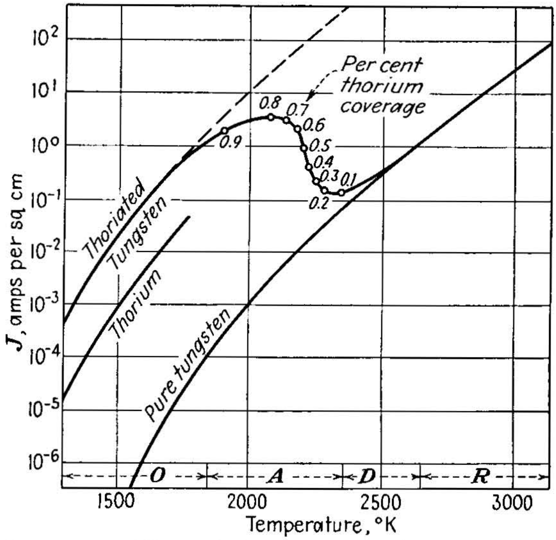


FIG. 4.11.—The emission of thoriated tungsten as a function of temperature. *O* indicates operating range of temperatures; *A*, activation range; *D*, diffusion range, and *R*, reduction range.

It is interesting to note that the thoriated tungsten filaments are usually operated at 1900°K, which is nearly the melting temperature of thorium, something that could not be done with the pure metal because of its softness at this high temperature. Also, the work function of thoriated tungsten filaments is 2.6 volts for a 100 per cent covered surface, and this work function is lower than the work function either of tungsten, 4.51 volts, or of thorium, 3.4 volts. The work function of thoriated tungsten is a linear function of the surface coverage given by $\phi = 4.51 - 1.9\theta$ volts, where θ is the fraction of the tungsten surface covered by thorium. The reason why the work function is reduced by having the metals in combination is that most of the electrons in the thorium layer are drawn toward the tungsten base. This produces a dipole layer on the surface, with its positive end outward. This means that in most of the surface region the electrostatic forces are outward, opposing the image forces and thus reducing the work function.

Thoriated tungsten surfaces are always carbonized to increase the life. It has been found that if some of the tungsten is converted to tungsten carbide (W_2C) the evaporation of thorium from its surface is greatly reduced.[1] The rate of evaporation of thorium from a tungsten carbide surface at 2200°K is only about one-sixth of that from an uncarbonized surface at this temperature. Carbonization may be achieved by heating the filament to a temperature of 1600°K in a vapor of some hydrocarbon such as naphthalene or acetylene. It may also be achieved by heating the filament to red heat in an atmosphere of hydrogen while in contact with a carbon surface. As the filament is converted to tungsten carbide, its electrical conductance decreases until when totally converted it is about 6 per cent of the original value. The electrical resistance is therefore an excellent index of the degree of conversion. In practice, it is found that the conversion cannot be carried beyond the point where the conductance is reduced to 80 per cent of its original value, for the tungsten carbide is so brittle that the filament would be dangerously weakened by further action.

The fact that the layer of thorium on tungsten is monatomic is evidenced by at least two aspects of the behavior of the composite emitter surface. (1) If the filament is deactivated by heating to a higher temperature after having been activated, the manner in which the emission reduces with time is independent of the length of time the film has been activated. This indicates that the activation beyond a certain point does not add any more emitting material to the surface, which can be the case only if the layer is monatomic and surplus atoms are lost by evaporation. (2) There is no discontinuity in the emission characteristics during the activation process.

Monatomic films other than thorium on tungsten may be used. It is found that they are not as stable as a thorium layer because of more rapid diffusion and evaporation, and hence they are not much used. Curves showing the emission characteristics of various combinations are

---

[1] KOLLER, L. R., "The Physics of Electron Tubes," 1st ed., McGraw-Hill, New York, 1934.

shown in Fig. 4.12. In this figure the notation *O-W* means that the emitting metal is on an oxidized tungsten surface.

4.4. Oxide Emitters. In 1904, Wehnelt discovered that copious electron emission could be obtained from alkaline-earth oxide coatings. The entire development of small vacuum tubes is based upon this discovery, for oxide coatings are used almost exclusively as a source of emission in them. The alkaline-earth metals that are readily available are barium, strontium, and calcium, and it is their oxides that have been found to give such high emission. Modern oxide coatings are usually a half-and-half mixture of the oxides of barium and strontium. Such a coating will give high emission at low temperatures with a high thermal efficiency; thus at 1000°K an emission of 100 ma per cm^2 at an efficiency of 20 ma per watt input is readily obtained. This is about the same emission as is given by a tungsten filament at 2300°K, but the emission efficiency here is only 1 ma per watt. The oxide coatings may be applied either to an indirectly heated cathode surface or directly to a filament. They are particularly well adapted to making specially shaped unipotential cathodes.

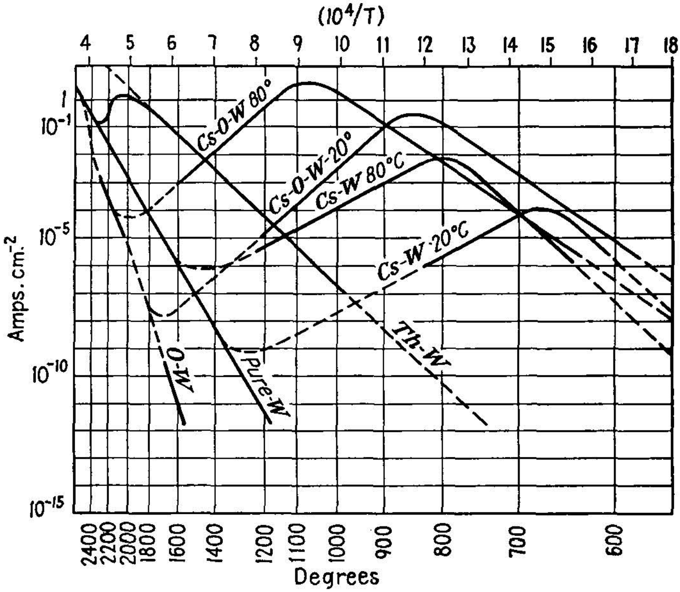


FIG. 4.12.—The emission of monatomic films on tungsten. (*After Dushman.*)

Theory of Oxide Emission. Oxide emission has been the subject of extensive study for the last 30 years though it has not been until recently that an explanation of the action has been available in fairly complete form.[1–5]

The accumulated evidence indicates that the emission takes place from particles of free metal on the surface of the oxide coating. The free metal is made available by the following mechanisms:

1. Chemical reaction of the oxide with the core metal.
2. Electrolytic reduction by the gradient of potential through the oxide coating.
3. Reduction of the oxide by positive-ion bombardment.

It was discovered early that the core metal played a part in the electron emission. It was even believed that electrons were liberated at the core. This was disproved by showing that there was no emission from the core metal when the coating was removed by mechanical shock. Further, the emission was shown to be independent of the size and shape of the core. Also, the photoelectric work function of the oxide surface was the same as the thermionic work function. However, the most conclusive evidence that the emission is from the surface is that the same emission characteristics are obtained from an oxide coating if metal is vaporized onto the surface as is obtained by the normal process of activation. Different core metals do, however, exhibit different effects upon the emission. In the order of their reaction titanium, tantulum, nickel, and molybdenum will react with the alkaline-earth oxides to produce core-metal oxide and free alkaline earth. The action is evidenced by the fact that oxides can be activated by heating alone. The titanium reaction is probably responsible for the excellent performance obtained with cores of "Konel" metal, which is an alloy of nickel, iron, cobalt, and titanium. The metal most used for core metals is nickel, which is preferred because of its excellent physical properties and low cost.

Free alkaline-earth metal is also made available by the electrolytic action associated with the passage of current through the coating. The earth oxides dissociate under the usual condition of polarity. The metal ion goes to the core, and the oxygen ion is liberated. This action can be detected by the liberation of oxygen.

Dissociation of the oxides is also caused by positive-ion bombardment.

---

[1] BLEWETT, J. P., Properties of Oxide Coated Cathodes, *Jour. Appl. Phys.*, vol. 10, Part I, October, 1939, pp. 668–679; Part II, pp. 831–848, December, 1939.

[2] DUSHMAN, S., Thermionic Emission, *Rev. Modern Phys.*, vol. 2, pp. 381–476, October, 1930.

[3] RIEMANN, *op. cit.*

[4] BECKER, *op. cit.*

[5] BLEWETT, J. P., Oxide Coated Cathode Literature, 1940–1945, *Jour. Appl. Phys.*, vol. 17, pp. 643–647, August, 1946.

Even in the best vacuums there are enough ions present to give an appreciable action. The fact that activation is greatly facilitated by application of a positive potential to the tube in processing is considered sufficient evidence of the existence of this action.

As with the atomic-film emitters the resultant work function is lower than that of the pure metals alone, and these are already very low. Reported values of the work function of oxides have shown a tremendous variation until recently, when improved vacuum techniques and a better understanding of the mechanism have given rise to some fairly consistent values. The work functions of the oxides are now believed to lie within 25 per cent of the following values:

BaO	1.1 volts
SrO	1.4 volts
CaO	1.9 volts
BaO + SrO	1.0 volts

Emission from the combination of barium and strontium oxides is seen to be better than from either one alone. The reduction in work function over that of the pure metals is again probably due to an electrical double layer formed by a monatomic coating of the pure metal on the oxide. Values of the emission constant A also show a great range of variation as reported by various observers. It has been found that both the emission constant and the work function change with the degree of activation of the oxide coatings. Both decrease with activation, and experimentally it is found that the work function is a linear function of the logarithm of the emission constant. Properly speaking, it is not correct to ascribe an emission constant to oxide coatings, for the emission law in this case is slightly different from the Richardson-Dushman law.[1] An equivalent emission constant is of the order of 0.01 amperes per cm^2 per deg.

Electron-microscope studies of oxide emission show that there is no relation between surface irregularities and emission.[2] Variations in work function are observed with orientation of crystal faces. The emission surface does not change much with degree of activation though the emission may change greatly. Emission is improved by reducing oxide particle size, as may be done by using colloidal particles.

In operation, an oxide cathode has to establish an equilibrium between rate of production of free emitting metal and evaporation of the same. This means the establishment of an equilibrium between electrolysis, diffusion, and evaporation. This latter will be disturbed if the temperature of the oxide or the amount of current is changed. Under normal

---

[1] BLEWETT, Properties of Oxide Coated Cathodes, Part I, *op. cit.*

[2] HEINZE, W., and S. WAGENER, Vorgänge bei Aktivierung von Oxydkathoden, *Zeit. fur Tech. Phys.*, vol. 17 (No. 12), pp. 645–653, 1936.

conditions the equilibrium adjusts itself to the current drawn so quickly that no change is evident. If, however, the cathode temperature is low or if the emission is partly contaminated or partly exhausted, there will be evident an adjustment of emission over a period of seconds or even minutes as the current drawn is suddenly changed.

The adjustment is of the following nature: If the voltage on a tube is increased, the current immediately increases and then drops slowly, coming to rest at a value between the previous and initial value. If the voltage is decreased, the current will immediately decrease and then slowly rise to a value between the previous and initial value.

Activation of Oxide Emitters. Since the alkaline-earth oxides are not stable in air, the coating must be applied to the cathode or filament in the form of a carbonate or hydroxide. The carbonates are most extensively used, being held to the surface with an organic binder. Coatings of a thickness of 0.010 to 0.020 in. work well. When a coating has been applied and the tube evacuated, the coating is activated by first heating it to a temperature of about 1500°K for a few minutes. This reduces the carbonates to oxides, and during this time copious CO_2 is evolved. Considerable thermal reduction also occurs, with attendant evaporation of liberated metal. The oxide coating is then operated at a temperature of about 1000°K with a potential of about 100 volts applied to an adjacent electrode through a protective resistor. Electrolysis and positive-ion bombardment then occur, and the emission will build up slowly to a final value, when the filament will be ready for use.

Various other methods of applying coatings may be used. Heating in air is recommended to eliminate the organic binder. For a water paste the coating should be baked in an inactive gas to get good adherence. Hydroxides, which are very good for coating tungsten, may be dipped and then baked in air to get a so-called "combined coating."

Specific Emission Characteristics. The lines of Fig. 4.5 show the behavior of oxide coatings in comparison with other emitters. The low work function is evident from the small negative slope of the curve. Emission as a function of power is shown in Fig. 4.7 in contrast with other emitters. The higher emission efficiencies are evident. The emission obtainable from oxide coatings has increased with the years. This may continue, though an increase over present values by more than a factor of 10 is not probable. Some comparative emission efficiencies are

	Ma per Cm² per Watt
Pure tungsten filament	2–10
Thoriated tungsten filaments	5–100
Oxide-coated indirectly heated cathodes	10–200
Oxide-coated filaments	200–1,000

Under normal conditions the life of an oxide coating should be several thousand hours. Cessation of emission is due to exhaustion of free metal in the oxide. In mixed coatings there is a preferential evaporation of the barium, which finally leaves the relatively less efficient strontium to give a greatly reduced emission.

Oxide coatings are more easily damaged or poisoned than any other type of coating. They are particularly susceptible to poisoning by oxygen. Emission may be reduced by several powers of 10 by the presence of oxygen at a pressure of 10^{-4} mm of mercury, while a pressure of 10^{-3} mm will inhibit emission completely. Oxide coatings are seldom used on tubes where they will be subjected to bombardment of more than 1,000 volts. Bombardment by particles of higher energy will disintegrate an oxide coating completely.

Transient Emission. The monatomic layer of barium of the oxide coating has tremendous instantaneous-emission potentialities. Such a layer may yield instantaneous emission as great as 100 amperes per cm^2. When short-time high voltages are applied, such large emission may be realized.[1,2] The high voltage exhausts the available emission in a time of the order of milliseconds. When this happens, the supply of free barium must be resupplied through processes of reduction and diffusion. Since this takes an appreciable time, a current-voltage plot of a diode operated under these conditions at 60 cycles exhibits pronounced exhaustion effects, giving rise to a loop in the retrace characteristic. When a very sharp pulse of voltage is applied to an emitting surface, the emitted current consists of a capacitive displacement component as well as the conduction component. As a result, the current pulse will generally have an initial peak with a subsequent rapid decay.

4.5. Schottky Effect. A departure from the Richardson-Dushman emission equation occurs when the emitting surface is subjected to a strong positive potential gradient. Effectively the field reduces the work function. As a result, the current from an emitter increases with the potential applied even though the temperature is kept constant and the emission is not affected by the space charge of the electrons.

The action may be understood by referring to Fig. 4.13, in which the effect of a constant gradient of potential upon the normal potential barrier at the surface of the emitter is shown. The combination of the constant gradient and the normal potential barrier is seen to give a new potential barrier, which has a maximum at a certain distance d_c from

---

[1] SCHADE, O. H., Analysis of Rectifier Operation. *Proc. I.R.E.*, vol. 31 (No. 7), pp. 341–361, 1943.

[2] COOMBES, E. A., Pulsed Properties of Oxide Cathodes, *Jour. Appl. Phys.*, vol. 17, pp. 647–654, August, 1946.

the surface. This distance is known as the "critical escape distance" because once an electron gets beyond this distance the electrostatic forces are outward rather than restraining and thus an electron keeps on moving. Upon equating the image field with the gradient, the maximum of the restraining potential is found to occur at a distance

$$d_c = \frac{1}{2}\sqrt{\frac{e}{4\pi\varepsilon_0 E}} \tag{4.5}$$

where e is the charge on the electron, E is the potential gradient, and ε_0 is the dielectric constant of free space of value 8.85×10^{-12} for rationalized mks units. The crest of the potential barrier has been reduced by the work the electron would have to do to overcome the image force from the surface from d_c to infinity. This amount of work is $\frac{e}{2}\sqrt{\frac{eE}{4\pi\varepsilon_0}}$ volts. The work function is further reduced the same amount owing to the fact that the potential at the distance d_c is reduced by the amount d_cE. The total reduction in the effective work function is thus $e\sqrt{\frac{eE}{4\pi\varepsilon_0}}$ volts. When this correction is made for the work function in the Richardson-Dushman equation, it is found that the ratio of the emitted current in the presence of the strong electric field to the normal emission current is given by

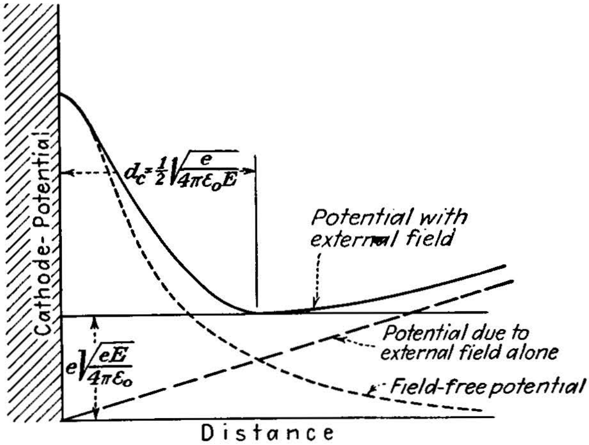


FIG. 4 13.—Diagram of the potential barrier associated with the Schottky effect.

$$\frac{J_E}{J} = \epsilon^{4.403E^{1/2}/T} \tag{4.6}$$

where J_E is the emission-current density in the presence of the strong electric field, J is the normal emission-current density, ϵ is the Napierian base 2.718, E is now the negative gradient of potential in volts per centimeter, and T is the temperature in degrees Kelvin. This equation may be verified experimentally by plotting the logarithm of J_E against the square root of E. The experimental results are found to give a good straight line for all but low values of gradient at which the current drops more rapidly than this simple theory predicts. The slope of the line is, of course, $\frac{4.403 \log_{10} \epsilon}{T}$.

4.6. Contact Difference of Potential. Another factor that occasionally enters the emission picture is "contact difference of potential." This term is given to the effect observed when two dissimilar metals are put in good electrical contact. It is found that a small potential difference will exist between the free surfaces of the two different metals. This difference of potential turns out to be the difference between the work functions of the metals and arises from the fact that electrons can move more readily from the metal of low work function to the metal of high work function than vice versa. The differential action results in an equilibrium that leaves the metal of low work function positively charged relatively to the metal of high work function by just the difference of the work functions. In ordinary vacuum tubes contact differences of potential are usually less than $\frac{1}{10}$ volt and so do not cause serious trouble except in special cases. Such differences of potential as may arise from contact of dissimilar metals will be most serious in such places as the cathode–control-grid circuit.

4.7. Secondary Emission. Another form of emission that plays an important role in vacuum tubes is secondary emission. This occurs when a surface is struck by electrons or ions of appreciable velocity. Secondary emission caused by the bombardment of electrons is the more important case and occurs whenever the striking electrons have energies corresponding to a few volts or more. When this happens, the striking electrons may knock one or more electrons out of the material, giving rise to a reverse component of current. The electrons knocked out of a material, known as "secondary" electrons, may number more than the "striking," or "primary," electrons. There is no violation of the conservation of energy law when this happens, for the velocity of the secondary electrons is for the most part very low. Secondary emission is commonly encountered in multiple-electrode tubes, where it has the effect of altering somewhat the normal primary-electron current characteristics. It occurs in cathode-ray tubes where the beam electrons hit the fluorescent screen, and is necessary there to complete the circuit for the current flow. It is deliberately used in a number of types of electron-multiplier tubes, where it makes possible a high amplification of current by a purely electronic action.

Secondary-emission characteristics of materials are measured by means of the apparatus shown schematically in Fig. 4.14. In the arrangement shown a beam of electrons is directed at a target inside of a sphere at a higher potential, which attracts the secondary electrons liberated at the target. The ratio of secondary- to primary-electron current can be read for any primary-electron potential.[1] For a long time there were great discrepancies in the reported secondary-emission characteristics of the various metals. It was evident that small traces of impurities or surface contaminations made a great difference in the secondary-emission characteristics. Techniques have now been refined to the point where the values reported by various investigators are fairly consistent. The average secondary-emission characteristics of the materials commonly used in vacuum tubes when only the ordinary precautions against contamination are taken are shown in Fig. 4.15.[2]

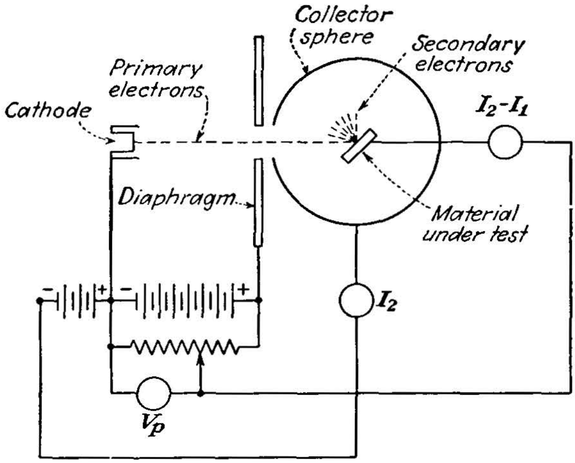


FIG. 4.14.—Apparatus for the measurement of secondary-emission characteristics.

Variation of Secondary Emission with Primary-electron Potential. In Fig. 4.16 are shown the secondary characteristics of the common metals presented in curve form, giving the ratio of secondary- to primary-elec-

---

[1] See KOLLATH, R., Sekundarelektronemission fester Körper, *Physik. Zeit.*, vol. 38, pp. 202–224, Mar. 15, 1937, for an excellent discussion of methods of measurement and results obtained up to that date.

[2] HARRIES, J. H. OWEN, Secondary Electron Radiation, *Electronics*, vol. 17, pp. 100–108, 180, September, 1944.

tron current as a function of the primary-electron potential as reported by Bruining and DeBoer.[1] These results probably are more reliable than any previously reported, for the investigators used a special apparatus in which the metal to be tested was evaporated onto the target in a vacuum just before the measurement was made. The results presented show lower ratios of secondary- to primary-electron current than those previously reported. This is probably due to the fact that with previous handling the metals became partly oxidized and oxidized surfaces are

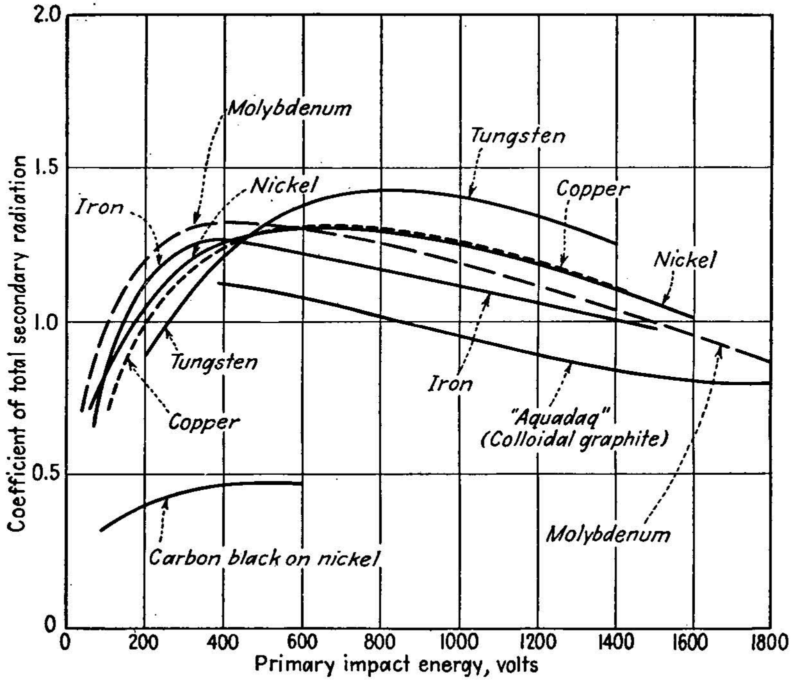


FIG. 4.15.—Secondary-emission characteristics of the metals under ordinary conditions. The curve shows the ratio of the number of secondary to primary electrons for various primary-electron impact velocities expressed in volts. (*After Harries.*)

---

[1] BRUINING, H., and J. H. DEBOER, Secondary Emission, Part I, Secondary Emission of Metals, *Physica*, vol. 5, pp. 17–30, January, 1938; Part II, Absorption of Secondary Electrons, *Physica*, vol. 5, pp. 901–912, December, 1938; Part III, Secondary Electron Emission Caused by Bombardment with Slow Primary Electrons, *Physica*, vol. 5, pp. 913–917, December, 1938; Part IV, Compounds with a High Capacity for Secondary Electron Emission, *Physica*, vol. 6, pp. 823–833, August, 1939; Part V, Mechanism of Secondary Electron Emission, *Physica*, vol. 6, pp. 834–839, August, 1939; Part VI, Influence of Externally Adsorbed Ions and Atoms, on the Secondary Electron Emission of Metals, *Physica*, vol. 6, pp. 941–950, October, 1939.

known to have higher secondary emission than those which are not. The curves of Fig. 4.15 show that all the metals have a low secondary emission at low primary-electron potentials. Most of the metals have a maximum secondary emission between 200 and 400 volts of primary potential, which then decreases slowly, becoming constant at a value between 50 and 95 per cent of the maximum value. Most of the uncontaminated metals have a maximum ratio of secondary- to primary-electron currents less than 1 though it should be remembered that metals as encountered in tubes are seldom uncontaminated and will have maximum ratios of the order of 1 to 5.

Although the complete theory of secondary-electron emission is as yet not worked out, a great deal is known of the mechanism.[1,2] When primary electrons strike a surface at right angles, they may knock electrons out of the atoms near the surface and those with velocity components directed toward the surface may be able to overcome the surface-potential restraints and escape from the metal. Each primary electron may shake up several atoms, thus giving rise to several electrons emitted per primary electron. It should be noted that the source of secondary electrons lies almost entirely in the electrons of the surface atoms and not in the free electrons of the metal. If a normally directed primary electron strikes a free electron, it cannot give it a component of velocity directed toward the surface. Electrons knocked out of atoms, however, may have such

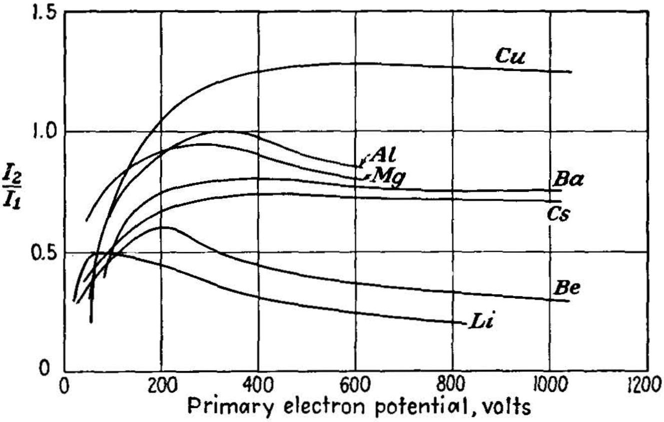


FIG. 4.16.—Secondary-emission characteristics of metals with inappreciable surface contamination. (*After Bruining and DeBoer.*)

---

[1] *Ibid.*, Part V.

[2] WOOLDRIDGE, D. E., Theory of Secondary Emission, *Phys. Rev.*, vol. 56, pp. 562–578, Sept. 15, 1939.

a component. As the potential of the primary electron is increased, it will at first knock out more and more secondary electrons. However, as the potential is further increased, the surface atoms are exposed to the primary-electron forces for a shorter time, *i.e.*, the so-called "collision diameter" decreases, and the primary electron will first knock electrons out of atoms when it has slowed down upon penetration into the metal. Thus at the maximum of emission it is believed that the majority of the secondary electrons are liberated a depth of several atoms into the metal.[1] Beyond this potential the primaries penetrate still farther into the metal, and the probability that the electrons knocked out of the atoms at this depth will reach the surface decreases, with the result that the secondary emission decreases.

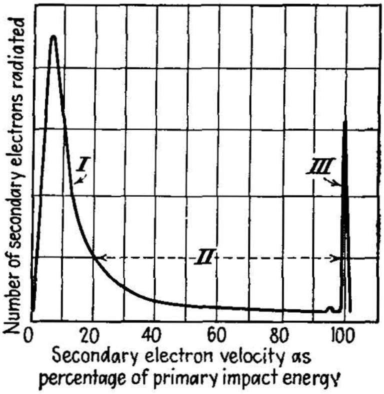


FIG. 4.17.—The relative velocity distribution of secondary electrons. About 90 per cent of the secondary electrons will have velocities in range I, 7 per cent in range II, and 3 per cent in range III.

Velocity Distribution of Secondary Electrons. In Fig. 4.17 is shown a typical curve of the distribution of velocities in the secondary electrons emitted from a metal. Most of the electrons, about 90 per cent, have velocities below 20 volts. The electrons naturally fall into three groups as indicated in the figure. These are as follows:

Group I—0 *to* 20 *volts.* This group comprises about 90 per cent of all the secondaries for primary potentials of 50 volts or more. There is a pronounced maximum in this group at about 10 volts. These are the electrons which are shaken out of the atoms as a result of the passage of the primary electrons and do not have much energy.

Group II—20 *volts to* 98 *per cent of the primary-electron potential.* These comprise about 7 per cent of the total secondary current. They represent high-energy electrons knocked out of atoms and elastic reflections of the primary electrons at a considerable depth in the metal.

Group III—98 *to* 100 *per cent of primary-electron potential.* This group comprises only about 3 per cent of the secondary current

---

[1] BRUINING, H., Depth at Which Secondary Electrons Are Liberated, *Physica*, vol. 3, pp. 1046–1052, September, 1936.

and has a maximum at about 99 per cent of the primary-electron potential. This group arises from elastic reflections of primary electrons from atoms near the surface of the metal, not really secondary electrons at all.

Another representation of the secondary-electron velocity distribution is obtained if potential between sphere and target of the apparatus of Fig. 4.14 is made negative instead of positive and the current of the sphere is measured against the retarding potential. The resultant curve is shown in Fig. 4.18. This curve is an average for measurements on various metals with primary-electron potentials in the range of 275 to 1,000 volts. Curves like those in Fig. 4.17 are obtained by taking the negative derivative of curves such as those in Fig. 4.18.

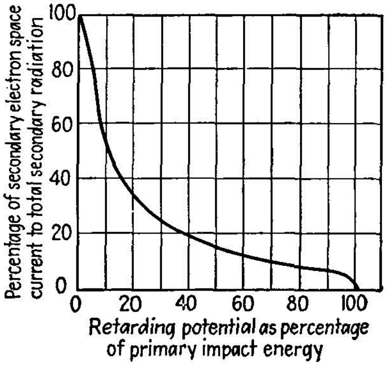


FIG. 4.18.—Collector current as a function of retarding potential of the secondary-emission measuring apparatus of Fig. 4.14.

Variation of Secondary Emission with Angle. When primary electrons strike a surface at right angles, it is found that secondary electrons are emitted at all angles. The spray of secondary electrons seems to follow very nearly a cosine law of distribution under all conditions.

When the primary electrons strike a metal surface at an angle, it is found that the distribution of the angle on the secondaries is still nearly a cosine-law variation. More important than this is the fact that the secondary- to primary-emission ratio increases as the primary electrons strike more nearly parallel to the surface. Some typical curves showing the variation of the secondary- to primary-emission ratio are given in Fig. 4.19. The increase in secondary emission with angle is largely due to the fact that at angles other than normal the primary electron may knock free electrons out of the metal as well as electrons out of atoms. The variation of emission is given quite closely by[1]

$$R_\theta = R_0 \epsilon^{p(1-\cos\theta)} \tag{4.7}$$

where θ is angle between normal and direction of primary electrons

R_θ is ratio of secondary to primary electrons at angle θ

R_0 is ratio of secondary to primary electrons at angle zero

ϵ is Napierian base 2.718

---

[1] BRUINING and DEBOER, *op. cit.*, Part II.

p is a coefficient that increases with primary potential and is proportional to the primary-electron penetration

Secondary Emission of Composite Layers. Certain combination surfaces have been found to have pronouncedly higher secondary emission than the pure metals. Such surfaces are the alkalihalides on a base of the alkali metal and alkali oxides on various metal bases. All these combinations show the same general secondary-emission characteristics as do the pure metals except that the current ratios instead of being in the vicinity of unity may be as high as 8 to 11. The velocity distribu-

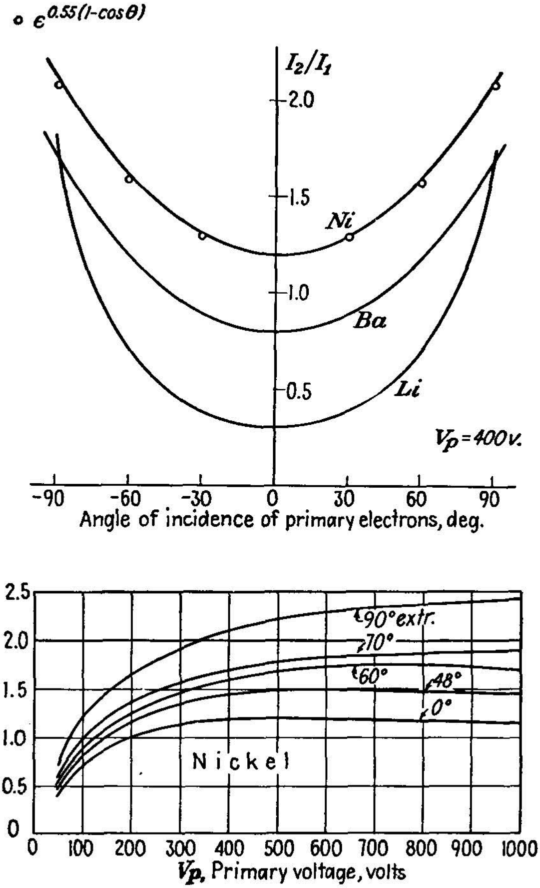


FIG. 4.19.—Variation of secondary-emission ratio with angle of primary impact. Note that the secondary-emission ratio increases as the angle of incidence becomes more nearly grazing. (*After Bruining and DeBoer.*)

tion for composite surfaces is much narrower than for the pure metals, *i.e.*, a given percentage of the total electrons are included in a lower range of velocities, 85 per cent in the first 3 volts. Below are given data on some of the alkali halides.[1]

TABLE III
MAXIMUM SECONDARY-EMISSION RATIOS OF ALKALI HALIDES

Compound	Maximum Ratio
LiF	5.6
NaF	5.7
CaF_2	3.15
NaCl	6.8
KCl	7.5
RbCl	5.8
CsCl	6.5
NaBr	6.25
NaI	5.5
KI	5.6

Of the alkali oxides, by far the best emitter is caesium oxide, partly reduced, on a base of silver. Some typical curves for alkali oxides are shown in Fig. 4.20. This same combination gives very high photoemission. Photoemissive surfaces are prepared in the same way.

In connection with composite surfaces it should be noted that a combination with a low work function does not necessarily have a high secondary- to primary-electron ratio, and vice versa. Thus tungsten with a work function of 4.52 volts has a maximum ratio of 1.5. Contamination with oxygen increases the work function to 9.25 volts but increases rather than decreases the maximum ratio.[2] This probably means that electrons are more readily knocked out of the surface atoms and so give increased secondary emission even though they require more energy to escape from the surface. For a given combination of elements, however, the secondary emission usually increases with decreasing work function. Thus, if caesium on caesium oxide on silver is contaminated with oxygen, the work function increases and the secondary emission decreases. Also, in the case of molybdenum partly coated with barium the work function passes through a maximum with a given percentage of the surface covered, as is evidenced by the photoelectric emission. The secondary emission passes through a maximum with the same

---

[1] Bruining and DeBoer, *op. cit.*, Part V.

[2] Zworykin, V. K., and G. A. Morton, "Television," p. 32, Wiley, New York, 1940.

percentage of surface coverage though the maximum is not nearly so pronounced.[1]

Secondary Emission of Insulators. Insulators as well as conductors may emit secondary electrons. Measurements on insulators are more difficult to make because the potential of the insulator cannot be measured directly. The characteristics can, however, be deduced from the potential that the insulator assumes relative to a spherical collector electrode when bombarded with electrons of different potentials. The general features of the secondary emission of insulators may be summed as follows:[2] Insulators exhibit curves of ratio of secondary- to primary-electron current versus primary-electron potential that are similar to those of the metals. Ratios usually exceed 1 over a considerable range of potentials, a maximum occurring between 300 and 800 volts. As with the metals, the ratio rises rapidly to a maximum and then drops slowly. As with the metals, most of the secondary electrons are emitted perpendicularly to the surface, following very nearly a cosine law of distribution

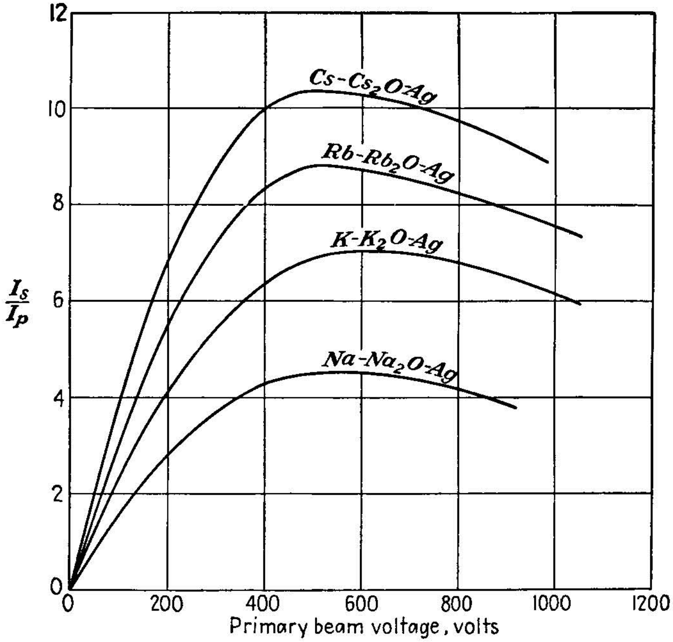


FIG. 4.20.—Secondary emission of the alkali oxides. (*Reprinted by permission from "Television" by V. K. Zworykin and G. A. Morton, Wiley, New York,* 1940.)

---

[1] BRUINING and DEBOER, *op. cit.*, Part VI.

[2] KOLLATH, *op. cit.*

regardless of the direction of the primary electrons. Upon bombardment at an angle the secondary- to primary-electron ratio increases as the primary electrons strike more nearly parallel to the surface up to a critical angle, beyond which the ratio drops to a small value less than unity and remains nearly constant. The critical angle depends upon the material and is a function of temperature, the angle with the normal increasing with temperature. The explanation of the sudden drop in emission with increased angle with the normal of primary-electron bombardment seems to be that a layer of negative charge forms on the surface which traps, by a space-charge action, the primary electrons and the secondary electrons they would have freed.

In normal action an insulator will have its potential influenced by its secondary-emission characteristics. The action will depend upon the primary-electron energy relative to the secondary-emission characteristics. Action can be divided into three cases as follows:

1. Primary-electron potential below that at which secondary- to primary-current ratio is unity. Here the number of secondary electrons emitted is less than the number of primaries, and so the insulator acquires a negative potential that is large enough to repel most of the primaries. This constitutes a blocking action. The insulator is finally in stable equilibrium at zero potential.
2. Secondary- to primary-current ratio greater than unity. Under this condition the insulator gives off more electrons than it acquires and so becomes more positive than its surroundings. When this happens, the insulator reattracts the slow secondaries and so remains a few volts more positive than the potential through which the primary electrons have been accelerated.
3. Primary-electron potential greater than that at which secondary- to primary-current ratio has dropped to unity. In such cases the insulator will gain more electrons than it loses and so will become more negative in potential until the primary electrons are retarded to the point where the ratio of secondary to primary current is unity. At this potential, the primary- and secondary-electron currents are equal, and the insulator is in stable equilibrium.

CHAPTER 5

DETERMINATION OF POTENTIAL FIELDS

THE fundamental theoretical technique necessary for the study of the internal behavior of a vacuum tube is that of determining the distribution of the electric potential within the tube. From the determination of the electric potential within a tube can be deduced the amplification factor of the tube, the focusing properties of the electrodes, and the current-voltage characteristics. In short, the determination of the distribution of the electric potential within a tube is the point of departure for the study of almost all its characteristics.

The methods of determining the potential fields of vacuum tubes are rather special. The most extensive information is obtained from conformal transformations and from solutions of the Laplace differential equation. The particular transformations and functional forms most frequently encountered in tubes are ordinarily given only a fraction of the total space allotted to the entire subject of electrostatics in books devoted to this subject. For this reason a brief review will be given of all the standard methods of determining potential fields, including some numerical and graphical methods, so that the elegance of the special methods mentioned will be appreciated.

5.1. Units and Dimensions. In this book there will be used the system of rationalized mks units. For this system the units of length, mass, and time are the meter, kilogram, and second, respectively and the electrical units are the usual practical ones—the volt, the ampere, the coulomb, etc. The term "rationalized" indicates that the factor 4π has been incorporated into the arbitrary constants in such a way that the greatest over-all simplicity of all relations is obtained. This is done in such a way that the factor 4π does not appear in relations involving plane geometry and rectangular coordinates but does appear in relations involving spherical geometry. A further feature of the rationalized mks system of units is that the equivalent dielectric constant of free space and the equivalent permeability of free space are not unity but have some specific values. These are the only two values that need to be known in this system to work practical problems, whereas in some of the other systems a whole table of conversion factors has to be invoked every time a practical problem is solved.

5.2. Fundamental Quantities and Definitions: Forces between Charges. All electrostatic relations are based upon the application of the observed effects of charges upon one another. Qualitatively, the observations are that there are two kinds of charges, that like repel and unlike attract, that the force between charges decreases as the distance between them increases. Quantitatively, all this is expressed by *Coulomb's law*,

$$F = \frac{q_1 q_2}{4\pi\varepsilon r^2} \tag{5.1}$$

where F is the radially directed force in newtons (1 newton equals 10^5 dynes) between charges q_1 and q_2 in coulombs, r is the distance between charges in meters, and ε is the so-called "dielectric constant" of the medium. The dielectric constant is equal to the product of the relative dielectric constant and the dielectric constant of free space,

$$\varepsilon = \varepsilon_r \varepsilon_0 \tag{5.2}$$

where ε_r is the relative dielectric constant as would be determined by the ratio of capacity of a condenser using the medium and free space as dielectric and ε_0 is the equivalent dielectric constant of free space whose value turns out to be 8.85×10^{-12} farad per meter in rationalized mks units.

The region in the vicinity of electric charges is referred to as the *electric field*. The *electric intensity* E at any point in such a field is the *force per unit charge* on a small test charge placed at the point. The intensity, which will also be shown to be the negative gradient of the electric potential, is a *vector* quantity in that it has both magnitude and direction.

Intensity at a distance r from a charge q is, by Coulomb's law,

$$|E| = \frac{q}{4\pi\varepsilon r^2} \tag{5.3}$$

Where more than one charge is concerned,

$$E_x = \frac{\sum_n q_n \cos(x,r_n)}{4\pi\varepsilon r_n{}^2} \tag{5.4}$$

$$E_y = \frac{\sum_n q_n \cos(y,r_n)}{4\pi\varepsilon r_n{}^2} \tag{5.5}$$

The summation must be taken by a summation of components where (x,r_n) is the angle between a line parallel to the x axis and the vector from the charge q_n to the point at which the intensity is being determined.

A line of force, or a line of flux, is a line drawn so that it has everywhere the direction of the electric intensity. Lines of flux originate on positive charges and terminate on negative charges. In the rationalized mks system of units *one line of flux emanates from every unit positive charge.* The density of the flux lines is known as the *displacement* or flux density. Displacement and intensity are related by the expression

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}^* \tag{5.6}$$

where $\boldsymbol{D}$ is the displacement, or number of flux lines per square meter, and ε is the dielectric constant of the medium. Equation (5.6) is, for homogeneous isotropic dielectrics, strictly analogous to the expression $\boldsymbol{B} = \mu \boldsymbol{H}$, which applies for magnetic fields.

The potential at any point in an electric field is defined as the *work per unit charge* required to bring a small positive test charge from infinity to the point in question (symbol V). Potential is a scalar quantity, *i.e.*, completely specified when its magnitude alone is given. Applying this definition to obtain the potential at a distance r from a charge q,

$$V = \int_{\infty}^{r} F\,dr = -\int_{\infty}^{r} \frac{q}{4\pi\varepsilon r^2}\,dr = \frac{q}{4\pi\varepsilon r} \tag{5.7}$$

The minus sign appears because the work is being done against the force. The potential obtained above is in volts if q is in coulombs and r is in meters. The work is independent of the path. The potential at a point due to a number of charges is equal to the sum of the potentials due to the separate charges,

$$V = \sum_{n} \frac{q_n}{4\pi\varepsilon r_n} \tag{5.8}$$

For a continuous distribution of charge over a surface,

$$V = \frac{1}{4\pi\varepsilon} \int \frac{\sigma}{r}\,da \tag{5.9a}$$

where σ is the surface density of charge, da is the element of area, and the integration is taken over the area of the surface. For a continuous distribution of charge throughout a volume,

$$V = \frac{1}{4\pi\varepsilon} \int \frac{\rho}{r}\,dv \tag{5.9b}$$

---

* Bold-faced capitals will be employed to designate vector quantities when used in the vector sense. Components of vectors are themselves vectors but may usually be treated as scalar quantities when dealt with separately.

where ρ is the volume density of charge, dv is an element of volume, and the integration is taken over the volume.

The difference of potential between two points in an electric field is defined as the work per unit charge required to bring a small positive test charge from one point to the other. This difference is independent of the path by which it is evaluated.

From the definition of potential it is seen that the intensity is the negative gradient of potential, the negative sign indicating that the force is exerted in a direction opposite to that of increasing potential. The gradient of the potential is a vector having the magnitude and direction of the maximum variation of potential. Thus

$$|\mathbf{E}| = -\frac{\partial V}{\partial s} \tag{5.10}$$

The force per unit charge in any general direction is given by

$$E \cos \alpha = -\frac{\partial V}{\partial l} \tag{5.11}$$

where α is the angle between the direction considered and the gradient of potential. Components of intensity are conveniently related to potential by

$$E_x = -\frac{\partial V}{\partial x} \tag{5.12a}$$

$$E_y = -\frac{\partial V}{\partial y} \tag{5.12b}$$

The form that components of intensity have in terms of derivatives of potential depends upon the coordinates in which the potential and distances are expressed. In all cases the component expressions corresponding to Eq. (5.10) have the form of the limiting value of the ratio of an increment of potential to an increment of length in the direction of the variable considered. Expressions for the intensity as a negative gradient of potential are given in Appendix II for the coordinate systems most commonly used.

5.3. Solution of Potential Fields by Summation of Intensities. The electric field around any distribution of charges may be found by summing the forces due to the charges by means of Eq. (5.4). Forces are best summed one component at a time. The procedure can usually be simplified by choosing the axes to take advantage of any symmetries. When an expression for each of the components of intensity has been found, the resultant intensity has a magnitude that is the square root of the sum of the squares of the components. The direction cosines of the resultant vector are given by the ratio of the respective components to the magnitude of the resultant.

Example: Find the electric intensity on the axis of a right-circular cylinder of radius a and length h at a distance x_0 from the end of the cylinder if the cylinder has a charge uniformly distributed throughout its volume of density ρ. In the configuration of Fig. 5.1 let x be the distance from the point P to the point on the axis corresponding to an element of volume in the cylinder. The elementary volume is given by

$$dv = r\,dr\,d\theta\,dx$$

and the corresponding element of charge is given by

$$dq = \rho\,dv$$

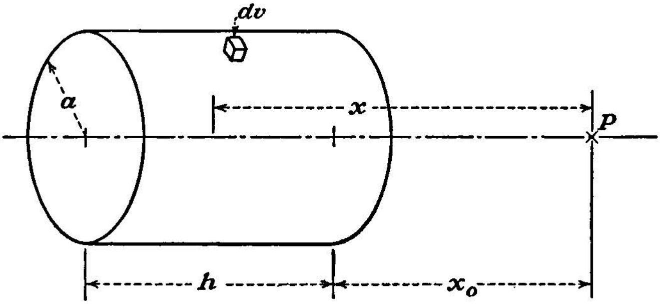


FIG. 5.1.—Notation for the evaluation of the axial intensity due to a cylindrical distribution of charge.

By symmetry there will be only an x component of intensity at the point P on the axis to which the element of charge will contribute

$$dE_x = \frac{\rho r\,dr\,d\theta\,dx}{r^2 + x^2}\,\frac{x}{(r^2 + x^2)^{1/2}}\,\frac{1}{4\pi\epsilon}$$

which will be recognized as being of the form

$$\frac{dq\cos\alpha}{R^2}$$

This differential expression must be integrated with respect to its three variables, θ from 0 to 2π, r from 0 to a, and x from x_0 to $x_0 + h$. When this triple integration has been performed, the resulting expression for the intensity on the axis is

$$E_x = 2\pi\rho\left[h - \sqrt{(x_0 + h)^2 + a^2} + \sqrt{x_0^2 + a^2}\right]\frac{1}{4\pi\epsilon}$$

5.4. Summation of Potentials. The potential at any point in a field may similarly be obtained by application of Eq. (5.7). This procedure is in general easier to apply than the direct evaluation of the intensity, for the summation for potentials is algebraic, whereas that for intensities must be vectorial. The expression for the components of intensity is, of course, derivable from the expression for potential.

Example: Find the potential at a distance c from the center of a spherical shell with inner and outer radii r_1 and r_2 and with a charge uniformly distributed throughout its volume of charge density ρ.

The element of volume in spherical coordinates is

$$dv = r^2 \sin\theta \, dr \, d\theta \, d\phi$$

where the symbols have the significance indicated in Fig. 5.2 and ϕ is the azimuthal angle. Then the potential at the point P due to the element of charge associated with the above element of volume is

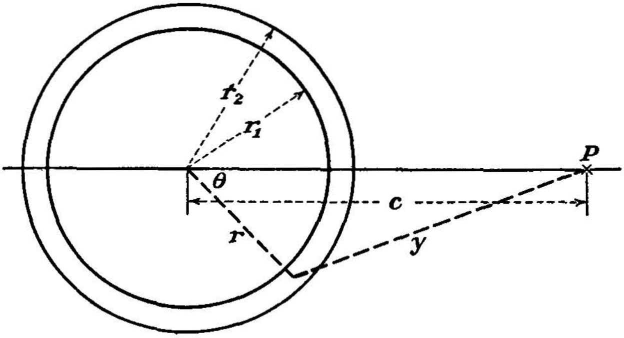


FIG. 5.2.—Notation for the evaluation of the potential due to a charge uniformly distributed throughout a spherical shell.

$$dV = \frac{\rho r^2 \sin\theta \, dr \, d\theta \, d\phi}{y} \frac{1}{4\pi\varepsilon}$$

It is convenient to use the distance y instead of the angle θ as a variable. The two quantities are related by the law of cosines

$$y = (c^2 + r^2 - 2cr\cos\theta)^{1/2}$$

so that, for constant r,

$$dy = \frac{cr \sin\theta \, d\theta}{y}$$

Making this substitution into the expression for the element of volume,

$$dV = \frac{\rho r}{c} \, dy \, d\phi \, dr \, \frac{1}{4\pi\varepsilon}$$

so that

$$V = \frac{\rho}{c} \int_{r_1}^{r_2} \int_0^{2\pi} \int_{c-r}^{c+r} r \, dy \, d\phi \, dr \, \frac{1}{4\pi\varepsilon}$$

The result of this integration gives

$$V = \frac{4\pi\rho}{3}\,\frac{r_2^3 - r_1^3}{c}\,\frac{1}{4\pi\varepsilon}$$

From this it is seen that the potential at the point P is the same as though the entire charge of the shell were concentrated at its center.

5.5. Gauss's Law. Gauss's law is one of the most useful relations in electrostatics. It enables one to determine quickly the field and potential around any symmetrical distributions of charge. The law may be stated as follows: *The integral of the normal outward component of electric flux over any closed surface is equal to a constant times the total charge enclosed by the surface.* For rationalized mks units, the constant is unity.

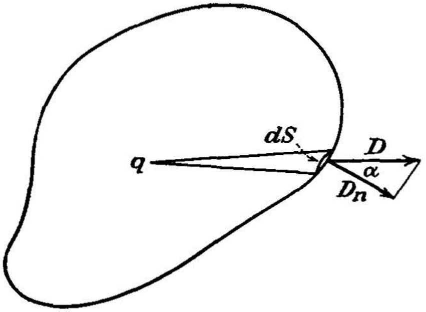


FIG. 5.3.—Notation for the evaluation of Gauss's law, Eq. (5.17).

Consider a closed surface S enclosing a single point charge q as shown in Fig. 5.3. Then the outward component of electric flux for the element of area dS is

$$D_n\,dS = D\cos\alpha\,dS \tag{5.13a}$$

$$D_n\,dS = \frac{q}{4\pi r^2}\cos\alpha\,dS \tag{5.13b}$$

It will be recognized that $\dfrac{dS}{r^2}\cos\alpha$ is the element of solid angle about the point charge intercepted by the area dS, since solid angle is measured by area intercepted on a unit sphere just as linear angle may be measured by arc length on a unit circle. Thus

$$d\Omega = \frac{dS\cos\alpha}{r^2} \tag{5.14}$$

where $d\Omega$ is an element of solid angle. Then

$$D_n\,dS = \frac{q\,d\Omega}{4\pi} \tag{5.15}$$

If this is integrated over the entire surface surrounding the point charge,

$$\int D_n\,dS = q \tag{5.16}$$

since there are 4π units of solid angle around a point.

Since the law of superposition holds for the potentials due to charges, the integral of the outward normal component of flux is equal to the total charge enclosed when the closed surface contains more than a single charge.

For a volume distribution of charge the law can be written

$$\int D \cos \alpha \, dS = \int \rho \, dv \tag{5.17}$$

where ρ is the volume charge density, v indicates volume, and the other symbols have the previous significance.

Example: Consider the case of a uniform distribution of charge on a circular wire of infinite length. From considerations of symmetry it is evident that the electric field will everywhere be radial and will be constant along the length of the wire. The equipotential surfaces will be cylinders concentric about the wire, and the flux lines will be straight radial lines.

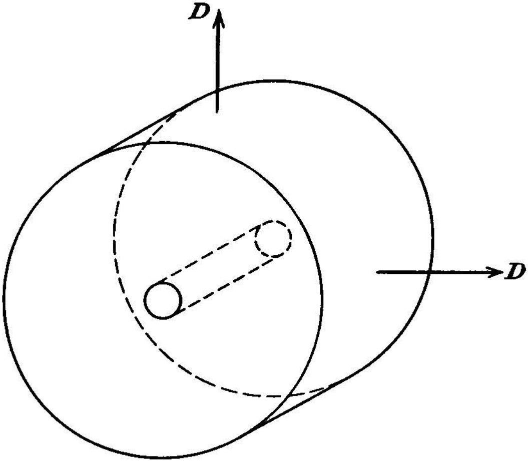


FIG. 5.4.—The flux associated with a linear distribution of charge.

Let the charge be uniformly distributed along the wire with a density of λ units per unit length. Draw a cylinder of radius r about the wire of radius a. Then the electric flux $D = \varepsilon E$ is everywhere outwardly directed as shown in Fig. 5.4. The integral of normal component of flux per unit length of this wire is equal to the product of the displacement and the area of the cylinder per unit length. This product must be equal to the linear charge density, so that

This gives the intensity at any distance r from a wire with a linear charge density.

The potential at any distance r from the wire is found by integrating the negative of the field with respect to r, giving

$$V = -\frac{\lambda}{2\pi\varepsilon}\ln r + C$$

The constant is necessary to adjust the potential to a prescribed value at some particular distance since the potential about a cylindrical wire, unlike that about a point charge, does not vanish at an infinite value of the radius.

In the case of two concentric cylinders of radii r_2 and r_1 having potentials V_2 and V_1, respectively, the potential between them is

$$V(r) = V_1 + (V_2 - V_1)\frac{\ln\frac{r}{r_1}}{\ln\frac{r_2}{r_1}}$$

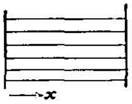
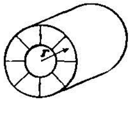
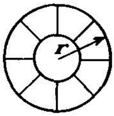
If

$$V_1 = 0$$

$$V(r) = \frac{V_2 \ln\frac{r}{r_1}}{\ln\frac{r_2}{r_1}}$$

From Gauss's law it may also be deduced that the field adjacent to a plane with a surface charge density σ is given by $\frac{\sigma}{\varepsilon}$ and is normal to the plane. It may also be verified that a charge uniformly distributed throughout a sphere or over the surface of a sphere looks to an observer outside the sphere as though the charge were all concentrated at the center of the sphere so that the laws for point charges hold.

The above results are summarized in the following table:

	Plane	Cylindrical	Spherical
Geometry			
Total positive charge............	$\sigma \times$ area	$\lambda \times$ length	q
Charge density..................	σ per unit area	λ per unit axial length	$\frac{q}{\text{area}}$
Intensity E....................	$\frac{\sigma}{\varepsilon}$	$\frac{\lambda}{2\pi\varepsilon r}$	$\frac{q}{4\pi\varepsilon r^2}$
Potential V....................	$-\frac{\sigma x}{\varepsilon} + C$	$-\frac{\lambda}{2\pi\varepsilon}\ln r + C$	$\frac{q}{4\pi\varepsilon r}$

5.6. Poisson's and Laplace's Equations. Poisson's and Laplace's equations are differential expressions of Gauss's law applied to an element of volume. Poisson's equation applies to regions containing charge. Laplace's equation is the same equation for the case of no charge. The equations are derived as follows:

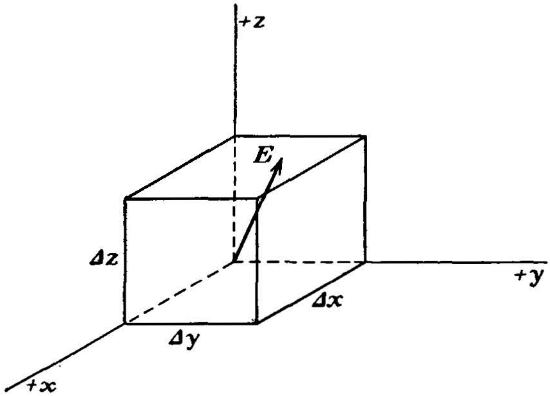


FIG. 5.5.—Notation used in the derivation of Laplace's equation in rectangular coordinates, Eq. (5.24).

Consider an element of volume in an electric field as shown in Fig. 5.5. If the intensity at the origin is E, then

$$\text{Flux into back face} = \varepsilon E_x \,\Delta y\, \Delta z$$

$$\text{Flux out of front face} = \varepsilon \left(E_x + \frac{\partial E_x}{\partial x}\,\Delta x\right) \Delta y\, \Delta z$$

$$\text{Net outward flux through front and back faces} = \varepsilon \frac{\partial E_x}{\partial x}\,\Delta x\,\Delta y\,\Delta z$$

Similarly

$$\text{Net outward flux through left and right faces} = \varepsilon \frac{\partial E_y}{\partial y}\,\Delta x\,\Delta y\,\Delta z$$

and

$$\text{Net outward flux through bottom and top faces} = \varepsilon \frac{\partial E_z}{\partial z}\,\Delta x\,\Delta y\,\Delta z$$

Upon combining these, the outward flux through all faces is

$$\varepsilon \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) \Delta x\,\Delta y\,\Delta z = \rho\,\Delta x\,\Delta y\,\Delta z \tag{5.18}$$

by Gauss's law where ρ is the volume charge density. The above equality is abbreviated

$$\text{Divergence } E = \nabla \cdot E = \frac{\rho}{\varepsilon} \tag{5.19}$$

in which the element of volume has been cancelled and the term "divergence" has been applied to the limiting value of the net outward flux per element of volume as the element of volume approaches zero.

But

$$\boldsymbol{E} = \text{negative gradient of } V \tag{5.20a}$$

frequently abbreviated

$$\boldsymbol{E} = -\nabla V \tag{5.20b}$$

or in component form

$$\begin{aligned} E_x &= -\frac{\partial V}{\partial x} \\ E_y &= -\frac{\partial V}{\partial y} \\ E_z &= -\frac{\partial V}{\partial z}^* \end{aligned} \tag{5.21}$$

Making these substitutions into Eq. (5.18),

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\varepsilon} \tag{5.22}$$

which is Poisson's equation. This is abbreviated

$$\nabla^2 V = -\frac{\rho}{\varepsilon} \tag{5.23}$$

In a region free of charge, $\rho = 0$ so that

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{5.24}$$

which is Laplace's equation. This is abbreviated

$$\nabla^2 V = 0 \tag{5.25a}$$

If the derivation is made in terms of general coordinates u_1, u_2, and u_3 with scale factors h_1, h_2, and h_3, respectively, so that an element of arc length is related to the coordinates and scale factors by

$$ds^2 = h_1^2\, du_1^2 + h_2^2\, du_2^2 + h_3^2\, du_3^2$$

then Laplace's equation assumes the general form

$$\nabla^2 V = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial u_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial V}{\partial u_1}\right) + \frac{\partial}{\partial u_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial V}{\partial u_2}\right) + \frac{\partial}{\partial u_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial V}{\partial u_3}\right)\right] \tag{5.25b}$$

Interpretations of Laplace's Equation. As has been mentioned before and as is evident from the development of the equation, Laplace's equation is a differential expression of Gauss's law for an element of volume.

---

* See Appendix II.

In the language of differential equations it says that the net electric flux emerging from an element of volume in a region free of charge is zero.

Another interpretation that can be given to Laplace's equation in the two-dimensional case is that it is an equivalent way of saying that the potential at any point in a field is the average of the potentials at four equally spaced surrounding points. Thus if there is given a set of curves of equal potentials in the vicinity of some electrodes, known as a "contour representation of potentials," then the potential at any point, say the point (2,2), is the average of the potentials at the four surrounding coordinate points, for the case assumed, the average of the potentials at the points (2,1), (3,2), (2,3), and (1,2). This property will be proved in a subsequent section.

Laplace's equation can also be interpreted in terms of the curvature of the potential profiles of a field configuration. Two-dimensional fields can be represented either by contours of equipotential or by potential profiles just as we can draw either a contour map or a set of profiles for a topographic representation of terrain. In the profile representation we draw potential as an ordinate against distance along some line as abscissa. It will be remembered from elementary calculus that the curvature of any curve is given by

$$\frac{1}{R} = K = \frac{\dfrac{d^2y}{dx^2}}{\left[1 + \left(\dfrac{dy}{dx}\right)^2\right]^{3/2}} \tag{5.26}$$

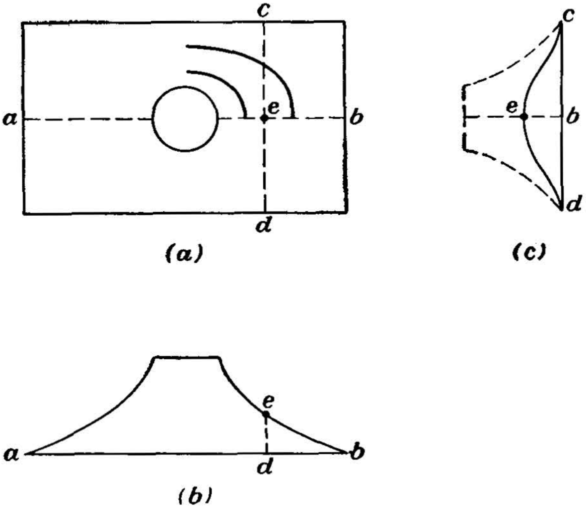
from which it is seen that the *sign* of the curvature is determined by the sign of the second derivative in the numerator since the denominator is always positive. If we now examine Laplace's equation in two dimensions,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial^2 y} = 0 \tag{5.27}$$

we see that the two terms may be interpreted as giving the sign of the curvature of the profiles in the x and y directions. By Eq. (5.27) the curvatures must be of opposite nature since the sum of the terms is zero; and hence if the profile in the x cut at some point in the field is concave upward, then the profile in the y cut at the same point must be concave downward.

Examination of a simple case will illustrate the property described above. In Fig. 5.6*a* is shown the contour representation for the case of a concentric line with a circular inner conductor and a rectangular outer conductor. The solid lines represent the equipotential lines or contours. In Fig. 5.6*b* is shown the potential profile along the line *ab*, and it will be seen that the profile is curved *away from the axis* at the point e. In Fig. 5.6c is shown the potential profile along the line cd, and it will be seen that the profile is curved *toward the axis* at the point e.

Solutions of Laplace's Equation in Two Dimensions. The form which the solutions of Laplace's equation take depends upon the coordinates in which the equation is expressed. For *rectangular coordinates* Laplace's equation has the form of Eq. (5.27). The solutions of this equation have the form



(a) (c) (b)

FIG. 5.6.—Example showing the relation between the curvatures of the profiles of a potential field.

$$V = (A \cos kx + B \sin kx)(C \cosh ky + D \sinh ky) \qquad (5.28a)$$

or

$$V = (A \cos ky + B \sin ky)(C \cosh kx + D \sinh kx) \qquad (5.28b)$$

The above results are arrived at by assuming that V has a solution of the form XY where X is a function of x alone and Y is a function of y alone. If the product XY is substituted for V in Eq. (5.27), there results upon differentiation and rearrangement

$$\frac{1}{X}\frac{d^2X}{dx^2} = -\frac{1}{Y}\frac{d^2Y}{dy^2} \qquad (5.29)$$

It is seen that the left-hand member is a function of x alone and that the right-hand member is a function of y alone. These can be equal only if each equals the same constant. If this constant is taken as $-k^2$, then we may write two component equations in the place of Eq. (5.29),

$$\frac{1}{X}\frac{d^2X}{dx^2} = -k^2 \tag{5.30}$$

and

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = +k^2 \tag{5.31}$$

The solution of Eq. (5.30) is

$$X = A \cos kx + B \sin kx \tag{5.32}$$

and the solution of Eq. (5.31) is

$$Y = C \cosh ky + D \sinh ky \tag{5.33}$$

Thus V is given by the product of X and Y, resulting in the solution of Eq. (5.28*a*) where multiple values of k as determined by imposed conditions are allowed. If the separation constant is chosen as $+k^2$ instead of $-k^2$, then the solution of Eq. (5.28*b*) results.

For the *polar coordinates* of Fig. 5.7 Laplace's equation has the form

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} = 0 \tag{5.34}$$

when the problem is one of axial symmetry. This has a solution in the form

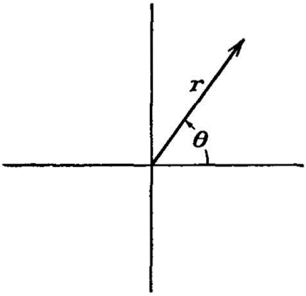


FIG. 5.7.—Polar coordinate notation.

$$V = (a \cos n\theta + b \sin n\theta)(cr^n + dr^{-n}) \tag{5.35}$$

as may be shown by the method demonstrated above using n^2 as the separation constant. When n equals zero the second factor in Eq. (5.35) is $c + d \ln r$.

For the *cylindrical coordinates* of Fig. 5.8 Laplace's equation, for cases of axial symmetry, has the form

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) + \frac{\partial^2 V}{\partial z^2} = 0 \tag{5.36}$$

This has a solution of the form

$$V = [aJ_0(kr) + bN_0(kr)](c \sinh kz + d \cosh kz) \tag{5.37}$$

where J_0 and N_0 are the zero-order Bessel and Neumann functions. Since the Neumann function of zero is infinite, this term is not often encountered in electronics problems. Most potential configurations have a finite potential along the axis of symmetry, as in electrostatic electron lenses where there is no conductor along the axis.

In order to apply the above solutions to definite problems it is necessary to evaluate the constants in such a way that the potential function fits the prescribed boundary conditions. If the constants can be selected so that the function fits all the boundaries (electrodes), then it will define the potential at all points in the field. The potential solutions frequently appear as a series summation of terms of the form indicated above.

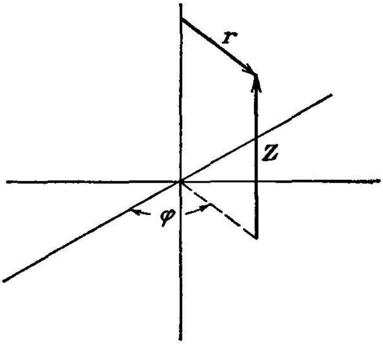


FIG. 5.8.—Cylindrical coordinate notation.

Difference Form of Laplace's Equation. We may write Laplace's equation in the form of a difference equation of which the differential equation is the limiting form. To do this we shall assume that the potential is known at a number of points whose spacing is finite though small. We shall assume that the points are at the intersections of a rectangular lattice as shown in Fig. 5.9 and that the spacing between the points is h,[1-3]. The conclusions that we shall draw from the difference equation set up on this basis will apply also to the differential equation and its solution.

Consider the first derivative of potential at the point (0) in the xy plane. The difference operators corresponding to the partial derivatives are given by

$$\frac{\partial V}{\partial x} = \frac{1}{h}(V_1 - V_0) \tag{5.38}$$

$$\frac{\partial V}{\partial x} = \frac{1}{h}(V_0 - V_3) \tag{5.39}$$

$$\frac{\partial V}{\partial y} = \frac{1}{h}(V_2 - V_0) \tag{5.40}$$

$$\frac{\partial V}{\partial y} = \frac{1}{h}(V_0 - V_4) \tag{5.41}$$

---

[1] MORSE, P. M., and HERMAN FESHBACH, "Methods of Theoretical Physics," Massachusetts Institute of Technology, 1946, pp. 139–147.

[2] SHORTLEY, G. H., and R. WELLER, The Numerical Solution of LaPlace's Equation, *Jour. Appl. Phys.*, vol. 9, pp. 334–348, May, 1938. Probably the best single reference on this subject.

[3] FROCHT, M. M., and M. M. LEVIN, A Rational Approach to the Numerical Solution of LaPlace's Equation, *Jour. Appl. Phys.*, vol. 12, pp. 596–604, August, 1941.

The difference operators corresponding to the second derivatives are given by

$$\frac{\partial^2 V}{\partial x^2} = \frac{1}{h^2}[(V_1 - V_0) - (V_0 - V_3)] \tag{5.42}$$

$$\frac{\partial^2 V}{\partial y^2} = \frac{1}{h^2}[(V_2 - V_0) - (V_0 - V_4)] \tag{5.43}$$

Upon substituting these values in Laplace's equation there results

$$V_0 = \tfrac{1}{4}(V_1 + V_2 + V_3 + V_4) \tag{5.44}$$

which states that the potential at the center of a square is the average of the potentials at the corners of the square.

It is possible to obtain numerical values of potential for various electrode configurations by means of Eq. (5.44). The procedure is to break up the field whose potential is desired into a suitable lattice, assume reasonable values of potential at each point in the lattice, and then apply Eq. (5.44) successively to each of the points, always using any new values of potential obtained. Successive applications of this procedure will correct any errors in the original assumptions, and the values of potential at any point will converge quite rapidly to the correct value. It is well to start with a coarse network and then make it finer.

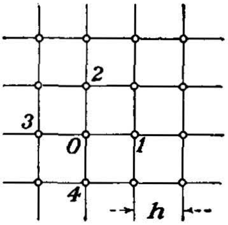


FIG. 5.9.—Arrangement of net points for the difference form of Laplace's equation in two-dimensional rectangular coordinates, Eq. (5.44).

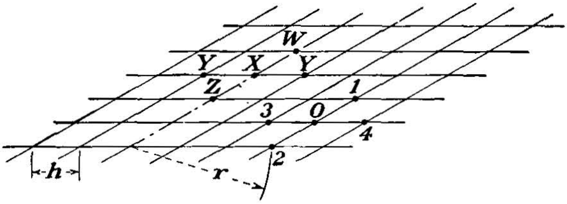


FIG. 5.10.—Arrangement of net points for the difference form of Laplace's equation in two-dimensional cylindrical coordinates, Eqs. (5.45) and (5.46).

The expression given in Eq. (5.44) was derived for two-dimensional rectangular coordinates. For two-dimensional problems of axial symmetry expressed in terms of the cylindrical coordinates of Fig. 5.8 that hold for electrostatic electron lenses, and the like, the corresponding expression for the lattice of Fig. 5.10 is

$$V_0 = \frac{1}{4}(V_1 + V_2 + V_3 + V_4) + \frac{h}{8r}(V_4 - V_3) \qquad (5.45)$$

in which the points 1 and 2 are on a line parallel to the axis and point 3 is closer to the axis than point 4. The expression [Eq. (5.45)] works for all parts of the field except points on the axis for which

$$V_x = \tfrac{1}{6}(V_w + V_z + 4V_y) \qquad (5.46)$$

It will be noticed that Eq. (5.45) reduces to Eq. (5.44) for large values of r.

The above manipulations for the cylindrical case can be simplified by a change of variable. If as a new variable there be taken

$$y = r^{1/2}V \qquad (5.47)$$

then the Laplace equation reduces to

$$\frac{\partial^2 y}{\partial r^2} + \frac{\partial^2 y}{\partial z^2} + \frac{y}{4r^2} = 0 \qquad (5.48)$$

The corresponding net-point equation is

$$y_0 = \frac{y_1 + y_2 + y_3 + y_4}{\left(4 - \dfrac{h^2}{4r^2}\right)} \qquad (5.49)$$

which is much simpler to apply than Eq. (5.45).

The case of two-dimensional polar coordinates can be reduced to the rectangular coordinate treatment by changing the variables according to

$$u = \ln_\epsilon r \qquad (5.50)$$

and

$$v = \theta \qquad (5.51)$$

For a lattice of equal increments of u and v, Eq. (5.44) applies directly. The reasons for this will become evident when the transformation $W = \ln_\epsilon Z$ has been studied.

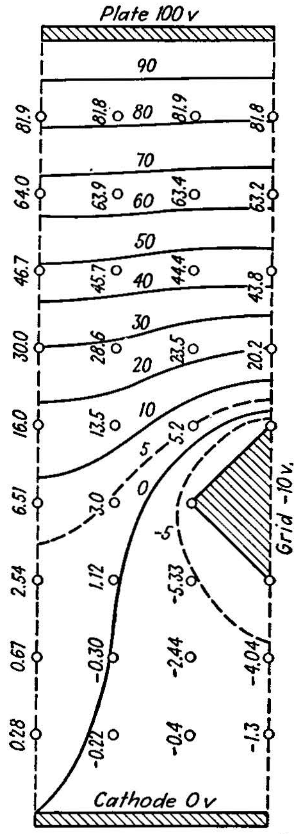


FIG. 5.11.—The potential field of a triode, calculated from the difference form of Laplace's equation, Eq. (5.44).

In Fig. 5.11 is shown the potential field inside a half section of a plane-electrode triode as calculated from repeated application of the difference form of Laplace's equation, Eq. (5.44).

5.7. Elastic-membrane Models of Potential. It is possible to represent two-dimensional potential problems having a *z-axis* symmetry by the elevation of a deformed elastic membrane. If an elastic membrane is uniformly stretched in all directions and leveled when suspended in a plane frame and is then deformed from its original plane by displacing the membrane distances proportional to electrode voltages with blocks shaped like the electrodes to be studied, then the displacement of the membrane at any point from the original plane is proprotional to the potential at that point in the field. In other words, the membrane is a topographic model of the potential field with vertical displacement proportional to potential. The deformed surface that is obtained is a very good representation of the potential field. This is because the surface will deform itself so that its area will be a minimum. Analytically this is expressed by making

$$S = \int\int \sqrt{1 + \left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}\, dx\, dy = \min \tag{5.52}$$

where z is the elevation and x and y are the coordinates in the horizontal plane. This is a problem in the calculus of variations that is converted into a problem in differential equations by applying to the equation for S that is in the form

$$S = \int\int F\left(x,y,z,\frac{dz}{dx},\frac{dz}{dy}\right) dx\, dy \tag{5.53}$$

The Euler differential equation

$$\frac{\partial}{\partial x} F_{z_x} + \frac{\partial}{\partial y} F_{z_y} + F_z = 0 \tag{5.54}$$

where the subscripts indicate differentiation with respect to the following factors:

$$z_x = \frac{\partial z}{\partial x}$$

$$z_y = \frac{\partial z}{\partial y}$$

$$F_{z_x} = \frac{\partial F}{\partial z_x}$$

$$F_{z_y} = \frac{\partial F}{\partial z_y}$$

$$F_z = \frac{\partial F}{\partial z}$$

Application of the Euler equation yields as the differential equation of the deformed surface[1-3]

$$\left[1+\left(\frac{\partial z}{\partial y}\right)^2\right]\frac{\partial^2 z}{\partial y^2}+\left[1+\left(\frac{\partial z}{\partial x}\right)^2\right]\frac{\partial^2 z}{\partial x^2}+2\,\frac{\partial^2 z}{\partial x\,\partial y}\,\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}=0 \quad (5.55)$$

which reduces to Laplace's equation,

$$\frac{\partial^2 z}{\partial x^2}+\frac{\partial^2 z}{\partial y^2}=0 \quad (5.56)$$

for $\frac{\partial z}{\partial x} \ll 1$ and $\frac{\partial z}{\partial y} \ll 1$. If the angles of all lines on the surface are kept below 6 deg with the horizontal, the departure in deformation from that representing the true potential at any point will be less than 1 per cent. Some practical considerations are of importance. A No. 30 rubber surgical dam makes a good membrane. It should be stretched enough so that it will be tight and not sag and yet not be too close to the rubber's elastic limit. A linear stretch of about ⅓ works well. It helps in obtaining a uniform stretch to mark coordinate lines on the sheet before stretching and then stretch so that these are straight and of the proper spacing.

The applications of the elastic-membrane model of potential are somewhat limited, for it is accurate only for small deformations, it can be used to represent only two-dimensional problems with a z-axis (stacking) symmetry—it cannot exactly represent problems with a rotational symmetry about an axis—and it cannot be modified to include space-charge effects. In spite of these limitations, models of this sort have been used extensively by various laboratories in their studies of potential fields and electron paths; in the latter regard it yields much information in a short time. The use of the membrane in determining electron paths will be mentioned in a later section. Figure 5.12 shows the elastic membrane model used in the Electrical Engineering Department of Stanford University.

5.8. Current-flow Models of Potential. The laws which govern the flow of current in a uniformly conducting medium are the same as those which govern the "flow" of electrostatic-flux lines in a vacuum. This

---

[1] Kleynen, P. H. J. A., Motion of an Electron in Two Dimensional Electrostatic Field, *Philips Tech. Rev.*, vol. 2 (No. 11), pp. 338–345, 1937. Original article on this subject.

[2] Strutt, M. J. O., "Moderne Mehrgitter-Elektronenroehren," pp. 3–6, Springer, Berlin, 1938.

[3] Zworykin, V. K., and J. A. Rajchman, The Electrostatic Electron Multiplier, *Proc. I.R.E.*, vol. 27, pp. 558–565, September, 1939.

makes it possible to set up current-flow models of electrode systems and to measure the potential at any point.

The equations for the components of current density in a continuous and uniform medium such as some electrolyte are given by Ohm's law in terms of the gradient of potential as

$$J_x = -g \frac{\partial V}{\partial x} \tag{5.57a}$$

$$J_y = -g \frac{\partial V}{\partial y} \tag{5.57b}$$

for the two-dimensional case, where J is current density and g is the specific conductivity of the medium.

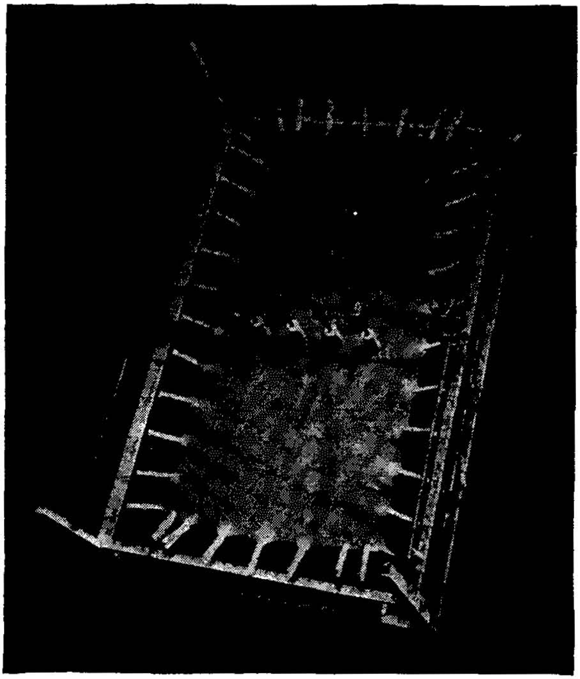


FIG. 5.12.—Elastic-membrane model of potential.

Since the flow of current in a medium of constant conductivity corresponds to an irrotational flow of an incompressible fluid, as much current will flow into any element of volume as flows out of it. This condition is sometimes expressed by saying that the divergence of the current is zero, which may be expressed mathematically as

$$\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} = 0 \tag{5.58}$$

Substitution of the components of Eq. (5.57) into Eq. (5.58) yields Laplace's equation in the form

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{5.59}$$

Upon comparing the above equations with those developed for the electrostatic field it is seen that an exact correspondence can be established. The relations may be tabulated in one-to-one correspondence as follows:

Currents	Electrostatic Fields
Quantities	
J Current density	D Displacement flux
g Specific conductivity	ε_r Dielectric constant
V Potential	V Potential
Relations	
$J_x = -g\frac{\partial V}{\partial x}, \qquad J_y = -g\frac{\partial V}{\partial y}$	$D_x = -\varepsilon\frac{\partial V}{\partial x}, \qquad D_y = -\varepsilon\frac{\partial V}{\partial y}$
$\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} = 0$	$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} = 0$
$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0$	$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0$

From the above tabulation it is seen that the correspondence between the current flow and electrostatic field is quite complete. It is thus necessary only to set up a current-flow model with electrodes geometrically similar to those of the electrostatic problem whose solution is desired and to measure the potential contours. The model is easily set up for two-dimensional problems by means of a flat tank. A weak solution of copper sulphate may be used as an electrolyte. This has a fairly good conductivity and has no polarizing action with copper electrodes. The

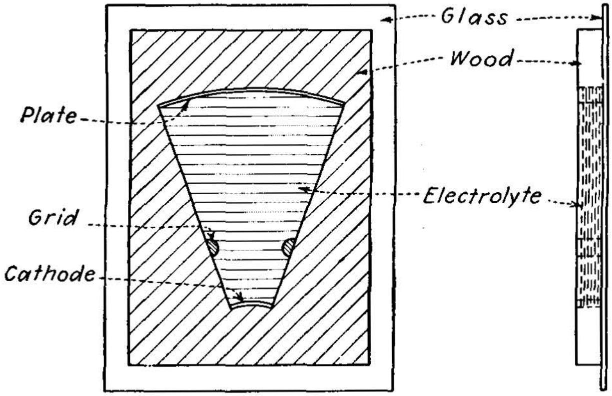


FIG. 5.13.—Current-flow model of a cylindrical triode.

equipotential contours can be traced with a probe connected to two resistances forming two arms of a bridge. The other arms of the bridge are in the electrolyte. By setting the external resistances any equipotential contour can be traced by observing the points at which a null indication is received. Tanks may be made of wood cemented to a glass plate. A large sheet of coordinate paper may be put under the glass to identify the location of points and thus facilitate their transfer to another sheet for plotting. It is also possible to use a pantograph for plotting directly. In using a flat tank it is absolutely necessary that the liquid be of the same depth at all points. Placing the tank upon a board with leveling screws makes it easy to level. Figure 5.13 shows a tank of the type described. This particular tank represents a section of a cylindrical electrode triode.

The arrangement of resistors used with triode current-flow models is shown in Fig. 5.14. The resistors R_p and R_g are used to set the relative positive plate and negative grid potentials. The resistors R_1 and R_2 in the bridge arms are used to determine the potential of the contour to be traced. If the resistors R_2 and R_1 are set in the ratio of 2 to 8, the probe will trace out the contour having 80 per cent of the plate-cathode potential, since the percentage voltage of the contour is given by $\dfrac{R_2}{R_1 + R_2}$. Headphones are conveniently used as balance detectors. It is also possible to use a cathode-ray oscilloscope. If the probe and R_1,R_2 junction are connected to the vertical plates and a voltage in phase with the electrode potentials is connected to the horizontal plates, there will result a straight-line Lissajous figure whose slope will be zero when the probe is in the proper position. The advantage of this arrangement is that the slope of the line will be negative or positive according to whether the probe is to one side or the other of the proper position. A low frequency of the order of 50 to 100 cycles should be used. If the frequency is too low, it is difficult to detect a null. If it is too high, the distributed capacities affect the balance.

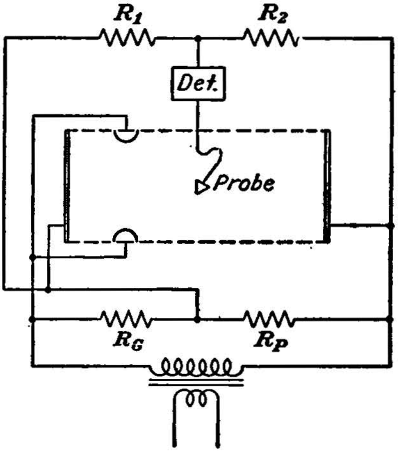


FIG. 5.14.—Circuit arrangement for measuring potential contours on a current-flow model of a plane-electrode triode section.

It should be observed in the model of Fig. 5.14 that the proper conditions of symmetry are obtained along the nonconducting boundaries which are indicated by dotted lines. Here the current flow must be parallel to the boundary, which ensures that the potential contours are at right angles to the nonconducting boundaries, since the equipotentials are perpendicular to the flow lines. It will also be true that the equipotentials will be perpendicular to all lines of symmetry running in the direction of the flow. Flow lines will be perpendicular to conducting surfaces. The flow will also be parallel to the top and bottom of the liquid layer since the air above and the glass below are nonconducting.

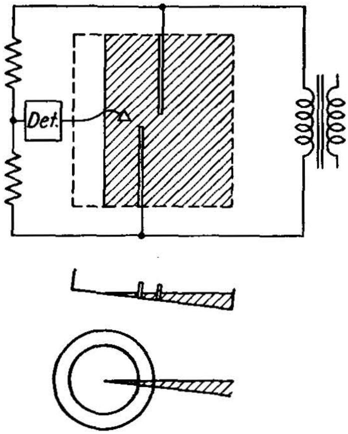


FIG. 5.15.—Arrangement for measuring potential contours on a current-flow model of an electron lens. The lens electrodes are cylinders of revolution that require a tilted tank to represent a wedge-shaped portion of the potential field. The edge of the wedge-shaped portion of electrolyte corresponds to the axis of the electrodes.

For problems involving axial symmetry such as are encountered in electron optics, a slightly different arrangement of electrodes must be used. Here it is necessary to reproduce conditions of axial symmetry and it is not correct to use a uniform depth of electrolyte as in Fig. 5.14 without special electrodes. To obtain correct results, either the electrodes or the volumetric shape of the electrolyte must be changed. It is possible to use a deep flat tank if the electrodes are shaped like portions of half cylinders with their edges at the surface of the electrolyte. For such an arrangement the probe should be kept at the surface of the electrolyte. A more convenient arrangement is to use a wedge-shaped electrolyte. Use of such a section corresponds to a pie-shaped section of small angle cut out of the field of revolution as shown in Fig. 5.15. The wedge-shaped volume of electrolyte is obtained by simply tilting a flat tank. Properly speaking the electrodes should be portions of cylinders, but if the angle of the electrolyte wedge is small enough, say less than 5 deg, they may be portions of planes without introducing any appreciable error.

5.9. Sketching of Flux and Potential Fields. The properties of electrostatic fields are such that, with a little practice, it is possible to sketch fields with considerable accuracy without recourse to mathematical methods. It is known, for instance, that flux and potential lines are everywhere at right angles to each other. Flux lines emerge at right angles from conducting surfaces. Potential lines near conductors tend to have the same shape as the conductor. These and other useful properties may be summarized as follows:

PROPERTIES USEFUL IN SKETCHING FIELDS

1. Flux and potential lines form orthogonal families of curves.
2. Flux lines are perpendicular to conductors at conductor surface.
3. Potential contours close to conductors tend to have the same form as the conductors.

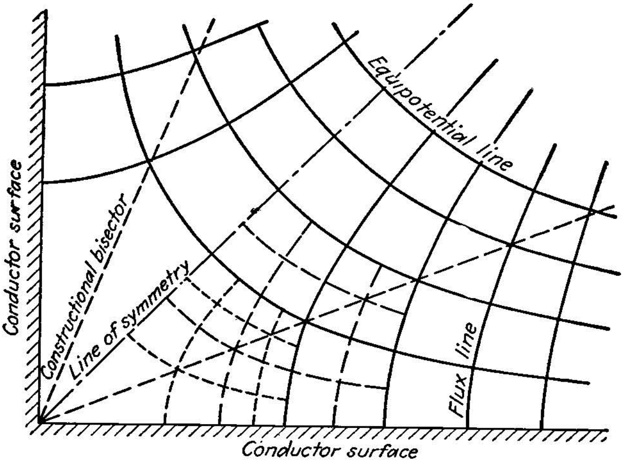


FIG. 5.16.—Sketch of flux and potential lines in an inside right-angled corner. This sketch was made by the method of Sec. 5.9 without mechanical or numerical aids.

4. Potential lines are parallel or perpendicular to lines of symmetry; constructional bisectors may exist.
5. Flux-potential patterns should be drawn with curvilinear squares, *i.e.*, a four-sided figure, with right angles at the corners and with equal average lengths of opposite sides, which maintains these properties upon infinite subdivision.

a. Same potential difference exists across each square.
b. Same flux passes through each square.
c. Each square has the same attraction for the conductor face.
d. Each square has the same energy storage.
e. Each flux line represents the same increment of capacity between electrodes.

Most of the above properties are self-evident. In Fig. 5.16 is shown a plot of the flux and potential inside of a right-angled corner. This plot was sketched, not calculated. It will be observed that all the curvilinear squares upon infinite subdivision will still be curvilinear squares. The principal line of symmetry is shown by the center line. The constructional bisectors are shown by dotted lines.

A flux plot to be of value should include

1. Shape of fields at large distances as well as small distances from the charges (conductors).
2. Location of all conductors and charges.
3. Geometrical symmetries of any kind.
4. All singular points, *i.e.*, "saddle" points, giving rise to a crossing of equipotential contours.

The above enumeration is actually quite general, and all these inclusions are not always necessary in electronic problems. Singular points occur where there is an apparent intersection of potential contours. This occurs only where the equipotential surface is saddle-shaped.[1,2]

5.10. Method of Conformal Transformations. The method of conformal transformation is based upon solutions of Laplace's equation in two-dimensional rectangular coordinates and functions of the complex variable $z = x + iy$. Most functions of the complex variable of the form

$$W = f(z) \tag{5.60}$$

are separable into real and imaginary parts

$$W = E(x,y) + iF(x,y) \tag{5.61}$$

in which each part is a solution of Laplace's equation. The two parts of the complex function, $E(x,y)$ and $F(x,y)$, further represent orthogonal families of curves. They may hence be taken as representing flux and equipotential lines. The functions having the above properties are known as *analytic* functions (to be defined more explicitly).

Every analytic function of the complex variable may thus be considered to represent the flux and potential field of some set of electrodes. Fields may further be transformed by means of analytic functions from one form to another. Thus, given a function that gives the field corresponding to one set of electrodes, the application of another function will transform this field into one corresponding to another set of electrodes. In the course of this transformation all the properties of flux and potential fields are preserved.

Analytic functions when used for making transformations have the property of preserving the angles between lines and of making corre-

---

[1] MOORE, A. D., Mapping of Magnetic and Electric Fields, *Elec. Jour.*, vol. 23, pp. 355–362, July, 1926.

[2] STEVENSON, A. R., and R. H. PARK, Graphical Determination of Magnetic Fields, *Trans. A.I.E.E.*, vol. 46, pp. 112–135, February, 1927.

sponding incremental areas similar in shape. It is for these reasons that the transformations are called "conformal."

Application of the method of conformal transformations usually takes the form of finding a transformation which converts the electrodes (equipotentials) of some simple field to the structure of which the field is desired. The field of the simple electrodes can usually be determined, and then the transformation converts the entire field to that of the more complex arrangement.

Conformal transformations are familiar to everyone in the form of maps. The surface of the earth may be mapped in many ways, which give apparently different shapes to the land masses. The different shapes are, however, merely different representations of the same thing. Most maps could be transformed from one form to another by means of conformal transformations, since the transformations would preserve the angles between river tributaries and keep the shape of small areas the same. An example of this idea is found in the logarithmic transformation, which, as will be shown, is capable of transforming a polar azimuthal equidistant projection of the Northern Hemisphere into what is approximately a Mercator projection of this hemisphere.

Complex Functions Satisfy Laplace's Equation. In studying conformal transformations it will first be shown that functions of the complex variable $z = x + iy$ are solutions of Laplace's equation in two-dimensional coordinates,

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0 \tag{5.62}$$

where

$$f(x,y) = f(x + iy) = f(z) \tag{5.63}$$

This follows since

$$\frac{\partial f}{\partial x} = \frac{df}{dz}\frac{\partial z}{\partial x} = \frac{df}{dz} \tag{5.64}$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{d^2 f}{dz^2} \tag{5.65}$$

Similarly

$$\frac{\partial f}{\partial y} = \frac{df}{dz}\frac{\partial z}{\partial y} = i\frac{df}{dz} \tag{5.66}$$

$$\frac{\partial^2 f}{\partial y^2} = -\frac{d^2 f}{dz^2} \tag{5.67}$$

It is evident that these partial derivatives are such as to satisfy Laplace's equation in the form of Eq. (5.62). The converse of this property is also true, *viz.*, that solutions of Laplace's equation in two-dimensional rectangular coordinates are expressible as functions of the complex variable.

Example: Let

$$\begin{aligned} f(z) &= Az^2 + Bz \\ &= Ax^2 + 2Aixy - Ay^2 + Bx + iBy \end{aligned}$$

Then

$$\frac{\partial f}{\partial x} = 2Ax + 2Aiy + 0 + B + 0$$

$$\frac{\partial^2 f}{\partial x^2} = 2A$$

And

$$\frac{\partial f}{\partial y} = 0 + 2Aix - 2Ay + 0 + iB$$

$$\frac{\partial^2 f}{\partial y^2} = -2A$$

Laplace's equation is seen to be satisfied.

It is also true that the real and imaginary parts of the function are solutions of Laplace's equation.

$$f(z) = E(x,y) + iF(x,y)$$

where

$$E(x,y) = Ax^2 - Ay^2 + Bx$$

and

$$F(x,y) = 2Axy + By$$

It is evident that

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} = 2A - 2A = 0$$

and that

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} = 0 - 0 = 0$$

Definition of Analytic Functions. The properties of functions of the complex variable will now be considered. It was mentioned above that a large group of functions had the desired properties, and such functions were referred to as "analytic functions." It will be remembered that, in the study of functions of the real variable, attention is usually restricted to functions which are continuous and functions of which the derivative at any point is independent of the direction in which we approach the point as we take the limit of the ratio of the increment of the function to the increment of the variable. Similarly in studying functions of the complex variable we shall restrict attention to functions having a derivative that is independent of the direction of approach to the point in question. This is necessary because only functions having this property also have the desired properties of potential functions. Mathematicians use the term *analytic* to describe such functions.

Consider

$$W = f(z) = f(x + iy) \tag{5.68}$$

Such a function is said to be *analytic* if it has a derivative that is independent of the direction of the increment of the variable Δz as it approaches zero.

For a single real variable

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} \tag{5.69}$$

For the complex variable $z = x + iy$

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\Delta z \to 0} \frac{\Delta f}{\Delta z} \tag{5.70}$$

Let

$$W = u + iv = f(x + iy)$$

If the function is analytic,

$$\lim_{\substack{\Delta z \to 0 \\ \Delta z = \Delta x + i\,\Delta y}} \frac{\Delta W}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta z = \Delta x + i0}} \frac{\Delta W}{\Delta x} = \lim_{\substack{\Delta y \to 0 \\ \Delta z = 0 + i\,\Delta y}} \frac{\Delta W}{i\,\Delta y} \tag{5.71}$$

In derivative form

$$\frac{dW}{dz} = \frac{\partial W}{\partial x} = -i\,\frac{\partial W}{\partial y} \tag{5.72}$$

But

$$W = u + iv$$

Therefore

$$\frac{\partial W}{\partial x} = \frac{\partial u}{\partial x} + i\,\frac{\partial v}{\partial x} \tag{5.73}$$

and

$$-i\,\frac{\partial W}{\partial y} = -i\left(\frac{\partial u}{\partial y} + i\,\frac{\partial v}{\partial y}\right) \tag{5.74}$$

$$-i\,\frac{\partial W}{\partial y} = \frac{\partial v}{\partial y} - i\,\frac{\partial u}{\partial y} \tag{5.75}$$

Hence

$$\frac{\partial u}{\partial x} + i\,\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\,\frac{\partial u}{\partial y} \tag{5.76}$$

Equating real and imaginary parts,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{5.77}$$

$$\frac{\partial v}{\partial x} = -\,\frac{\partial u}{\partial y} \tag{5.78}$$

These equations are known as the "Cauchy-Riemann conditions" and serve to identify analytic functions.

Dividing Eq. (5.77) By Eq. (5.78),

$$\frac{\dfrac{\partial u}{\partial x}}{\dfrac{\partial u}{\partial y}} = -\frac{\dfrac{\partial v}{\partial y}}{\dfrac{\partial v}{\partial x}} \tag{5.79}$$

which is the orthogonality condition for two functions since the derivative of a function of x and y is

$$\frac{dy}{dx} = -\frac{\dfrac{\partial f}{\partial x}}{\dfrac{\partial f}{\partial y}} \tag{5.80}$$

and curves are perpendicular if the derivative of one curve is the negative reciprocal of the derivative of the other. If we take derivatives of the Cauchy-Riemann equations with respect to x and y, respectively, then

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x\,\partial y} \tag{5.81}$$

$$\frac{\partial^2 v}{\partial x\,\partial y} = -\frac{\partial^2 u}{\partial y^2} \tag{5.82}$$

Subtracting these gives

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{5.83}$$

or Laplace's equation holds for the real part of the function. Similarly, Laplace's equation holds for the imaginary part.

To summarize, an analytic function is one whose derivative is independent of the direction of the increment of the variable as the increment approaches zero. For such a function the Cauchy-Riemann conditions hold. Analytic functions have real and imaginary parts which are orthogonal to each other and each one of which is a solution of Laplace's equation.

It will be recognized that functions may be analytic except at certain points just as functions of a real variable may be continuous except at certain points. Such points are frequently those at which the function has a pole, *i.e.*, assumes an infinite value. It is possible to use such functions if the regions in which the function is not analytic are excluded from consideration.

A serious limitation of the method of conformal transformations is that it is not always possible to find the transformation which will convert one set of electrodes to another. In general, there is no definite method by which the transformation which fits a set of electrodes can be found. An exception to this remark is the Schwartz-Christoffel transformation, which transforms the real axis in the W plane into any polygon in the Z plane, but this transformation does not find much use in the field of electronics. However, it is also possible to use methods of successive approximations and series expansions. *Fortunately, the transformations necessary for the most important vacuum-tube problems are known.*

The Logarithmic Transformation. The transformation that solves the problem of determining fields in the plane-electrode triode is known. It is the logarithmic transformation

$$W = \ln_\epsilon Z \tag{5.84}$$

where $\epsilon = 2.718$ is the Napierian base.
This is analytic for all finite values of x and y other than zero.

The nature of the logarithmic transformation can best be understood by studying its component relations. It is most convenient to use polar coordinates in the Z plane and rectangular coordinates in the W plane. Thus let

$$Z = r\epsilon^{i\theta} \tag{5.85}$$

and

$$W = u + iv \tag{5.86}$$

In these coordinates

$$u + iv = \ln_\epsilon r + i\theta \tag{5.87}$$

so that the component equations relating the real and imaginary parts are

$$u = \ln_\epsilon r \tag{5.88}$$

$$v = \theta \tag{5.89}$$

or, solved for r and θ,

$$r = \epsilon^u \tag{5.90}$$

$$\theta = v \tag{5.91}$$

This function is readily proved to be analytic for finite values of the argument by application of the Cauchy-Riemann equations.

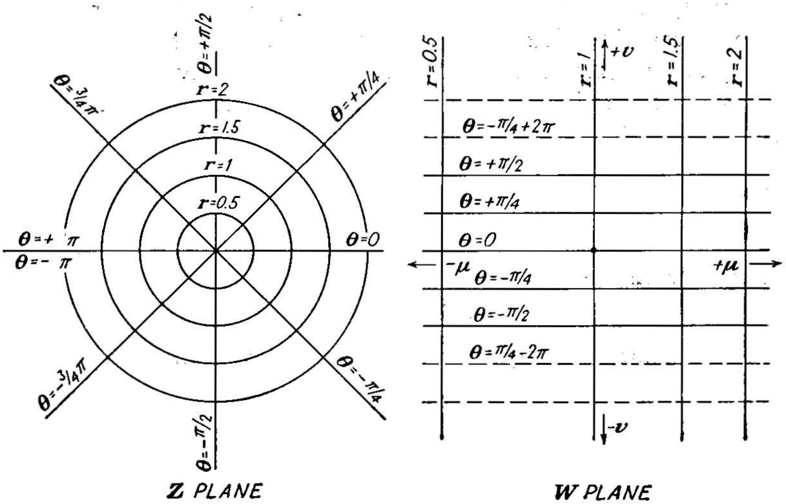
Examination of the v component of W shows that it is multiple-valued, in fact infinitely so. This occurs because any angle in the Z plane can be written as an angle less than 2π plus any integral multiple of 2π. The angle θ can be written as $\theta + 2\pi n$, where n is any positive or negative integer. Thus, corresponding to any point in the Z plane there are an infinite number of points in the W plane evenly spaced by a distance 2π along a vertical line.

From Eq. (5.88) it is seen that any circle about the origin in the Z plane, $r = k$, transforms into a line parallel to the v axis in the W plane, $u = \ln k$.* Circles with radii less than 1 give lines in the left half of the

---

* Hereafter, the notation $\ln r$ will be used to designate the natural logarithm of r.

W plane since the logarithms of numbers less than 1 are negative, and circles with radii greater than 1 give lines in the right half of the *W* plane since the logarithms of numbers greater than 1 are positive. Any radial line through the origin, $\theta = k$, transforms into a set of lines in the *W* plane parallel to the *u* axis and spaced a distance 2π, $v + 2\pi n = k$. These relations are shown in Fig. 5.17. From this it is seen that a single point in the *Z* plane such as $r = 1.5$, $\theta = \frac{\pi}{4}$ transforms into a series of points $u = \ln 1.5$, $v = \frac{\pi}{4} + 2n\pi$ in the *W* plane. Thus a single point in the *Z* plane that may be taken as representing a line charge transforms into a row of line charges evenly spaced in the *W* plane. This gives the arrangement corresponding to a grid of evenly spaced parallel wires and is the basis for the plane-electrode representation.



Z PLANE W PLANE

FIG. 5.17.—The logarithmic transformation, $W = \ln_\epsilon Z$.

The nature of the logarithmic transformation is better understood if the transformation be effected in a series of steps. Imagine the *Z*-coordinate plane to be a stretched elastic membrane. If the polar-coordinate diagram of the *Z* plane shown in Fig. 5.18*a* be split along the negative *x* axis and the upper and lower edges be rotated clockwise and counterclockwise, respectively, the pie-shaped section of Fig. 5.18*b* will result. If now the point on the pie is stretched to the left and the outer edge is compressed, the configuration shown in Fig. 5.18*c* results. Finally the left and right edges are made the same length and are stretched to negative and positive infinity, respectively, to give a strip of the *W* plane as shown in Fig. 5.18*d*.

The nature of the logarithmic transformation is also well illustrated by relationship between certain types of maps. Thus, if the Northern Hemisphere on a polar azimuthal equidistant projection be taken as the Z plane, then the Northern Hemisphere on a Mercator projection corresponds very closely to the $W = \ln_\epsilon Z$ plane. It will be recognized that each of these two common maps is but a different representation of a part of the earth's surface. In Fig. 5.19 is shown a polar azimuthal projection of the Northern Hemisphere. In Fig. 5.20 is shown a Mercator projection of the Northern Hemisphere.

The polar azimuthal equidistant projection is made by unfolding the earth's surface and stretching it out until it is a plane tangent to the pole with distances from the pole made equal to the great-circle distances on the actual sphere. This is indicated in Fig. 5.21. The longitude lines become straight lines through the pole, and the latitude circles remain circles.

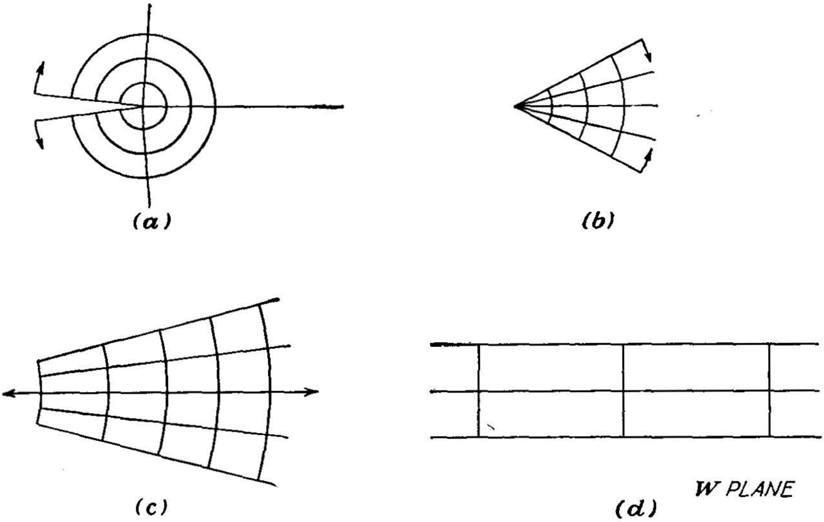


FIG. 5.18.—Steps in a progressive transformation from the Z to the $\ln_\epsilon Z$ plane.

Mercator's projection is approximated by surrounding the earth with a circular cylinder tangent to the earth at the equator as in Fig. 5.22 and extending to infinity in both directions. Points on the earth's surface are then projected onto this cylinder by drawing a line from the earth's center through the point in question and extending it until it hits the cylinder. The cylinder is then cut and unfolded to give a plane surface. The latitude circles on the sphere become a series of parallel straight lines on the Mercator projection. The longitude circles become another set of equidistant parallel straight lines perpendicular to the latitude lines.

It is easily seen that the latitude circles, $r = k$, in the polar azimuthal equidistant projection become straight lines parallel to the equator, $u = \ln k$, in the Mercator projection. The longitude lines through the pole in the polar azimuthal equidistant projection, $\theta = K$, become a set of evenly spaced lines perpendicular to the equator in the Mercator projection, $v = K$. The pole, which is the center of things in the polar azimuthal equidistant projection, recedes to infinity in the Mercator projection. Distortions in the different representations are evident. The polar azimuthal equidistant projection gives its most accurate representation near the pole but stretches out the equator disproportionately, causing Africa to be too wide. The Mercator projection is most accurate in the band around the equator but causes areas near the poles to be disproportionately large. Greenland on a Mercator projection looks larger than South America but is actually only one-tenth as large.

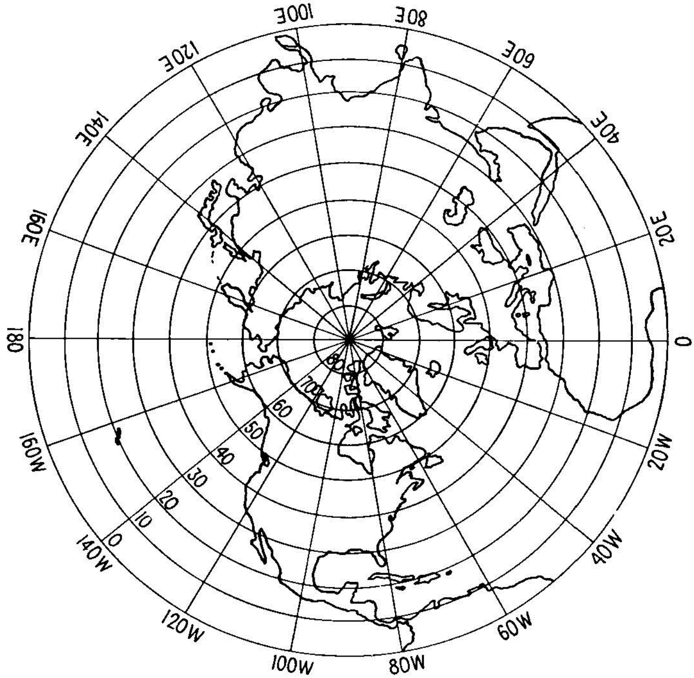


FIG. 5.19.—The polar azimuthal equidistant projection of the Northern Hemisphere. This may be considered as a polar-coordinate representation of the Z plane.

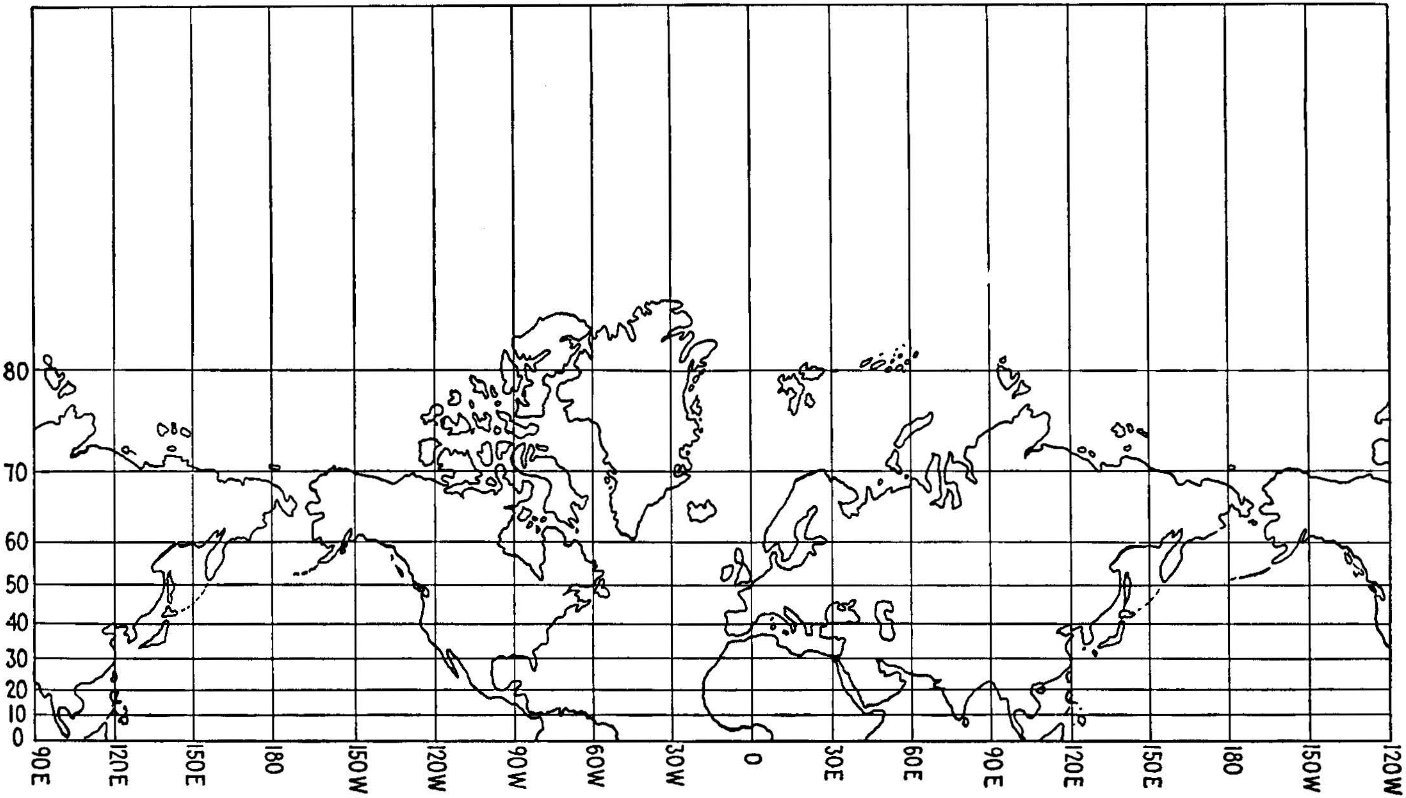


FIG. 5.20.—The Mercator projection of the Northern Hemisphere. This map results from applying the transformation $W = \ln_e Z$ to the polar azimuthal equidistant projection of Fig. 5.19.

Observe, however, that angles and the similarity of small areas are preserved.

The Function $W = Z^{1/n}$. The simple power function given by raising Z to some rational fractional power is the function that gives the fields inside of a cylindrical triode. As usually written, this function is

$$W = Z^{1/n} \tag{5.92}$$

It may also be written

$$W = \epsilon^{\frac{\ln Z}{n}} \tag{5.93}$$

or

$$\ln W = \frac{1}{n} \ln Z \tag{5.94}$$

but the form of Eq. (5.92) is preferred.

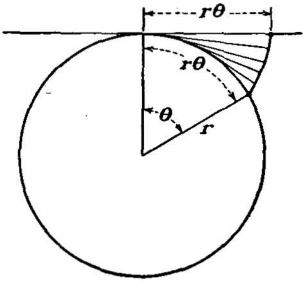


FIG. 5.21.—Construction of the polar azimuthal equidistant projection.

The nature of the power function may best be understood by examining the form of the function for a specific value of n. Consider the case of $n = 2$. Then

$$W = Z^{\frac{1}{2}} \tag{5.95}$$

or

$$Z = W^2 \tag{5.96}$$

Using rectangular components for both Z and W,

$$Z = x + iy \tag{5.97}$$

and

$$W = u + iv \tag{5.98}$$

Making these substitutions,

$$x + iy = u^2 + i2uv - v^2 \tag{5.99}$$

from which, by equating real and imaginary parts, the component equations are

$$x = u^2 - v^2 \tag{5.100}$$

and

$$y = 2uv \tag{5.101}$$

These component equations satisfy the Cauchy-Riemann conditions since $\frac{\partial x}{\partial u} = \frac{\partial y}{\partial v} = 2u$ and $\frac{\partial x}{\partial v} = -\frac{\partial y}{\partial u} = -2v$. Letting x and y assume various constant values, it is seen that the component equations (5.100) and (5.101) represent two families of orthogonal hyperbolas previously shown in Fig. 5.16. For a better comparison the Z and W planes are shown in Fig. 5.23, in which corresponding flux and potential lines are indicated. It will be seen from this figure that the upper half of the Z plane transforms into the first quadrant of the W plane. The transformation is double-valued, *i.e.*, one point in the Z plane transforms into two points in the W plane. For example, the point (0,4) in the Z plane transforms into the point (1.414,1.414) and also the point $(-1.414, -1.414)$ in the W plane. For most purposes only the first, or "principal," value of the multiple values is used, though all of them have the correct mathematical properties. It can further be seen that if the polar representation of points is used the angle of the point in the W plane is half the angle of the corresponding point in the Z plane and the radius vector of a point in the W plane is the square root of the radius vector of the corresponding point in the Z plane.

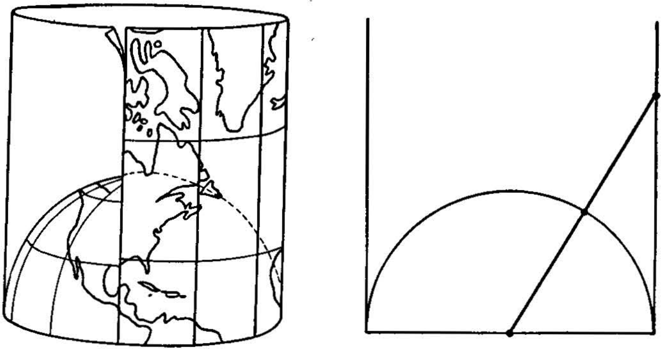
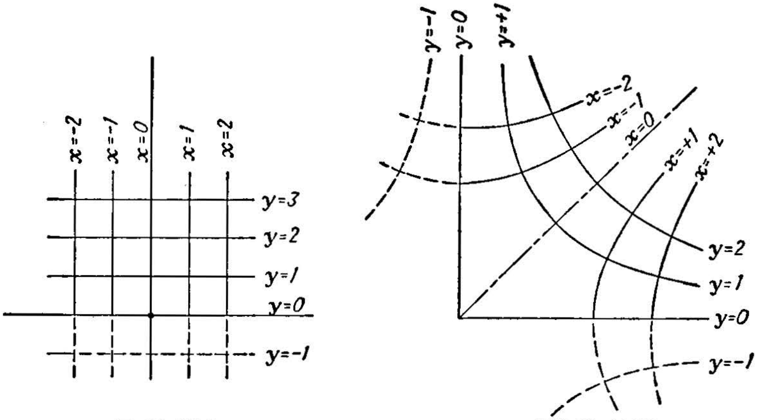


FIG. 5.22.—Construction of the Mercator projection.



Z PLANE W PLANE

FIG. 5.23.—The transformation $W = Z^{1/2}$.

In the general case of the function $W = Z^{1/n}$, the function is n-valued if n is an integer. As a result, the upper half of the Z plane transforms into a segment of the W plane having an angle $\frac{\pi}{n}$. Angles in the W plane are $\frac{1}{n}$th the corresponding values in the Z plane (principal values), and radius vectors have a magnitude in the W plane that is the nth root of the radius vector of the corresponding points in the Z plane.

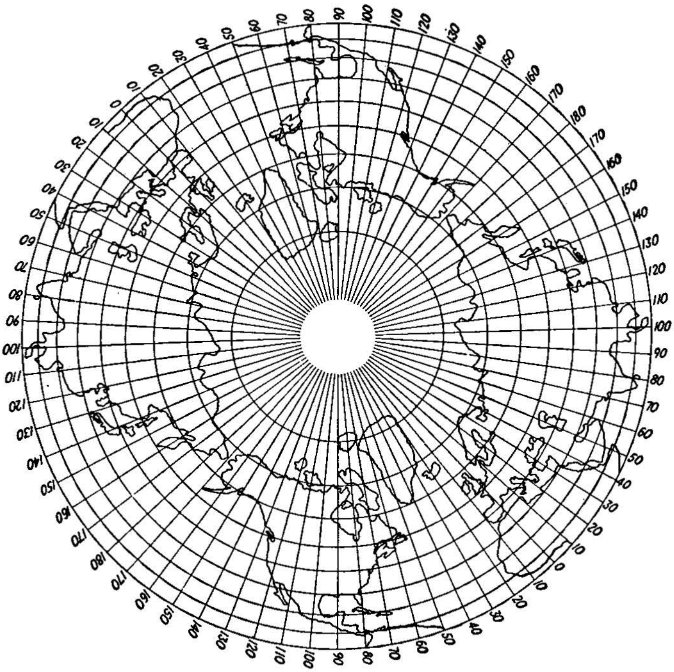


FIG. 5.24—Polar azimuthal equidistant projection of the Northern Hemisphere transformed by $W = Z^{1/2}$.

The component relations are not readily written in rectangular components for any general integral value of n. In polar form, however, they are quite simple. Let

$$Z = r\angle\theta = r\epsilon^{i\theta} \tag{5.85}$$

as before; and let

$$W = R\angle\phi = R\epsilon^{i\phi} \tag{5.102}$$

Then the component equations in polar form are

$$R = r^{1/n} \tag{5.103}$$

and

$$\phi = \frac{\theta}{n} \tag{5.104}$$

The nature of the transformation $W = Z^{1/n}$ may be indicated as a kind of deformation of the Z plane. Upon comparing the W- and Z-plane representations in Fig. 5.23, it is seen that if the upper half of the Z plane be cut along the negative x axis and if the upper edge of the negative x axis be swung clockwise 90 deg and the lower half of the negative x axis be swung counterclockwise 90 deg then the W-plane representation will result if the intermediate regions are allowed to deform accordingly. A set of polar maps can also be drawn to illustrate the nature of the transformation. In Fig. 5.24 are shown maps illustrating the nature of the transformation $W = Z^{1/2}$. It is seen that the representation is double-valued and that the scale of distances from the pole in the W-plane representation is quadratic rather than linear; the land areas are pushed out from the pole toward the equator though the map as a whole differs surprisingly little from the usual representation.

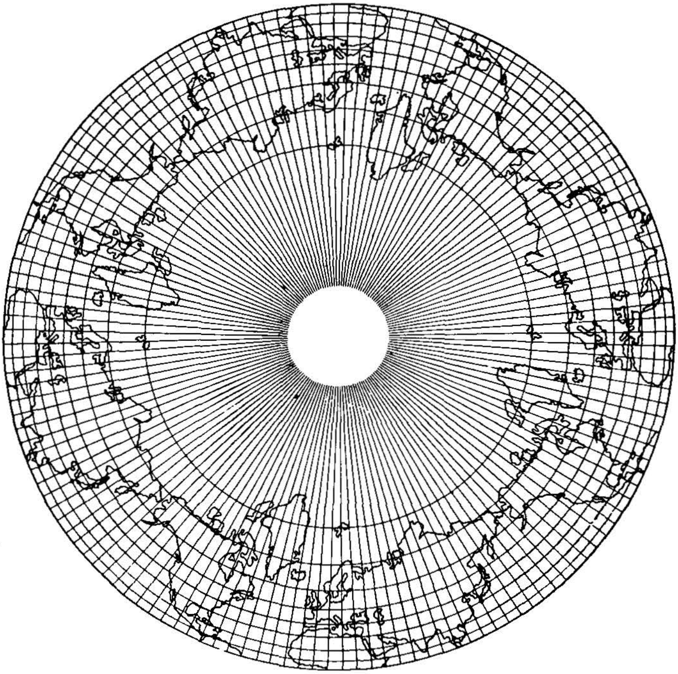


Fig. 5.25—Polar azimuthal equidistant projection of the Northern Hemisphere transformed by $W = Z^{1/4}$.

In Fig. 5.25 are shown maps illustrating the nature of the transformation $W = Z^{1/4}$. This transformation is quadruple-valued, *i.e.*, every point in the Northern Hemisphere is repeated four times in the W-plane representation. The scale of distances from the pole is quartic in the W plane, with the result that the land masses are compressed strongly near the equator.

An inkling of how this transformation is used is obtained if we consider that in the polar azimuthal equidistant or Z-plane representation a cathode wire be located at the North Pole, a grid wire be located at Iceland, and the equator be a circular plate surrounding both. Then we have a simple tube structure with one cathode wire, one grid wire, and one plate. If then the transformation $W = Z^{1/4}$ be used, the corresponding W-plane representation has one cathode wire at the pole as before, a surrounding plate at the equator as before, but *four* grid wires located at the four Icelands, which are evenly spaced around the 66° latitude circle. If the potential field can be found in the Z plane, then it can be transformed into the W plane just as the land outlines have been transformed. This is what Chap. 7 is mostly about.

CHAPTER 6

LAWS OF ELECTRON MOTION

ALL electronic devices depend for their action upon the effect of applied electric or magnetic fields upon electron flow within the device. The applied fields may control the direction or the magnitude of the current flow or both. In this chapter there will be studied the effect of fields upon the electron paths when the electrons are present in small enough number so that their presence does not change the applied fields. In a subsequent chapter there will be studied the effect of fields upon electron flow when the electrons are present in sufficiently large numbers to influence the fields.

6.1. Electron in a Uniform Electric Field. An electron in a uniform electrostatic field experiences a constant force in the direction of increasing potential. As a result, the laws governing an electron starting from rest are the same as those which apply to a body falling freely under the influence of gravity until very high velocities are reached. From Newton's second law,

$$m \frac{d^2x}{dt^2} = -Ee \qquad (6.1)$$

where m is mass of the electron, 9.107×10^{-31} kg

x is distance, meters

t is time, sec

$-E = \frac{dV}{dx}$ is gradient of potential, volts per meter

e is *magnitude* of the charge of the electron,

1.602×10^{-19} coulomb

A first integration of Eq. (6.1) gives

$$v = \frac{dx}{dt} = -\frac{e}{m} Et \qquad \text{meters per sec} \qquad (6.2)$$

the constant being zero because the velocity is taken as zero when time is zero. A second integration gives

$$x = -\frac{1}{2}\frac{e}{m} Et^2 \qquad \text{meters} \qquad (6.3)$$

in which the constant is again zero for an electron starting from rest at a point of zero potential.

If time is eliminated between Eqs. (6.2) and (6.3), there results

$$\tfrac{1}{2}mv^2 = eV \tag{6.4a}$$

where $V = -Ex$ is the potential through which the electron has fallen. Equation (6.4*a*) states that the kinetic energy acquired by an electron starting from rest is equal to the potential energy which it has lost. Solving for v,

$$v = \sqrt{\frac{-2Exe}{m}} = \sqrt{\frac{2eV}{m}} \qquad \text{meters per sec} \tag{6.4b}$$

The relation between Eqs. (6.2), (6.3), and (6.4*a*) and the corresponding equations for a falling body is evident. It is seen that the quantity $-\dfrac{eE}{m}$ is the equivalent of the gravitational constant.

If the values for charge and mass be substituted and all quantities be expressed in practical units, then

$$v = 1.758 \times 10^{11}Et \qquad \text{meters per sec} \tag{6.5}$$
$$x = 0.879 \times 10^{11}Et^2 \qquad \text{meters} \tag{6.6}$$
$$v = 5.93 \times 10^5 \sqrt{V} \qquad \text{meters per sec} \tag{6.7a}$$

where v is velocity, meters per sec
E is gradient, volts per meter
V is potential, volts
x is distance through which the electron has been accelerated

The above expressions are not accurate for potentials exceeding 30,000 volts.

The ratio of the charge to the mass of the electron is so high that a small voltage will impart a tremendous velocity to the electron. It takes only three-tenths of a microvolt to give an electron a velocity of 700 mph which is approximately the velocity of sound. Although the speeds of electrons are very high, their energy is low because of their minute mass.

Electron speeds are frequently expressed in terms of the corresponding voltage. Energies are also designated in terms of electron volts, 1 electron volt being equal to 1.602×10^{-19} watt-sec. An electron that has "fallen" through 1 volt of potential is said to have acquired an energy of 1 electron volt.

If an electron enters a region of uniform field at a point x_0 with an initial velocity v_0 parallel to the field, then

$$v = -\frac{e}{m} Et + v_0 \qquad \text{meters per sec} \tag{6.8}$$

$$x = -\frac{e}{2m} Et^2 + v_0 t + x_0 \qquad \text{meters} \tag{6.9}$$

where x_0 is the initial distance in meters.

Eliminating t between Eqs. (6.8) and (6.9),

$$\tfrac{1}{2}m(v^2 - v_0^2) = e(V - V_0) \tag{6.10a}$$

or

$$v = 5.93 \times 10^5 \sqrt{V - V_0} \qquad \text{meters per sec} \tag{6.10b}$$

Equations (6.4), (6.7), and (6.10*b*), which give velocity in terms of potential, are not restricted to uniform fields or to one-dimensional fields. This is due to the fact that these equations express the conservation of energy and hence are independent of the electron path and the nature of the potential field.

6.2. Initial Velocity Not Parallel to Field. When an electron enters a region of a uniform field with an initial velocity that is at an angle with the gradient of potential, the electron follows a parabolic trajectory. This is because it experiences a constant force in the direction of the gradient and no force at right angles to this. The case is analogous to the mechanical case of a projectile fired from a gun in the absence of friction. The projectile is subjected to a constant downward force but has no force affecting the component of velocity parallel to the earth's surface.

The differential equations for the components of electron motion when the electron meets a retarding component of field are

$$\frac{d^2y}{dt^2} = -\frac{e}{m} E \tag{6.11}$$

and

$$\frac{d^2x}{dt^2} = 0 \tag{6.12}$$

The initial conditions that determine the solution of these equations are as follows:

When $t = 0$,

$$\frac{dy}{dt} = v_0 \cos \theta \qquad y = 0$$

$$\frac{dx}{dt} = v_0 \sin \theta \qquad x = 0$$

where θ is the angle that the initial velocity makes with the gradient of potential and v_0 is the initial velocity.

A first integration gives

$$\frac{dy}{dt} = -\frac{e}{m}Et + v_0 \cos\theta \tag{6.13}$$

and

$$\frac{dx}{dt} = v_0 \sin\theta \tag{6.14}$$

A second integration gives

$$y = -\frac{e}{2m}Et^2 + v_0 \cos\theta\, t \tag{6.15}$$

and

$$x = v_0 \sin\theta\, t \tag{6.16}$$

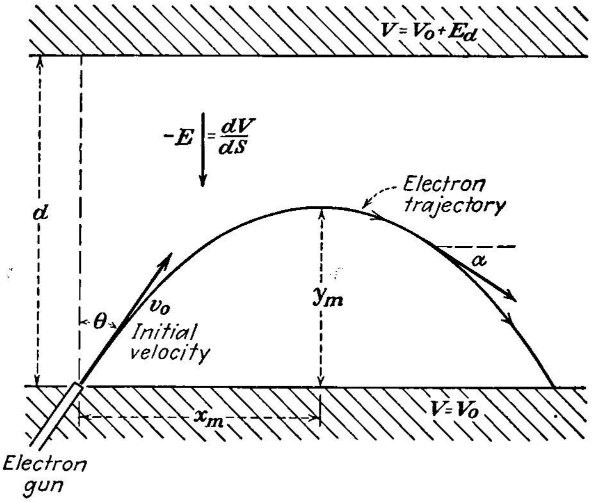


FIG. 6.1.—Parabolic electron trajectory in a uniform electric field. This case results from the injection of an electron with an initial velocity into a region where the electric field has a uniform retarding action.

Elimination of the time factor between Eqs. (6.15) and (6.16) gives the equation of the parabolic trajectory

$$y = \frac{-Ex^2}{4V_0 \sin^2\theta} + \frac{x}{\tan\theta} \tag{6.17}$$

where V_0 is the potential corresponding to the initial velocity. This is observed to be the equation of a parabola in x and y and to be independent of the system of electrical units used. The notation used in all the above equations corresponds to that shown in Fig. 6.1.

The slope of the trajectory at any point is given by

$$\tan \alpha = \frac{-Ex}{2V_0 \sin^2 \theta} + \frac{1}{\tan \theta} \tag{6.18}$$

where α is the angle that the tangent to the parabola makes with the horizontal axis.

The maximum height to which the electron rises is

$$y_m = \frac{V_0 \cos^2 \theta}{E} \tag{6.19}$$

and the horizontal displacement corresponding to this is

$$x_m = \frac{2V_0 \sin \theta \cos \theta}{E} \tag{6.20}$$

6.3. Electrostatic Deflection of Cathode-ray Beams. An application of the situation analyzed in the last section is found in the deflecting plates of the ordinary cathode-ray tube. Here a stream of electrons enters a region of a uniform field, traverses a parabolic path while under the influence of this field, and leaves the region between the plates at a different angle from that at which it enters. It then travels in a straight line until it hits the fluorescent screen.

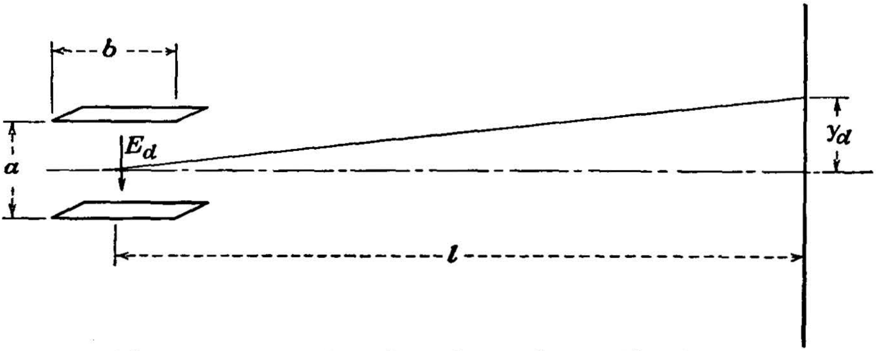


FIG. 6.2.—Electrostatic deflection of an electron beam. The electron enters the region of deflecting field at right angles to the field. The trajectory is parabolic between the plates.

In this case, as may be seen by reference to Fig. 6.2, the electron enters the deflecting field at right angles, making the angle θ equal to 90 deg. The potential gradient is V_d/a. For this condition the slope of the trajectory upon emerging from the plates after a distance of travel b in the horizontal direction is, by Eq. (6.18)

$$\tan \alpha = \frac{V_d b}{2aV_0} + \frac{1}{\infty} \tag{6.21}$$

where V_d is the potential between plates and V_0 is the potential corresponding to the initial velocity. But

$$\tan \alpha \cong \frac{y_d}{l} \tag{6.22}$$

so that

$$y_d = \frac{l\, b V_d}{2 a V_0} \tag{6.23}$$

This expression is only an approximate one, for it neglects the fringing effect of the flux lines around the end of the deflecting plates.

In most cathode-ray tubes the deflecting plates are not parallel but slope apart so that the electron in passing between them is subjected to a constantly decreasing gradient. When this is the case, the expressions obtained previously cannot be used and the problem must be solved anew. This is readily done by setting the gradient between the plates equal to

$$\frac{dV}{dy} = \frac{V_d}{a_1 + \dfrac{(a_2 - a_1)}{b}\, x} \tag{6.24}$$

where a_1 and a_2 are the separations of the ends of the deflecting plates where the beam enters and leaves, respectively. Other symbols have their previous significance. The expression for the crosswise acceleration involving this factor is then integrated to obtain the crosswise component of velocity at the point where the beam emerges from between the deflecting plates. The ratio of the crosswise to the axial velocity multiplied by the distance to the fluorescent screen is then equal to the screen displacement. This has the form

$$y = \frac{blV_d}{2V_0 a_1} \frac{ln \left(\dfrac{a_2}{a_1}\right)}{\left(\dfrac{a_2}{a_1} - 1\right)} \tag{6.25}$$

which reduces to Eq. (6.23) when $a_1 = a_2$. From this it is seen that the effect of spreading the deflecting plates at one end is to decrease the deflection. If the separation of the plates is increased 50 per cent at the far end, the deflection is decreased to 81.2 per cent of its value for the parallel plates having the near-end spacing. The deflection for divergent plates is, however, slightly greater than for parallel plates having their spacing equal to the average spacing of the divergent plates. Equation (6.25) is still in error because it takes no account of the flux fringing at the ends of the deflecting plates.[1,2] The effect of the fringing is to increase the effective length of the plates.

6.4. Relativity Correction for Velocity. The general expression developed in Sec. 6.1 giving electron velocity as proportional to the square root of potential is valid only for velocities low compared with the velocity of light. This is due to the fact that according to the theory of relativity the mass of a particle changes with its velocity, and in the derivation of the expressions of Sec. 6.1 the mass was assumed constant.

One of the postulates of the theory of relativity is that nothing can move with a speed greater than the velocity of light. As a consequence of this upper limit on velocity, it is seen that a body subjected to a constant force must have its mass increase as it is accelerated, or otherwise its velocity would increase indefinitely and finally violate the postulate by exceeding the velocity of light. If, however, the mass of the particle increases as its velocity increases, a constant force produces an acceleration that decreases with velocity and permits the possibility of an upper limit to velocity.

Another conclusion of the theory of relativity is that matter and energy are equivalent. Mass may be considered a manifestation of energy. To relate this to the remarks of the previous paragraph, the energy expended in accelerating an electron manifests itself as an increase in its mass. From this idea, the law for the change of mass with velocity and the corresponding law for velocity in terms of potential are readily deduced.

Mass and energy are related by a factor c^2, where c is the velocity of light.

$$w = c^2m \tag{6.26}$$

where w is energy in watt-seconds, c is the velocity of light, 3×10^8 meters per sec, and m is mass in kilograms.

Consider the increase in mass that an electron experiences as it is accelerated. Then

$$c^2\,dm = dw = F\,ds \tag{6.27}$$

where dm is the increase in mass, dw is the energy expended in accelerating the particle, F is the applied force, and s is the distance factor.

According to Newton's second law,

$$F = \frac{d}{dt}(mv) \tag{6.28}$$

---

[1] See also BENHAM, W., Inclined Deflecting Plates, *Wireless Eng.*, vol. 13 (No. 148) pp. 10–13, 1936.

[2] HINTERBERGER, O., Correction for End Effects in Oscilloscope Deflecting Plates, *Zeit. für Phys.*, vol. 105, pp. 501–512, July, August, 1937.

this being the general statement of the law when a variation in mass is encountered. Making this substitution into Eq. (6.27), and integrating

$$c^2 \int dm = \int \frac{d(mv)}{dt}\, ds = \int v\, d(mv) \tag{6.29}$$

Equating the integrands and separating variables,

$$\frac{dm}{m} = \frac{v\, dv}{c^2 - v^2} \tag{6.30}$$

which integrates to give

$$[\ln m]_{m_0}^{m} = \left[\ln \frac{1}{(c^2 - v^2)^{1/2}}\right]_0^{v} \tag{6.31}$$

giving the result sought,

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{6.32}$$

where m_0 is the rest mass of the electron.[1] It is seen that at low velocities the mass is practically the rest mass. As the velocity increases, the mass increases, slowly at first and then quite rapidly. At one-tenth the velocity of light (2,600 volts) the mass has only increased by ½ of 1 per cent. The mass tends to become infinite as the velocity of light is approached.

The expression for mass as determined by the velocity can now be applied to obtain an expression for velocity as a function of potential. This is best done by equating the expressions for potential and kinetic energy, the latter involving the general expression for the mass as a function of the velocity.

$$\text{Potential energy, } Ve = \text{kinetic energy, } -e \int E\, ds \tag{6.33}$$

But

$$\int E\, ds = \int \frac{E\, ds}{dt}\, dt = \int Ev\, dt \tag{6.34}$$

---

[1] This is what is known as the "transverse mass" of the electron because it is the effective mass of the electron to transverse deflection where the magnitude of the velocity is not changed appreciably. It should be distinguished from the "longitudinal mass," which has the value $\dfrac{m_0}{\left(1 - \dfrac{v^2}{c^2}\right)^{3/2}}$, which is the effective mass that an electron presents to longitudinal acceleration where the mass as well as the velocity changes.

In this expression

$$-eE = \frac{d}{dt}(mv) = m\frac{dv}{dt} + v\frac{dm}{dt} \tag{6.35}$$

Hence

$$Ve = \int m\frac{dv}{dt}v\,dt + \int v\frac{dm}{dt}v\,dt \tag{6.36}$$

$$Ve = \int mv\,dv + \int v^2\,dm \tag{6.37}$$

If now the general expression for m as a function of velocity be substituted and the integrals be evaluated between the limits of 0 and v on the variable v, there results

$$Ve = m_0c^2\left[\frac{1}{\left(1 - \frac{v^2}{c^2}\right)^{1/2}} - 1\right] \tag{6.38}$$

This is readily solved for velocity.

$$v = c\sqrt{1 - \frac{1}{(1 + 1.965 \times 10^{-6}V)^2}} \tag{6.39}$$

The corresponding expression for mass as a function of potential is

$$m = m_0(1 + 1.965 \times 10^{-6}V) \tag{6.40a}$$

The results of the above analysis deserve considerable study. Consider first the way in which the mass varies. Referring to Eq. (6.32), it is seen that at very small velocities the mass is practically the rest mass. As the velocity is increased, the mass at first increases parabolically with the velocity,

$$m = m_0\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) \tag{6.40b}$$

This expression is approximately correct until the velocity reaches one-tenth the velocity of light. At this velocity the mass has increased only ½ of 1 per cent.

From Eq. (6.40a) it is seen that the mass increases linearly with the potential. This happens because of the energy relation, which requires that the potential energy acquired manifest itself as an increase in mass. At about 500,000 volts the mass of the electron has doubled. This voltage is not ordinarily reached in ordinary tubes. At 5,000 volts the mass has increased by 1 per cent.

The velocity of the electron follows the low-voltage law of Eq. (6.7) until very large voltages are reached. Even at 100,000 volts the velocity

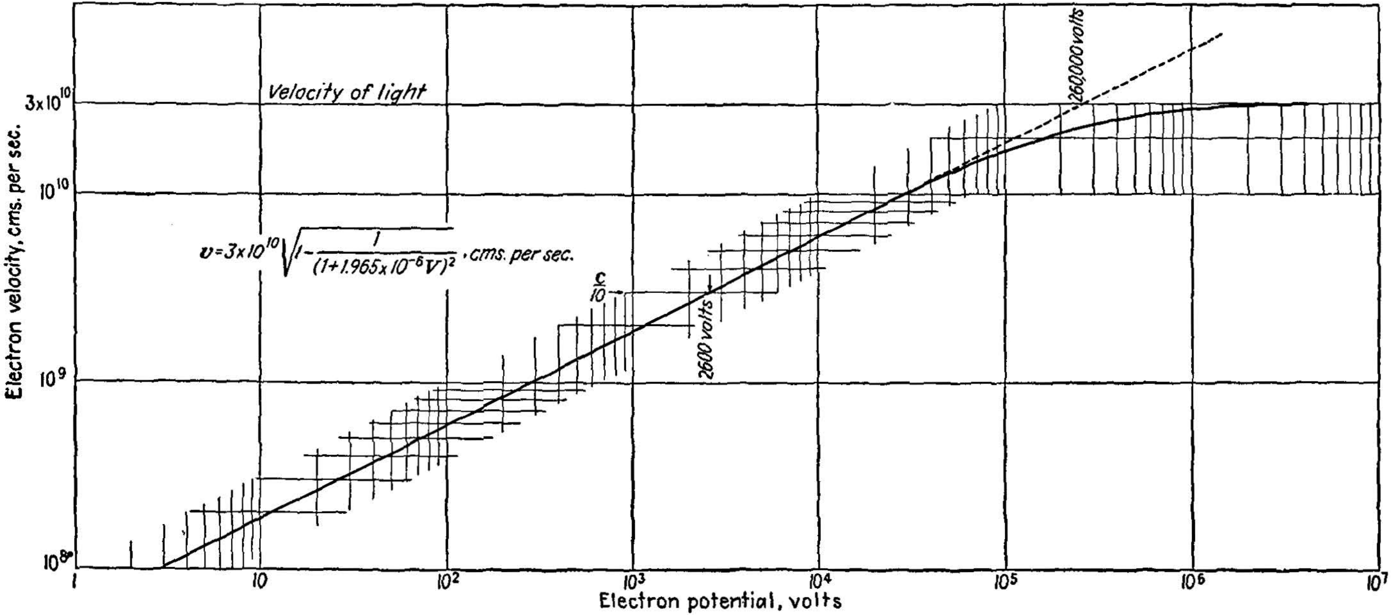


FIG. 6.3.—Electron velocity as a function of accelerating potential. For voltages up to about 30,000 volts, the velocity is proportional to the square root of potential. Above this voltage relativity considerations apply, with the result that the velocity increases less rapidly and assumes a limiting value equal to the velocity of light. At 3,000,000 volts the electron velocity is 99 per cent of the velocity of light.

has dropped only 7½ per cent from the low-voltage value given by Eq. (6.7*a*), which may be written

$$\frac{v}{c} = \frac{\sqrt{V}}{506} \tag{6.7b}$$

By 1,000,000 volts, however, the velocity has reached 93 per cent of the velocity of light. Above 1,000,000 volts the velocity becomes closer and closer to the velocity of light but experiences no rapid change with voltage.

The above relations are shown in Fig. 6.3. On this log-log plot it is seen that velocity follows the half-power law of potential well up to about 100,000 volts. Between 100,000 and 1,000,000 volts the change from the half-power law occurs, and above 1,000,000 volts the velocity is practically constant. Several convenient reference points may be taken from this curve. An electron reaches one-tenth the velocity of light at about 2,600 volts. If there were no change of mass with velocity, the electron would reach the velocity of light at about 260,000 volts.

6.5. Two-dimensional Electric Fields. Electrons are frequently exposed to fields that are not uniform but that are two-dimensional or more. It is generally quite difficult to determine exactly what the electron path is by analytical methods.

The fundamental differential equations involved are quite simple, but they are usually difficult if not impossible to solve. In rectangular coordinates the differential equations are

$$\frac{d^2x}{dt^2} = -\frac{e}{m} E_x(x,y) \tag{6.41}$$

and

$$\frac{d^2y}{dt^2} = -\frac{e}{m} E_y(x,y) \tag{6.42}$$

where

$$E_x = -\frac{\partial V\,(x,y)}{\partial x} \tag{6.43}$$

and

$$E_y = -\frac{\partial V\,(x,y)}{\partial y} \tag{6.44}$$

When these equations can be solved, they give the components of electron displacement parametrically in terms of t.

When the potentials are given in two-dimensional circular-cylinder coordinates with an axial symmetry, as is the case in most electron-optical problems, the equations have the same form as those above. It is necessary only to substitute r for x and z for y to get the corresponding equations for this case.

For two-dimensional polar coordinates, such as are used for the cylindrical triode, the equations are quite different and still more difficult to solve. In terms of a radial variable r and an angular variable θ the differential equations of motion are

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = \frac{e}{m}\frac{\partial V}{\partial r} \tag{6.45}$$

$$r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt} = \frac{1}{r}\frac{e}{m}\frac{\partial V}{\partial \theta} \tag{6.46}$$

These equations are most readily obtained by applying the Lagrangian operator to the energy equation, which in these coordinates has the form

$$\frac{m}{2}\left[\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2\right] = eV \tag{6.47}$$

The difficulty in solving these two-dimensional problems arises from the fact that the variables in the component equations are rarely separable.

Example: One of the few two-dimensional problems that can be solved exactly is that of an electron released from a point on the side of an interior right-angled conducting corner at zero potential. The potential configuration is shown in Fig. 5.16. The equation for the potential is $V = kxy$ so that the components of electric intensity are $E_x = -ky$ and $E_y = -kx$. The differential equations of motion are then

$$\frac{d^2x}{dt^2} = \frac{e}{m}ky \tag{6.48}$$

and

$$\frac{d^2y}{dt^2} = \frac{e}{m}kx \tag{6.49}$$

It is convenient to make the substitution $\frac{e}{m}k = \omega^2$. If each equation is differentiated twice and the relations from the original equations substituted, there result

$$\frac{d^4x}{dt^4} = \omega^2 x \tag{6.50}$$

and

$$\frac{d^4y}{dt^4} = \omega^2 y \tag{6.51}$$

in which a separation of the variables has been achieved. When these equations are solved subject to the initial conditions that when

$$t = 0$$
$$x = 0$$
$$y = a$$

$$\frac{dx}{dt} = 0$$

$$\frac{dy}{dt} = 0$$

there results

$$x = \frac{a}{2}(\cosh \omega t - \cos \omega t) \tag{6.52}$$

and

$$y = \frac{a}{2}(\cosh \omega t + \cos \omega t) \tag{6.53}$$

The above solutions may be obtained either by standard methods or by the operational calculus. The nature of the solution is more apparent if the component displacements are referred to the line $y = x$, that is, if the system be rotated 45 deg clockwise. When this is done,

$$x_1 = \frac{a}{\sqrt{2}} \cosh \omega t \tag{6.54}$$

$$y_1 = \frac{a}{\sqrt{2}} \cos \omega t \tag{6.55}$$

This same result may be obtained more quickly if the original potential field be rotated 45 deg clockwise before formulating the differential equations. When this is done, the field has the form shown in Fig. 6.4 and the potential is

$$V = \frac{k}{2}(x_1^2 - y_1^2) \qquad \text{and} \qquad E_x = -kx_1, \qquad E_y = +ky_1$$

The differential equations are then

$$\frac{d^2x_1}{dt^2} = \omega x_1 \tag{6.56}$$

$$\frac{d^2y_1}{dt^2} = -\omega y_1 \tag{6.57}$$

Here the variables are already separated in equations of lower order; and when these are solved subject to the conditions that when

$$t = 0$$

$$x_1 = \frac{a}{\sqrt{2}}$$

$$y_1 = \frac{a}{\sqrt{2}}$$

$$\frac{dx_1}{dt} = 0$$

$$\frac{dy_1}{dt} = 0$$

the same solution as was obtained above results.

Examination of the solution shows that the electron oscillates about the line of symmetry while moving outward at a constantly increasing rate. It is also seen that the *shape* of the trajectory is independent of the strength of the field and also of the charge and mass of the electron. This is a general characteristic of such problems. The transit time, however, does depend upon all three of these factors. This means that a heavier particle starting under the same conditions will trace out the same path but be slower in doing so.

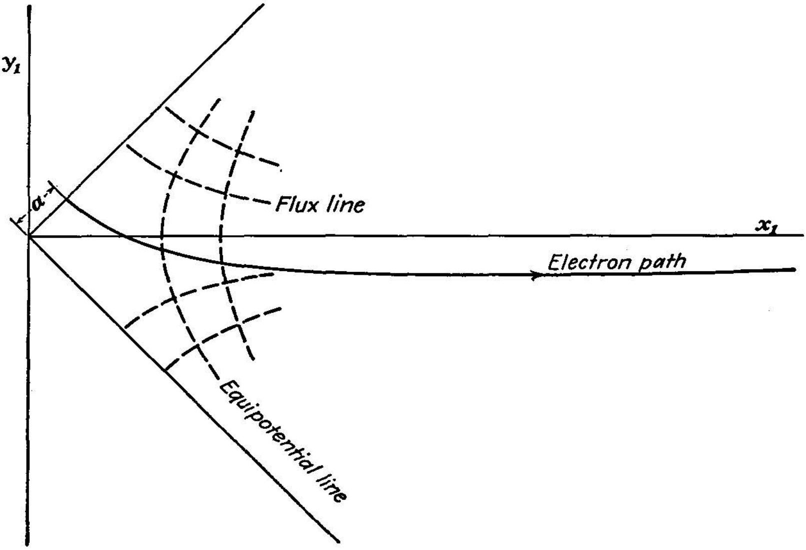


FIG. 6.4.—Path of an electron released from a point on the wall of a right-angled corner. Note that the electron does not follow a flux line but, because of its finite mass, overshoots the line of symmetry and subsequently oscillates about it.

If the general differential equations (6.41) and (6.42) are combined with the energy equation

$$\frac{m}{2}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right] = eV \tag{6.58}$$

and the factor t be eliminated between them, there results a differential equation in the coordinates x and y alone,

$$2V(x,y)\,\frac{d^2y}{dx^2} = \left[E_x\frac{dy}{dx} - E_y\right]\left[1 + \left(\frac{dy}{dx}\right)^2\right] \tag{6.59}$$

This equation is no easier to solve than those previously given, but several important properties of electron trajectories can be deduced from it.

1. The mass and charge of the electron do not appear in the equation. This means that the path taken will be independent of these factors.
2. The equation is not changed if either voltage or distance is changed by a constant factor. This means that the path will be the same for all magnitudes of voltage as long as the form of the field is not changed.
3. If the tube structure is enlarged by any factor, then the trajectory will be enlarged by the same factor.

6.6. Electron in a Uniform Magnetic Field. An electron in motion constitutes a minute electric current of magnitude $-ev$, where e is the magnitude of the charge on the electron and v is its velocity. As such, an electron in a magnetic field experiences a sidewise force just as does a wire carrying current. The magnitude of this force in newtons is $Bev \sin \theta$, where B is the magnetic-flux density in webers per square meter (1 weber per meter^2 equals 10^4 gausses) and θ is the angle between the vectors representing the field and the velocity, the latter being in units of meters per second. When the electron enters the field at right angles to it, the force is simply Bev directed at right angles to the velocity. The relative directions of field, velocity, and force are shown in Fig. 6.5. The force changes the direction but not the magnitude of the velocity and in this case is continuously exerted at right angles to the instantaneous velocity because the direction of the force changes with the direction of the velocity. This fulfills the conditions necessary for a circular motion of the electron in a plane normal to the magnetic field.

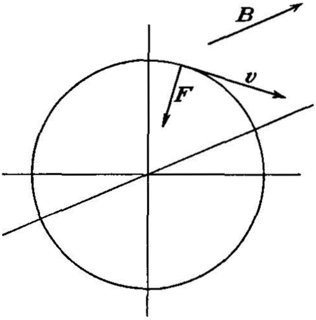


FIG. 6.5.—The direction of the force on an electron relative to the velocity and magnetic field that produce it. The force is the vector product of the magnetic-flux density and the velocity. If B is turned into v, then F advances like a right-hand screw.

The force developed by the magnetic field may be considered as a centripetal force that must equal the centrifugal force developed by the circular motion of the electron. Equating these forces,

$$Bev = \frac{mv^2}{R} \qquad \text{newtons} \tag{6.60}$$

where R is the radius of the circular electron path. From this the radius of the circular path is

$$R = \frac{m}{e}\frac{v}{B} \qquad \text{meters} \tag{6.61}$$

It is more convenient for application of this formula to express the physical quantities in numerical form and to use the potential corresponding to the velocity. With these changes the expression for the radius of the circular path becomes

$$R = 3.37 \times 10^{-6} \frac{\sqrt{V}}{B} \qquad \text{meters} \qquad (6.62)$$

where V is the electron velocity in equivalent volts and B is the magnetic-flux density in webers per square meter (10^4 gausses). This relation shows that, the stronger the field and the smaller the velocity of the electron, the smaller the circle in which it moves. The results of this relation are compactly presented in the nomogram of Fig. 6.6.

If the particle is not an electron but an ion of mass m_p and with n times the charge of the electron, the radius is given by

$$R_p = \frac{3.37}{B} \times 10^{-6} \sqrt{\frac{m_p}{m_e} \frac{V}{n}} \qquad \text{meters} \qquad (6.63)$$

where m_e is the mass of an electron and B is magnetic-flux density in webers per square meter.

Since the radius of the circle followed by the particle is proportional to the velocity, the period corresponding to one loop is independent of the initial velocity and depends only upon the magnetic field. The period is given by the circumference of the circle divided by the velocity,

$$T = \frac{2\pi m}{e} \frac{1}{B} \qquad \text{sec} \qquad (6.64)$$

In practical units this is

$$T = \frac{35.5}{B} \qquad \text{micromicroseconds} \qquad (6.65)$$

The value of the period can be obtained from the nomogram of Fig. 6.6 by observing that the period in microseconds is the same as the radius in centimeters when the potential is 11.22 mv.

For particles with a mass m_p and having n times the charge of the electron the period is

$$T = \frac{35.5}{nB} \frac{m_p}{m_e} \qquad \text{micromicroseconds} \qquad (6.66)$$

The fact that the period is independent of the velocity is significant and useful. If a number of electrons of different velocities be injected into a uniform magnetic field, they will trace out circles of different size but they will all return to the starting point at the same time. Use is made of this property in magnetic focusing of electron beams.

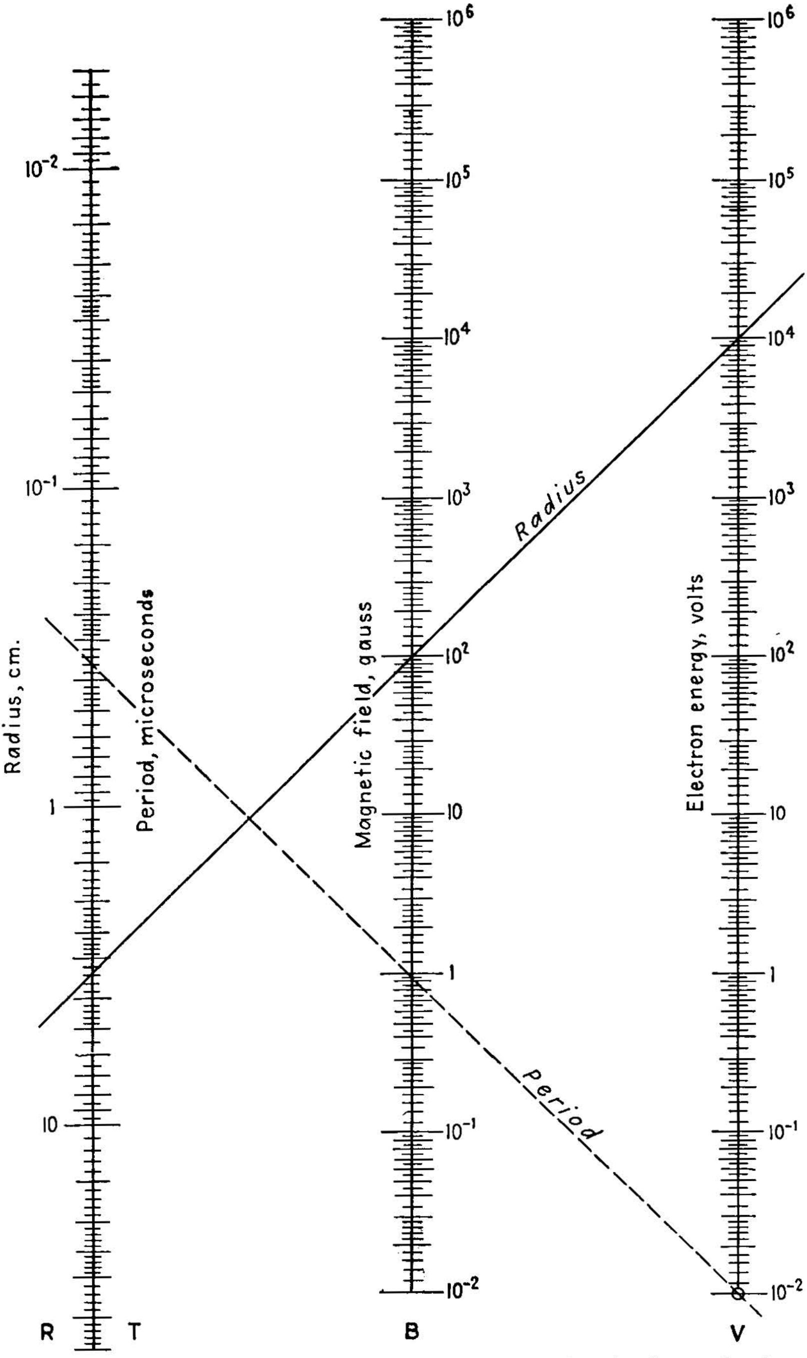


FIG. 6.6.—Nomographic chart giving the radius of the circular path of an electron in a uniform magnetic field as a function of the magnetic field strength and the electron energy in volts, Eq. (6.62). The chart also gives the period of a single rotation as a function of magnetic field strength, Eq. (6.65).

If an electron enters a magnetic field at an angle θ with the field, then there is a component of velocity parallel to the field, $v \cos \theta$, that is unaffected by the magnetic field. The other component of velocity, that normal to the field, $v \sin \theta$, produces a circular motion which combines with the parallel motion to give a helical path. The radius of the helix is given by

$$R = \frac{3.37 \times 10^{-6} \sqrt{V} \sin \theta}{B} \qquad \text{meters} \tag{6.67}$$

where V is in volts and B is in units of webers per square meter (10^4 gausses). The pitch of the helix is given by the product of the parallel component of velocity and the period as determined from Eq. (6.65).

$$p = \frac{21.2 \times 10^{-6} \sqrt{V} \cos \theta}{B} \qquad \text{meters} \tag{6.68}$$

It will be observed that for small angles the pitch does not vary much with the angle. Hence, if a magnetic field is placed parallel to a beam of electrons in a cathode-ray tube, the electrons will return to positions corresponding to their original relative position in a distance p along the beam. This is the principle of magnetic focusing, which is used to keep electron beams from spreading. All the electrons trace out helical paths of different radii but of the same pitch. Magnetic focusing cannot do more than reproduce the original beam diameter, and the field must be adjusted to produce this effect at the point desired.

6.7. Behavior of Electrons in Nonuniform Magnetic Fields. The paths followed by electrons in nonuniform magnetic fields are extremely complex. Little can be said about them except in certain simple limiting cases. *In all cases the magnitude of the velocity will be unchanged because no energy is added to or taken from the electron when subjected to the influence of a steady magnetic field alone.* In contrast, the direction of the velocity can experience very involved changes. The general form of the force equation depends upon the components of field and velocity. An x component of force results from a y component of field and a z component of velocity and also from a z component of field and a y component of velocity. Upon writing the components of force in terms of components of acceleration the general differential equations for three-dimensional rectangular components are

$$\left.\begin{aligned} \frac{m}{e}\frac{d^2x}{dt^2} &= B_y \frac{dz}{dt} - B_z \frac{dy}{dt} \\ \frac{m}{e}\frac{d^2y}{dt^2} &= B_z \frac{dx}{dt} - B_x \frac{dz}{dt} \\ \frac{m}{e}\frac{d^2z}{dt^2} &= B_x \frac{dy}{dt} - B_y \frac{dx}{dt} \end{aligned}\right\} \tag{6.69}$$

When the components of field vary from point to point, these equations are practically insoluble.

In electron-optics work, circular cylindrical coordinates r, θ, and z are used where the coordinates specify radial distance, angle, and axial distance, respectively. Here the equations have the same general form as Eqs. (6.69) but are quite different in their specific appearance. They are

$$\left.\begin{aligned} \frac{m}{e}\left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right] &= B_\theta \frac{dz}{dt} - B_z \frac{r\,d\theta}{dt} \\ \frac{m}{e}\left[\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right)\right] &= B_z \frac{dr}{dt} - B_r \frac{dz}{dt} \\ \frac{m}{e}\left[\frac{d^2z}{dt^2}\right] &= B_r \frac{r\,d\theta}{dt} - B_\theta \frac{dr}{dt} \end{aligned}\right\} \tag{6.70}$$

where the terms in the brackets on the left-hand side of the equations are the components of acceleration in the r, θ, and z directions, respectively.

Example: It is a known property that low-velocity electrons in a strong magnetic field will describe a tightly coiled spiral path which wraps itself around one of the flux lines and will thus follow the magnetic field. This property will be proved in the case of the magnetic field around a long, straight wire carrying current.

In this case there is only a θ component of field of magnitude $\frac{\mu I}{2\pi r}$, where I is the wire current. The r and z components of field are zero. For this condition, neglecting constants, Eqs. (6.70) become

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = \frac{I}{r}\frac{dz}{dt} \tag{6.71}$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = 0 \tag{6.72}$$

$$\frac{d^2z}{dt^2} = -\frac{I}{r}\frac{dr}{dt} \tag{6.73}$$

These equations cannot be solved exactly, but the nature of the path can be closely determined by some judicious approximations and observations. Integrating Eq. (6.73) with respect to time,

$$\frac{dz}{dt} = -I \ln\left(\frac{r}{r_0}\right) \tag{6.74}$$

where r_0 is a constant of integration. Integrating Eq. (6.72)

$$r^2\frac{d\theta}{dt} = C_\theta \tag{6.75}$$

where C_θ is a constant of integration. Substituting Eqs. (6.74) and (6.75) into Eq. (6.71),

$$\frac{d^2r}{dt^2} = \frac{I}{r}\left[-I \ln\left(\frac{r}{r_0}\right)\right] + \frac{C_\theta{}^2}{r^3} \tag{6.76}$$

This expression would be difficult to integrate exactly, but the form of the variation in r can be determined. It is seen that there is a value of r for which the acceleration is zero and that for values of r slightly less than this the acceleration is positive, while for values of r slightly greater than this the acceleration is negative. This means that, if the initial r component of velocity is small, the electron will oscillate about the value of r for which the acceleration is zero. Hence the expression for the r component of position is of the form

$$r = r_1 - a \sin kt \tag{6.77}$$

With this variation of r the z component of velocity is also seen to be periodic of small variation from Eq. (6.74), and hence z oscillates about its original value. Similarly the θ component of velocity is periodic and of small variation from Eq. (6.75) so that the value of $\frac{d\theta}{dt}$ is a constant with a superimposed periodic variation. The net result of these component displacements is that the electron will spiral around a flux line in some fashion, keeping a constant average value of r and z, and progress in the θ direction with a constant average velocity as shown in Fig. 6.7. Use of this property is made in television pickup tubes of the Orthicon type.[1]

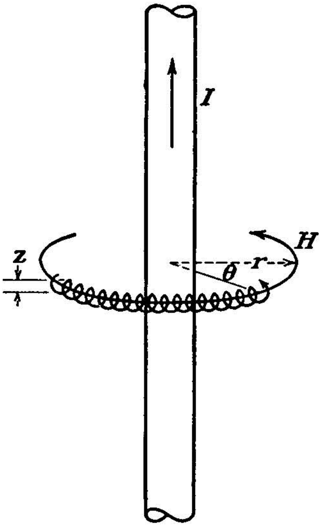


FIG. 6.7.—The motion of a low-velocity electron about a magnetic flux line. In the absence of strong electric fields, low-velocity electrons will spiral about magnetic flux lines.

6.8. Combined Electric and Magnetic Fields. When an electron is subjected to the combined action of both electric and magnetic fields, the paths tend to become quite complex. Some simple cases can be studied, however.

When an electron starts from rest under the influence of parallel electric and magnetic fields, the electron moves in the direction of the electric field and is unaffected by the magnetic field. The path in this case is a straight line, and the electron behaves as though the magnetic field did not exist.

If an electron with a given velocity is injected into a region containing electric and magnetic fields at right angles to each other and each at right

---

[1] ROSE, A., and H. IAMS, Television Pickup Tubes Using Low-velocity Electron Beam Scanning, *Proc. I.R.E.*, vol. 27, pp. 547–555, September, 1939.

angles to the initial velocity, then there is a certain ratio of electric- to magnetic-field strength for which the electron is not deflected in its path. This occurs when the force due to the electric field is equal and opposite to that produced by the magnetic field. For this condition

$$\frac{E}{B} = 5.93 \times 10^5 \sqrt{V_0} = v \qquad \text{meters per sec} \tag{6.78}$$

where v is the original velocity and V_0 the potential that produced it. As long as the above relation holds, the electron moves in a straight line. If any of the quantities involved is changed, the electron will be deflected from the straight-line path.

If an electron starts from rest in the presence of uniform electric and magnetic fields that are mutually perpendicular, it first experiences a force in the direction of the electric field and is unaffected by the magnetic field because of the low velocity. As it acquires velocity, it is deflected sidewise by the magnetic field. This action turns it around and brings it to rest at a point corresponding to its original position but displaced to one side. If the electron is then free to move, the action is repeated and the resulting path is a cycloid. The cycloidal nature of the path can be seen by considering that, if the magnetic field were moving in a direction mutually perpendicular to the electric field and to itself at a velocity given by Eq. (6.78), then to an observer moving with the magnetic field the effects of the two fields would cancel as far as forces parallel to the electric field were concerned. To this same observer the electron would behave as though it were injected into a magnetic field alone with a velocity given by Eq. (6.78) in a direction opposite to that of the observer's motion, and the resulting path would be a circle to this observer. To someone standing still relative to the fields the motion would be a circular motion combined with a translational motion, which in this case because of the equality of the velocity components gives rise to a cycloidal path.

For the relative position of the fields shown in Fig. 6.8, where B is in the negative z direction, the differential equations of motion are

$$\frac{d^2y}{dt^2} = \frac{e}{m}\frac{dV}{dy} - \frac{Be}{m}\frac{dx}{dt} \tag{6.79}$$

and

$$\frac{d^2x}{dt^2} = \frac{Be}{m}\frac{dy}{dt} \tag{6.80}$$

These equations are more simply written in the form

$$\ddot{y} = (a - \omega\dot{x}) \tag{6.81}$$
$$\ddot{x} = \omega\dot{y} \tag{6.82}$$

where the dots indicate derivatives with respect to time and $a = \dfrac{e}{m}\dfrac{dV}{dy}$ and $\omega = \dfrac{Be}{m}\cdot$ The initial conditions are that, when t is zero, y, x, $\dot{y}$, and $\dot{x}$ are also zero.

Integrating Eq. (6.82) with respect to time,

$$\dot{x} = \omega y \tag{6.83}$$

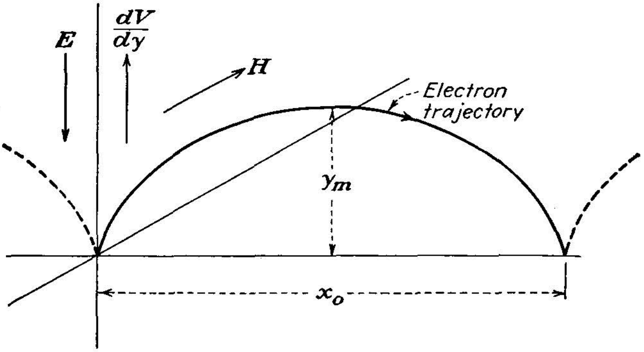


FIG. 6.8.—The cycloidal path resulting when an electron is liberated at zero velocity in crossed uniform electric and magnetic fields. The electron progresses in the positive x direction when the gradient of the electric field is in the positive y direction and the magnetic field is in the negative z direction.

since, when t equals zero, $\dot{x}$ and y are also zero. When this value of $\dot{x}$ is substituted in Eq. (6.81), there results

$$\ddot{y} = a - \omega^2 y \tag{6.84}$$

This can be solved either by standard methods or by the operational calculus to give

$$y = \frac{a}{\omega^2}(1 - \cos \omega t) \tag{6.85}$$

and the corresponding expression for x is from an integration of Eq. (6.83),

$$x = \frac{a}{\omega^2}(\omega t - \sin \omega t) \tag{6.86}$$

The last two equations above give the motion of the electron parametrically in terms of t. The motion is seen to consist of a uniform translation in the x direction with a superimposed circular motion.

The maximum displacement in the y direction is

$$y_m = \frac{2a}{\omega^2} = 1.138 \times 10^{-11} \frac{\frac{dV}{dy}}{B^2} \tag{6.87}$$

The displacement in the x direction corresponding to one cycle of the motion is found by substituting the value of time that restores the value of y to zero. This occurs for ωt equal to 2π so that

$$x_0 = \pi y_m \tag{6.88}$$

which is also to be expected from the ratio of the circumference to the radius of the generating circle that produces the cycloidal motion.

When an electron is injected into a region with uniform electric and magnetic fields at right angles to each other but with a finite initial velocity normal to the magnetic field, it will follow a trochoidal path in a plane normal to the magnetic field. Geometrically the trochoidal path is generated by a point on the rim of a wheel that is rolling along a straight line on a smaller diameter hub. The cycloid is the special case of the trochoid for which the diameters of the rolling and tracing circles are the same.

The differential equations for the case of an initial velocity are the same as for the cycloidal case [Eqs. (6.81) and (6.82)]; but in this case the initial conditions are different, and the form of the solution is hence different. When t is zero, y and x are zero, but $\dot{y} = v_{0y}$, $\dot{x} = v_{0x}$. Hence the first integration of Eq. (6.83) for the configuration of Fig. 6.8 gives

$$\dot{x} = \omega y - v_{0x} \tag{6.89}$$

When this substitution is made in Eq. (6.81) and this expression integrated twice to obtain the value for y, there results

$$y = \frac{(a - \omega v_{0x})(1 - \cos \omega t)}{\omega^2} + \frac{v_{0y}}{\omega} \sin \omega t \tag{6.90}$$

Substituting this in Eq. (6.89) and integrating to get the corresponding expression for x,

$$x = \frac{at}{\omega} + (1 - \cos \omega t)\frac{v_{0y}}{\omega} - \frac{a - \omega v_{0x}}{\omega^2} \sin \omega t \tag{6.91}$$

The two equations (6.90) and (6.91) determine completely the nature of the trochoidal path. The corresponding expressions for the components of velocity are

$$\dot{y} = v_{0y} \cos \omega t - \frac{a - \omega v_{0x}}{\omega} \sin \omega t \tag{6.92}$$

$$\dot{x} = \frac{a}{\omega} + v_{0y} \sin \omega t - \frac{a - \omega v_{0x}}{\omega} \cos \omega t \tag{6.93}$$

From the velocity-component equations it is seen that there is a constant x component of velocity of magnitude $\frac{a}{\omega}$. This corresponds to the translational velocity of the circles that generate the trochoidal motion. To achieve this translational velocity the radius of the rolling circle must be $\frac{a}{\omega^2}$ since the angular velocity of the velocity vectors is ω. The initial conditions also require that the instantaneous velocity of rotation of a point on the tracing circle be equal to the vector difference of the initial velocity and the translational velocity. This relation is shown by the vector diagram of Fig. 6.9.

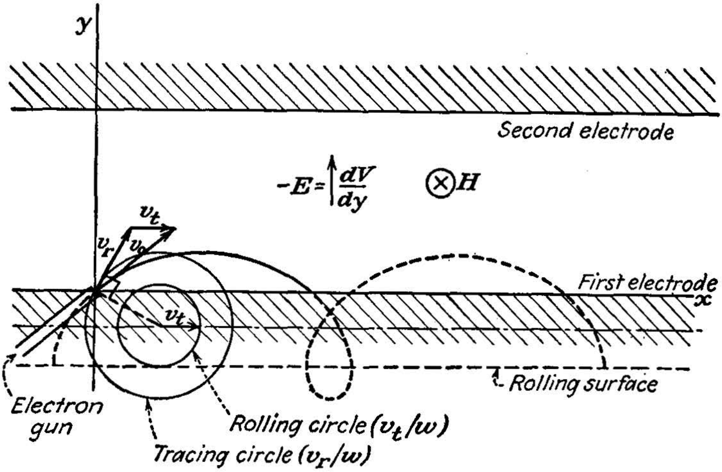


FIG. 6.9.—Trochoidal electron path resulting when an electron is injected with a finite velocity into a region of uniform crossed electric and magnetic fields. The electron will progress in the positive x direction when the gradient of the electric potential is in the positive y direction and the magnetic field is in the negative z direction.

If the scale of the velocity-vector diagram be taken the same as that of the diagram showing the generating circles and the resulting path, the electron path can be constructed graphically in quite a simple manner. It will be observed that the terms in Eqs. (6.90) and (6.91) giving the instantaneous displacements are the same as those in Eqs. (6.92) and (6.93) for the component velocities except for the ω factors. The center of the generating circles in the initial position is given by rotating the rotational vector v_r, 90 deg in a clockwise direction. The radius of the tracing circle is then given by the length of the rotational vector v_r, and the radius of the rolling circle is given by the length of the vector

v_t, the translational velocity. With the rolling and tracing circles and their initial position given, the path is readily constructed geometrically for any case. This construction is also illustrated in Fig. 6.9.

6.9. Approximate Numerical and Graphical Methods for Determining Electron Paths. The number of cases in which the motion of an electron under the influence of applied fields can be determined exactly is actually quite small and restricted to very simple cases. Hence the need exists for methods that will give an approximate answer when the fields are more complex, as they usually are.

Method of Joined Circular Segments. When an electron is moving through a potential field, the instantaneous radius of curvature of its path is determined by its velocity and by the sidewise force that is exerted on it by the field.[1] The sidewise force exerted on the electron depends upon the component of the gradient of potential normal to the instantaneous direction. This component of the gradient will be designated by $\nabla_n V$. The actual sidewise force is $e\nabla_n V$. This force must equal the centrifugal force of the electron in its instantaneous circular motion, and this is given by $\frac{mv^2}{R}$, where R is the instantaneous radius of curvature. Equating these two forces and substituting $2eV$ for mv^2,

$$\frac{2eV}{R} = e\nabla_n V \tag{6.94}$$

from which

$$R = \frac{2V}{\nabla_n V} \tag{6.95}$$

This is the instantaneous radius of curvature of the electron path at any point in the field, as shown in Fig. 6.10, on the assumption that the electron started from rest at a point of zero potential. It will be observed that the radius of curvature is independent of the mass and charge of the electron and also of the scale of potential, checking the conclusions drawn from the differential equations of the electron path.

By calculating the radius of curvature at a point in the field by Eq. (6.95), drawing a small segment of arc, and then applying this process repeatedly a good approximation to the actual curve traced by the electron is obtained. The potential at any point in the field is easily obtained, and the normal component of gradient is the projection of the vector giving the magnitude and direction of the greatest variation of potential upon a line normal to the electron's path. The method is subject to cumulative error unless the average potential and average normal

---

[1] Salinger, H., Tracing Electron Paths in Electric Fields, *Electronics*, vol. 10, pp. 50–54, October, 1937.

gradient over each segment of arc is used. If the segment of arc drawn at each step is kept a constant fraction of the radius of curvature, say one-twentieth, the error will not be great. It is also possible to derive simple expressions for the position of the next step in terms of the displacement and change of angle when the radius of curvature is so large that the arc segment is not easily drawn, as is frequently the case.

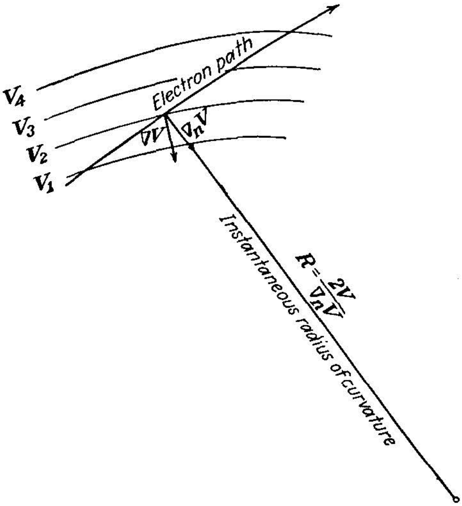


FIG. 6.10.—The instantaneous radius of curvature of an electron path in a region of varying potential. The instantaneous radius of curvature is equal to twice the potential at the point in question divided by the component of the gradient of potential perpendicular to the path.

Several ingenious gadgets have been devised that make the application of the principle outlined above purely a mechanical one.[1,2] These make use of a double probe in a current-flow model that has been set up to give the electric field involved. The double probe picks up a voltage proportional to the component of gradient in the direction of its alignment, and the average potential of the probes gives the potential at the point. The probe is connected to a small cart attached to a pantograph. The cart is steered in such a way that the instantaneous curvature of path which it is tracing is determined by the relation Eq. (6.95). Adjust-

[1] GABOR, D., Mechanical Tracer for Electron Trajectories, *Nature*, vol. 139, p. 373, February, 1937.

[2] LANGMUIR, D., Automatic Plotting of Electron Trajectories, *Nature*, vol. 139, pp. 1066–1067, June 19, 1937.

ments on the steering are made as continuously as possible from the information picked up by the double probe as it traces out the path. These devices are capable of considerable accuracy.

Use of Elastic-membrane Model of Potential to Determine Electron Paths. In a previous chapter it was pointed out that the elevation of the surface of a stretched elastic membrane approximated closely the solution of Laplace's equation when the deformations were small. Actually such a model of potential fields is of more use in determining electron paths than in solving potential problems. This is because it is found that the laws governing the motion of a small sphere rolling on the membrane are strictly analogous to the laws governing the motion of an electron in an electric field, and hence the path of such a sphere is a good approximation of the path of an electron in the corresponding electric field.

Except for frictional effects the kinetic energy picked up by a small sphere, say a $\frac{3}{16}$-in. ball bearing, is equal to the potential energy it has lost owing to its change in elevation. This is exactly what happens to the electron. In the case of the mechanical model, however, the kinetic energy is divided between translational and rotational components. As long as the sphere rolls with a given circle of contact, the proportionality between these two components of the kinetic energy is constant and the path of the sphere will be similar to that of the electron. Although it is difficult to prove mathematically, it can readily be shown by experiments with a large sphere on a hard, curved surface that the sphere will turn relatively sharp corners and finish with the same rolling circle of contact as it had initially. The sphere can actually change its direction by about 300 deg without losing its original circle of rolling contact. If the angle is more than 300 deg, the turn introduces a spinning action that spoils the energy relations indicated above. Actually, it is the radius of curvature of the path rather than the angle that matters. Roughly, the limiting radius of the path is five times the radius of the sphere.

Application of the Principle of Least Action. In many electron-trajectory problems use can be made of the principle of least action. This principle states that in a potential field of the type encountered in vacuum-tube problems a particle will move between two points by such a path that the action, defined as the integral of momentum with distance, will have a minimum value. This means that, if the paths are known and conditions are such that only one electron goes through each point, contours of constant action calculated from the defining integral will be everywhere perpendicular to the electron paths. Hence, if the electron paths are known, contours of constant action can be found that give the path of all electrons. If the electron paths are not known, as is generally the case, it is still possible to calculate the contours of constant action by methods involving successive approximation.

This is done by assuming a path of the electrons that is known to be close to the true path and then calculating the action along the assumed path. The process is easy, for the square root of potential can be substituted for velocity so that action is given by

$$A = \int mv\,ds = \sqrt{2em} \int \sqrt{V}\,ds \tag{6.96}$$

When the action along the assumed paths has been calculated, a first approximation to the action function has been obtained and corresponding contours of constant action can be drawn. If now curves be drawn normal to these contours of constant action, these will give a more accurate representation of the electron paths than those originally assumed. The second approximation to the action function can then be calculated along the improved paths and the process repeated to give any desired degree of accuracy. This is seen to be a perturbation process between action and potential.

In actual application in cases where the electron deflections are slight the first step of the process gives results that are sufficiently accurate.[1,2] The errors involved compensate because of the fact that, when the assumed path is shorter than the actual path, the potentials involved are smaller. In the determination of electron paths in tubes the assumption of straight-line paths initially is usually sufficiently good for cases in which the electron deflections are slight. The method is not accurate in the vicinity of any line of symmetry.

---

[1] LANGE, H., Current Division in Triodes and Its Significance in the Determination of Contact Potential, *Zeit. für Hochfrequenz.*, vol. 31, pp. 105–109, 133–140, 191–196, 1928.

[2] SPANGENBERG, KARL, Current Division in Plane-electrode Triodes, *Proc. I.R.E.*, vol. 28, pp. 226–236, May, 1940.

CHAPTER 7

ELECTROSTATIC FIELD OF A TRIODE

7.1. Method of Solution. The electrostatic fields within tubes are most readily obtained by means of the conformal transformations given in a previous chapter. These transformations give potential configurations that represent closely the fields encountered in tubes, whose electrode configuration is somewhat idealized. The cathode is assumed to be a plane or cylindrical surface, which it rarely is in practice. The electrodes are assumed to be infinite in length and breadth so that tube constants per unit area evaluated on this assumption do not include end effects.

It should be pointed out that, since the solutions obtained are not mathematically exact, various degrees of approximation are possible. In general, the more accurate the solution, the more complex and cumbersome the expressions obtained. Where extreme accuracy is desired, the method of conformal transformations is used as a starting point for series representations. Imaging or series procedures may also be used, but these have not proved of great value as a general method.

In spite of the above-mentioned departures from exactness the formulas obtained by the application of the method of conformal transformations meet the accuracy requirements of modern engineering.

7.2. Electrostatic Field of a Plane-electrode Low-mu Triode. The field of a low-mu triode may be determined by a method outlined by Maxwell.[1] Vacuum tubes had not yet been invented in Maxwell's time, but his analysis of the electrostatic field about a shielding screen of parallel wires is readily applied to the problem of the triode field.

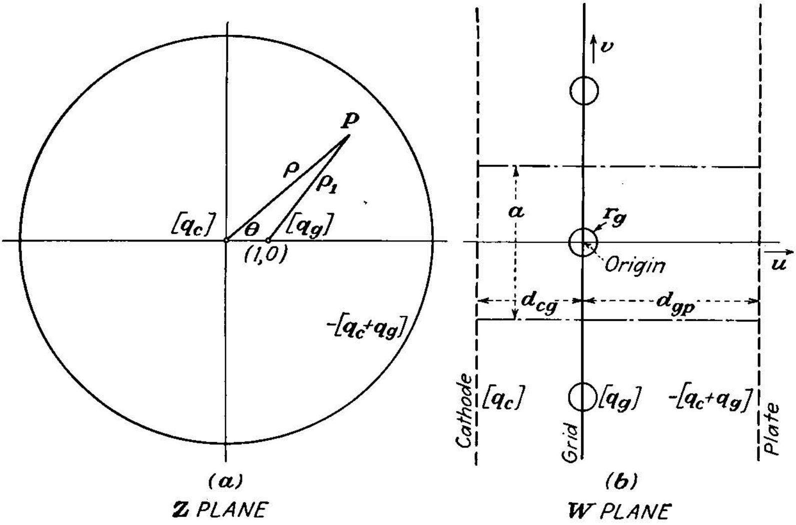
The field analysis is based upon the Z-plane configuration shown in Fig. 7.1*a*. This consists of two line charges located within a large cylinder. One line charge is located at the origin and has a linear-charge density of $+q_c$. The other is located at the point (1,0) and has a linear-charge density of $+q_g$. The field at great distances from these lines is nearly circular and may be fitted to that of a circular electrode whose radius is large compared with the distance between the line charges. It may be seen that the Z-plane representation represents a simple tube

---

[1] Maxwell, J. Clerk, "Electricity and Magnetism," 3d ed., Vol. I, Sec. 203, Cambridge, London, 1904.

with a cathode wire at the origin and a single grid wire at the point (1,0) surrounded by a circular plate. This simple tube has electrical characteristics the same as those of the plane-electrode and cylindrical-electrode structures that may be derived from it.

To obtain the field within the plane-electrode tube it is necessary to obtain an expression for the field in the Z plane of Fig. 7.1*a* and then transform this expression by the logarithmic transformation to fit that of the electrode configuration of Fig. 7.1*b*, which closely represents the structure of a practical tube. The potential at any point in the Z plane is given by the sum of the potentials resulting from each of the line charges. Polar coordinates will be used in the Z-plane relations.



(*a*) Z PLANE

(*b*) W PLANE

FIG. 7.1.—Elementary triode and equivalent plane-electrode triode.

The potential at any point (ρ,θ) is given by

$$V_z = -\frac{q_g}{2\pi\varepsilon_0}\ln\rho_1 - \frac{q_c}{2\pi\varepsilon_0}\ln\rho + C \tag{7.1}$$

where ρ_1 is the distance from the point in question to the grid-wire charge at the point (1,0) and C is a constant that adjusts the level of potential, and *ln* will be used hereafter to denote the natural logarithm. Making use of the law of cosines,

$$V_z = -\frac{q_g}{4\pi\varepsilon_0}\ln(\rho^2 + 1 - 2\rho\cos\theta) - \frac{q_c}{4\pi\varepsilon_0}\ln\rho^2 + C \tag{7.2}$$

The logarithmic transformation with a suitable coefficient will be used. The coefficient is selected so that in the plane-electrode structure of Fig. 7.1*b* the grid wires are spaced a distance a.

$$W = \frac{a}{2\pi} \ln Z \tag{7.3}$$

The component parts of this equation are

$$u = \frac{a}{2\pi} \ln \rho \qquad \text{or} \qquad \rho = \epsilon^{\frac{2\pi u}{a}} \tag{7.4}$$

$$v = \frac{a\theta}{2\pi} \qquad \text{or} \qquad \theta = \frac{2\pi v}{a} \tag{7.5}$$

in which u and v are the real and imaginary parts of W, respectively, and ρ and θ are the polar coordinates in the Z plane. Making the above substitutions in Eq. (7.2),

$$V_w = -\frac{q_g}{4\pi\varepsilon_0} \ln \left(\epsilon^{\frac{4\pi u}{a}} + 1 - 2\epsilon^{\frac{2\pi u}{a}} \cos \frac{2\pi v}{a} \right) - \frac{q_c}{4\pi\varepsilon_0} \ln \epsilon^{\frac{4\pi u}{a}} + C \tag{7.6}$$

The above expression gives very closely the potential inside of a plane triode. Examination of its form will show that the equipotential lines in the vicinity of the origin and the points $(0, \pm na)$ are circles, one set of which may be fitted to the grid wires. For large positive and negative values of u the equipotentials are almost planes that may be fitted to the plate and cathode planes, respectively. The general potential expression of Eq. (7.6) gives potential in terms of the charges q_g, q_c, and the constant C. For application it is also necessary to evaluate these constants in terms of the electrode potentials.

To evaluate the constants of Eq. (7.6) let the plate plane be located at $u = +d_{gp}$ where $d_{gp} \geqq a$. When this relation between d_{gp} and a holds, the second and third terms of the argument of the first logarithm will be less than 1 per cent of the first term and may thus be neglected. Making the substitution $u = +d_{gp}$ into Eq. (7.6),

$$V_p = -\frac{d_{gp}q_g}{a\varepsilon_0} - \frac{d_{gp}q_c}{a\varepsilon_0} + C \tag{7.7}$$

Let the cathode plane be located at $-d_{cg}$, where $d_{cg} \geqq a$. In this case the first and third terms of the first logarithm argument will be small compared with 1 so that the first term is substantially zero. Making the substitution $u = -d_{cg}$ into Eq. (7.6),

$$V_c = 0 + \frac{d_{cg}q_c}{a\varepsilon_0} + C \tag{7.8}$$

Let the grid wires be located at the points $(0, \pm na)$ and be of radius r_g and potential V_g. If $r_g \leqq \frac{a}{20}$, the potentials at points $(0, r_g)$ and $(r_g, 0)$

differ by only a few per cent. Into Eq. (7.6) substitute the values $u = 0$, $v = r_g$.

$$V_g = -\frac{q_g}{2\pi\varepsilon_0}\ln\left(2\sin\frac{\pi r_g}{a}\right) + C \tag{7.9}$$

If the cathode potential is set at a reference level of potential of zero, then

$$C = -\frac{d_{cg}q_c}{a\varepsilon_0} \tag{7.10}$$

from which Eqs. (7.7) and (7.9) become

$$V_p = -\frac{d_{gp}q_g}{a\varepsilon_0} - \frac{(d_{cg} + d_{gp})q_c}{a\varepsilon_0} \tag{7.11}$$

and

$$V_g = -\frac{\ln\left(2\sin\frac{\pi r_g}{a}\right)q_g}{2\pi\varepsilon_0} - \frac{d_{cg}q_c}{a\varepsilon_0} \tag{7.12}$$

It is already possible to obtain the amplification factor of the tube from Eqs. (7.11) and (7.12). The amplification factor of a tube is the ratio of the plate voltage to the negative of grid voltage for a condition of cutoff.[1] In terms of the electric field within the tube, cutoff exists when the gradient of potential at the cathode is zero, which in turn occurs when the cathode charge is zero. If q_c is made zero in the above two equations and the ratio taken as indicated,

$$\mu = \frac{-2\pi d_{gp}}{a\ln\left(2\sin\frac{\pi r_g}{a}\right)} \tag{7.13}$$

If Eqs. (7.11) and (7.12) are solved for q_c and q_g in terms of V_p and V_g and the expressions simplified by use of Eq. (7.13),

$$q_c = -\frac{\varepsilon_0 a(V_p + \mu V_g)}{(d_{gp} + d_{cg} + \mu d_{cg})} \tag{7.14a}$$

and

$$q_g = \frac{\varepsilon_0 a\mu[(d_{gp} + d_{cg})V_g - d_{cg}V_p]}{d_{gp}(d_{gp} + d_{cg} + \mu d_{cg})} \tag{7.14b}$$

The expression for μ of Eq. (7.13) given above is the simplest expression that adequately gives the amplification factor of a plane-electrode triode. Examination of this expression shows that the amplification factor increases as the grid-plate distance increases. This is in accord

---

[1] For a more general definition of the amplification factor see the chapter on Triode Characteristics.

with physical reasoning since the more remote the plate is the more influence the grid has. Amplification factor is also increased if the grid-wire spacing is decreased since this makes the grid more effective in controlling the off-cathode gradient of potential. Amplification is also increased if the grid-wire radius is increased, as would be expected. It will be observed that according to Eq. (7.13) the amplification factor is independent of the cathode-grid distance. This is approximately true as long as the approximations made in deriving the expression are not exceeded, *i.e.*, as long as the cathode-grid distance is not less than the grid-wire spacing. This may be understood by considering that the cathode charge is zero at cutoff. Thus for a cutoff condition all the flux lines originating on the grid terminate on the plate, and though some of them start toward the cathode they turn and end on the plate so that as long as the cathode is not too close to the grid the field pattern is not disturbed and the amplification factor is independent of cathode-grid distance. This interpretation will be discussed further in connection with equipotential contours and potential-profile plots.

Contour Representation of Potential Field. The form of the potential field resulting from the equations developed above may be best studied by examining the plots of the equipotential lines. A group of these equipotential contour plots of a typical plane triode are shown in Fig. 7.2 for various potentials. The contours of Fig. 7.2*a* show the field configuration for the case of the grid biased beyond cutoff. It will be observed that the gradient of potential at the cathode is negative. In the line of the grids the potential is increasingly negative in moving from cathode to grid. Along this same line the potential is increasingly positive in moving from grid to plate. Along the line from cathode to plate midway between grid wires the potential is at first negative and then positive. The dotted lines shown represent the boundary between the various types of equipotential lines. In the area within the dotted lines including the grid wires the equipotential lines are closed curves about the grid wires. In the other areas the equipotential lines run continuously from one section of the triode to adjoining sections, always on one side of the grid plane.

The other equipotential plots may be interpreted in a similar fashion. It will be observed that all the plots have some features in common. The equipotential lines close to the grid wires are nearly circular in all cases. The equipotential lines close to the plate and cathode are nearly straight lines. The equipotential lines may be divided into two groups, those which completely enclose the grid and those which run along from section to section. It will be observed that in some cases the equipotential lines of the second type listed above cross the grid plane between

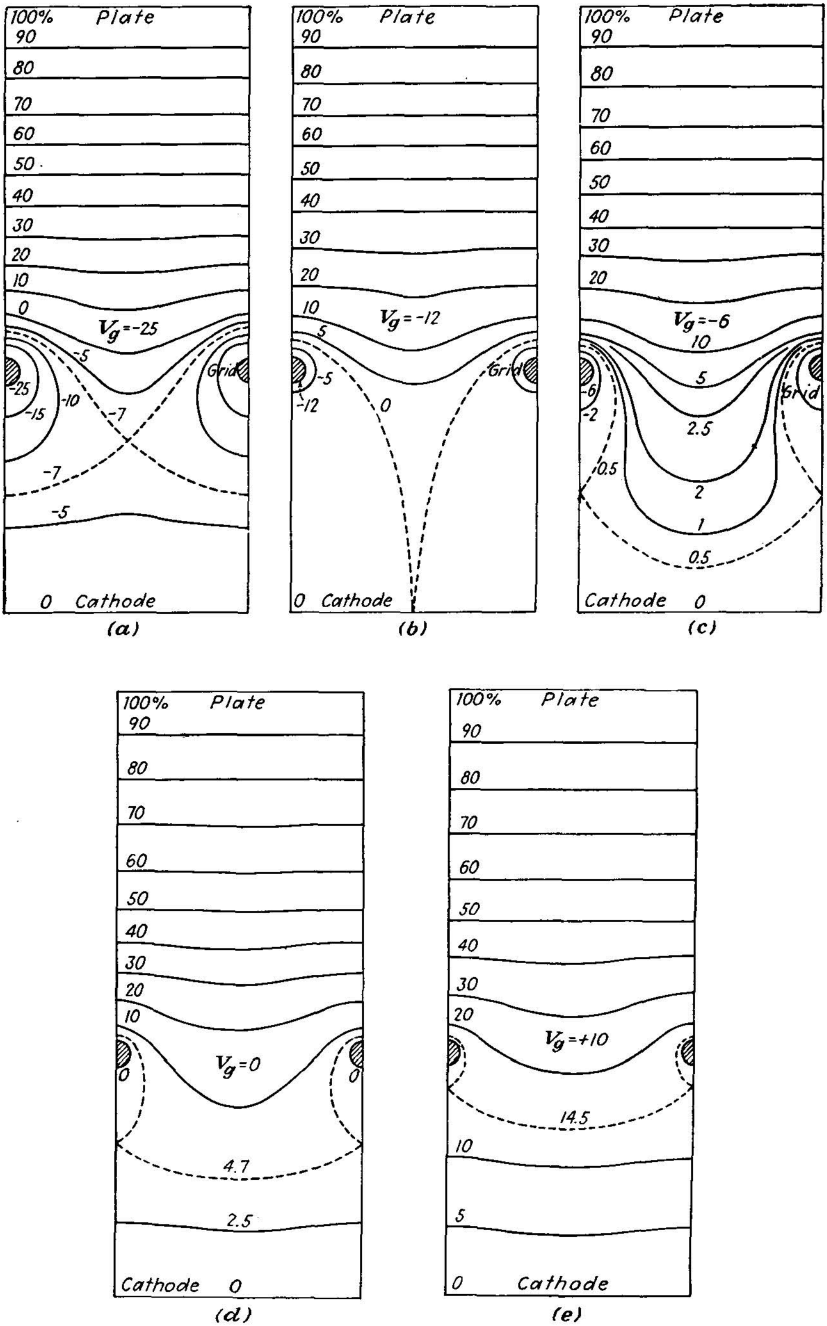


FIG. 7.2.—Equipotential contours in the plane-electrode triode: (*a*) grid beyond cutoff potential; (*b*) grid at cutoff potential; (*c*) grid negative at half cutoff value; (*d*) grid at zero potential; (*e*) grid positive.

grid wires but always pass the grid wires on the same side in moving from section to section. This behavior is observed in the case of the 10 per cent contour for $V_g = 0$. In all the cases shown the equipotentials are bowed toward the cathode through the grid plane. Only if the grid is more positive than the potential that gives a uniform positive potential gradient from plate to grid will the contours that cross the grid plane be bowed toward the plate. The equipotential plots shown in Fig. 7.2 were obtained with an electrolytic-trough model of potential. The equipotentials calculated from Eq. (7.6) would be almost the same in shape. For comparison a contour plot calculated by Eq. (7.6) is shown in Fig. 7.3. This plot represents an extreme condition of potential and dimensions. The grid-wire radius is $\frac{1}{20}$ of the grid-wire spacing. It will be observed that the grid-wire contour is not quite circular. It is of proper width in the plane of the grid wires but is longer in the direction at right angles to this. Because of this distortion of shape, which increases as the ratio of grid-wire diameter to grid-wire spacing increases, the formula for the amplification factor of Eq. (7.13) becomes inaccurate when the above ratio, known as the *screening fraction*, becomes greater than $\frac{1}{10}$. In the following section a more accurate formula is given, which is good up to screening fractions of $\frac{1}{6}$.

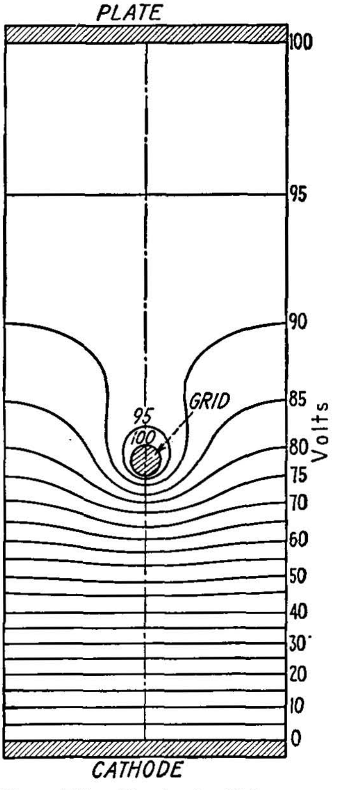


FIG. 7.3.—Equipotential contours in a plane-electrode triode with equal positive grid and plate potentials.

Profile Representation of Potential Field. The potential fields of a low-mu triode may also be studied by reference to profile representations of potential. These curves show how the potential varies along certain lines within the tube. The most common profile representations are shown along lines running from cathode to plate. In particular, two profiles are particularly informative. These are the profiles through the grid wire in a line running from cathode to plate at right angles to each of the latter, and in a line midway between grid wires. Such profiles are shown in Fig. 7.4.

In Fig. 7.4*a* are shown the profiles for a condition of a tube biased beyond cutoff. Here it is seen that the gradient of potential at the cathode is negative, thus making it impossible for electrons to leave the cathode. This is true because most electrons that do succeed in getting away because of some initial velocity are driven back by the negative gradient of potential. In the line of the grids the potential goes strongly negative until it reaches the negative grid potential. Beyond the grid the gradient is positive. In the line between the grid wires the potential is pulled negative at first by the presence of the negative grid, and it then becomes positive. It will be observed that the potential profiles are straight lines near the cathode and also near the plate. Further, it is only in the vicinity of the grid that there is a great variation in the value of potential in moving parallel to the cathode and plate planes.

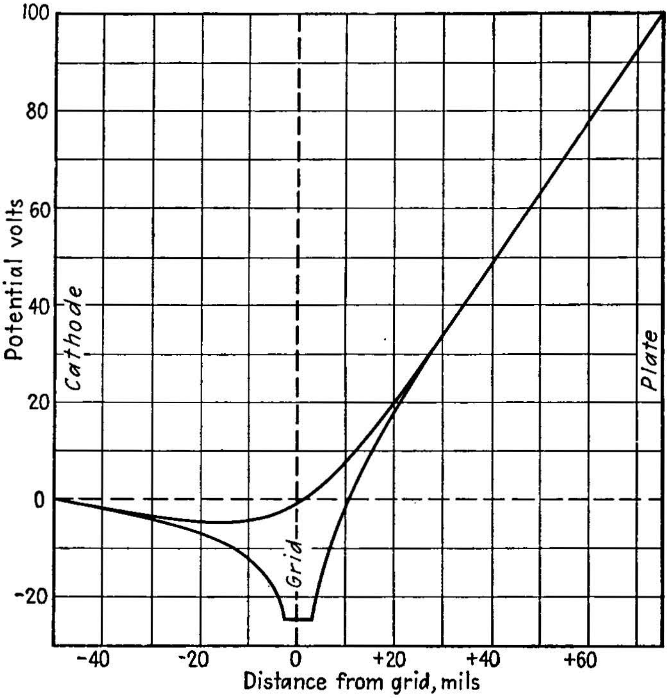


FIG. 7.4*a*.—Potential profiles of a plane-electrode triode, with grid at twice the cutoff value of potential.

In Fig. 7.4*b* are shown the potential profiles for the case of the grid biased to approximately cutoff. Here it is seen that the gradient at the cathode is zero. In the line of the grid wires the potential first goes negative to the value of grid potential and then positive. In the line between the grid wires the potential becomes increasingly positive in moving from cathode to plate. In this representation the amplification factor of the tube is given by the ratio of positive plate to negative grid potential. It is evident from these profiles why the amplification factor is independent of the cathode-grid distance provided that this is not too small. Up to a distance of about half the cathode-grid spacing the potential on both profiles is substantially zero for the particular dimensions shown. Hence in this particular case a cathode at zero potential could be put at any distance greater than half the cathode-grid distance shown without changing the shape or position of the potential profiles to the right of the profiles. The curves of Fig. 7.4*b* show the potential conditions that will just allow current to flow.

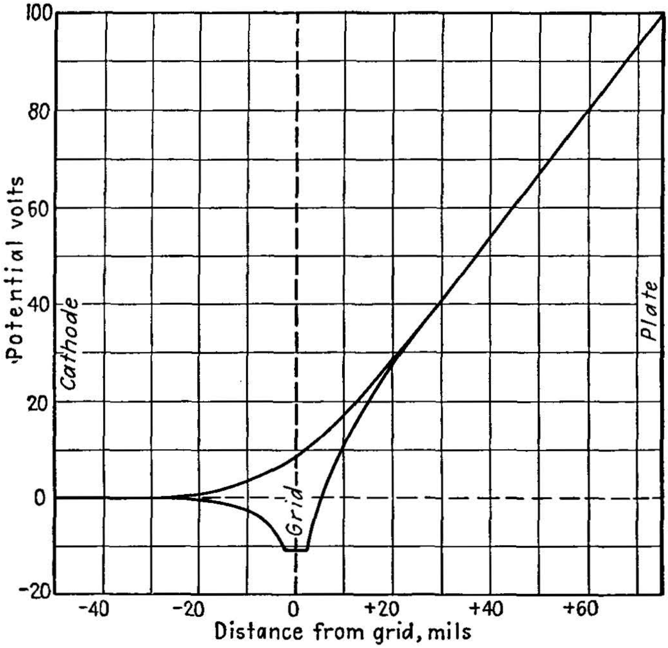


FIG. 7.4*b*.—Potential profiles of a plane-electrode triode, with grid at the cutoff value of potential.

In Fig. 7.4*c* are shown profiles for a negative grid potential greater than that which gives the cutoff condition. Here the gradient of potential at the cathode is positive even though the grid is negative. The curves shown represent the potentials that would exist in the absence of current, say in a cold tube. Although this condition of potential would permit current to flow, the actual flow would depress the profiles in the vicinity of the cathode, as will be described in a later chapter. In Fig. 7.4*d* the grid is at zero potential, and it is now possible for electrons to reach the grid, which has not previously been possible for nega-

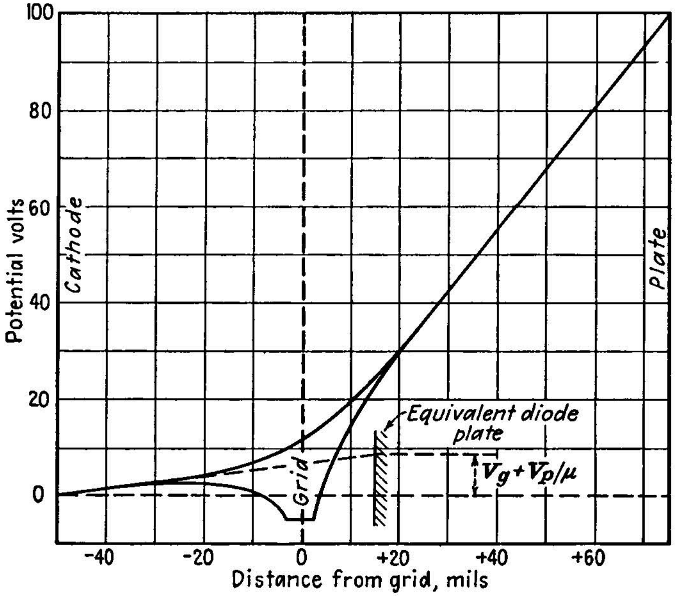


FIG. 7.4*c*.—Potential profiles in a plane-electrode triode, with grid negative at half the cutoff value of potential, which is the usual Class A operating condition.

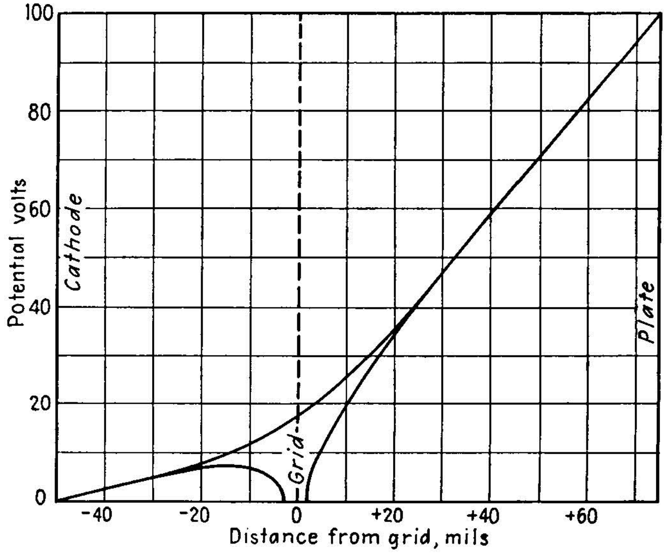


FIG. 7.4*d*.—Potential profiles in a plane-electrode triode, with grid at zero potential.

tive grid potentials. In Fig. 7.4*e* is shown an extreme condition of positive grid potential. Here the grid is as positive as the plate. This condition may be reached at the peak of the cycle in Class C power amplifiers.

In all cases the profiles are straight lines in the vicinity of the cathode and plate. For a condition of grid potential more negative than that of cutoff the slope of the straight-line portion at the cathode is negative. Above cutoff it is positive. The general form of the profiles corresponds to that which one would expect from a deformed elastic membrane. In each case the grid pushes a hole in what would otherwise be a straight-line profile from cathode to plate. Curvature requirements are met here. It will be observed that when one profile is concave upward the other is concave downward.

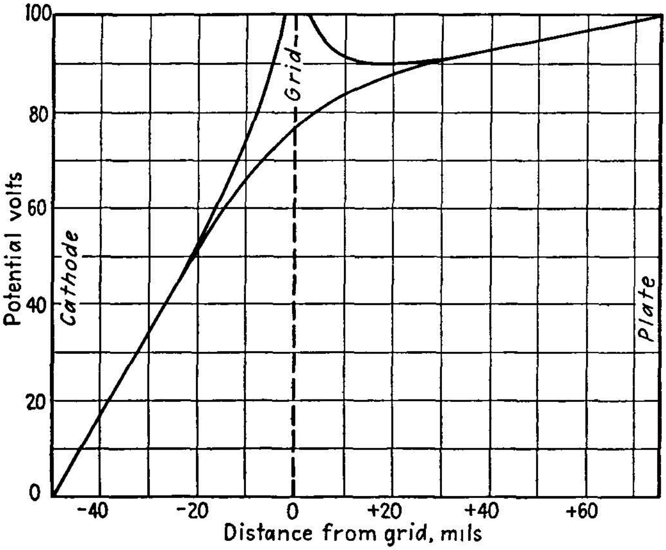


FIG. 7.4*e*.—Potential profiles in a plane-electrode triode, with grid and plate at the same positive potential. This condition may exist at the peak of the current pulse in a Class C amplifier.

7.3. Electrostatic Field of a Low-mu Cylindrical-electrode Triode. The same fundamental tube configuration as was used for the plane-electrode triode in Fig. 7.1*a* can be used to develop the cylindrical-electrode triode. In this case, however, the transformation equation takes the form

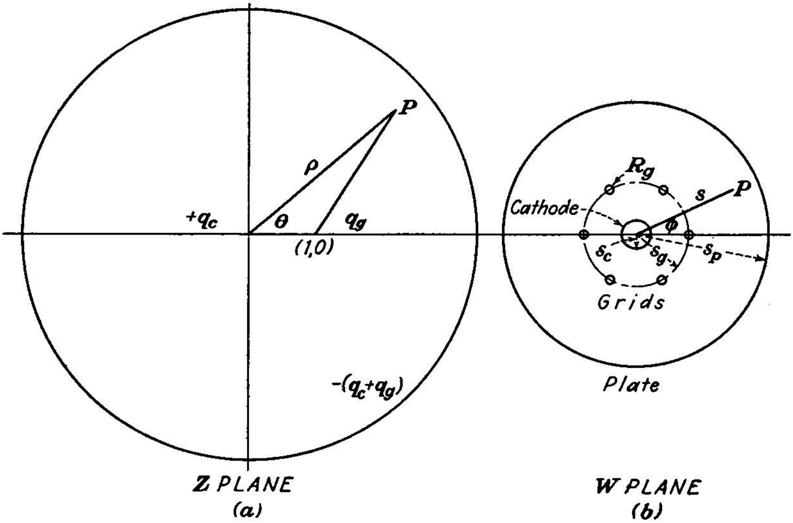
$$W = s_g Z^{1/N} \tag{7.15}$$

to give the electrode arrangement of a cylindrical triode with N grid wires as shown in Fig. 7.5*b*. Let the coordinates in the Z plane be and θ and in the W plane s and ϕ; let s_g be the radius of the grid-wire circle. Then the components of the transformation equation in polar coordinates are

$$\rho = \left(\frac{s}{s_g}\right)^N \tag{7.16}$$

and

$$\theta = N\phi \tag{7.17}$$



Z PLANE
(a)

W PLANE
(b)

FIG. 7.5.—Elementary triode and equivalent cylindrical-electrode triode.

As before, the equation for the potential at any point in the Z plane is

$$V_z = -\frac{q_g}{4\pi\varepsilon_0}\ln(\rho^2 + 1 - 2\rho\cos\theta) - \frac{q_c}{4\pi\varepsilon_0}\ln\rho^2 + C \tag{7.2}$$

Substitution of the component transformation equations gives

$$V_w = -\frac{q_g}{4\pi\varepsilon_0}\ln\left[\left(\frac{s}{s_g}\right)^{2N} + 1 - 2\left(\frac{s}{s_g}\right)^N\cos N\phi\right] - \frac{2q_c}{4\pi\varepsilon_0}\ln\left(\frac{s}{s_g}\right)^N + C \tag{7.18}$$

This gives the equation of a potential field in which the contours are circles close to the origin and at great distances from the origin. The contours are also circles about the points $\left(s = s_g, \phi = \frac{2\pi k}{N}\right)$, where k assumes integral values from zero to N.

The three sets of circles can be fitted to the cathode, plate, and grid wires, respectively. To fit the cathode to one of the circles close to and about the origin let $s = s_c \ll s_g$. Then

$$V_c = 0 - \frac{2q_cN}{4\pi\varepsilon_0} \ln \left(\frac{s_c}{s_g}\right) + C \tag{7.19}$$

To fit the large circles centered at the origin to the plate electrode let $s = s_p \gg s_g$. Then Eq. (7.18) becomes

$$V_p = -\frac{2Nq_g}{4\pi\varepsilon_0} \ln \left(\frac{s_p}{s_g}\right) - \frac{2Nq_c}{4\pi\varepsilon_0} \ln \left(\frac{s_p}{s_g}\right) + C \tag{7.20}$$

To fit one of the small circles about the point $(s_g,0)$ to one of the grid wires let $s = s_g$ and $\phi = \frac{R_g}{s_g}$, where R_g is the grid-wire radius. Then

$$V_g = -\frac{2q_g}{4\pi\varepsilon_0} \ln \left(2 \sin \frac{NR_g}{2s_g}\right) + C \tag{7.21}$$

The three equations (7.19), (7.20), and (7.21) express the electrode potentials in terms of the cathode and grid-wire charges. For the W-plane representation the charges are those of one pie-shaped section of angle $\frac{2\pi}{N}$.

As before, the amplification factor may be found by setting the cathode charge and potential zero and taking the ratio of plate to negative grid potential. From this operation

$$\mu = -\frac{N \ln \left(\frac{s_p}{s_g}\right)}{\ln \left(2 \sin \frac{NR_g}{2s_g}\right)} \tag{7.22}$$

The way in which the amplification factor of a cylindrical triode varies with the various electrode dimensions can be seen by inspection of Eq. (7.22). As the number of grid wires is increased, the amplification factor increases since N appears as a linear factor in the numerator and as a logarithmic factor in the denominator. This is to be expected from physical reasoning since an increase in the number of grid wires increases the effectiveness of the grid in controlling the off-cathode gradient and hence in controlling the current. The amplification factor increases as the plate radius is increased, also to be expected since this makes the plate less effective in controlling the current. The amplification factor increases as the radius of the grid-wire cylinder decreases since the factor in the numerator is more effective than that in the denominator. The amplification factor also increases as the grid-wire radius is increased. Because of the way in which the factors combine, high amplification factors may be obtained more readily with cylindrical-electrode structures than with plane-electrode structures.

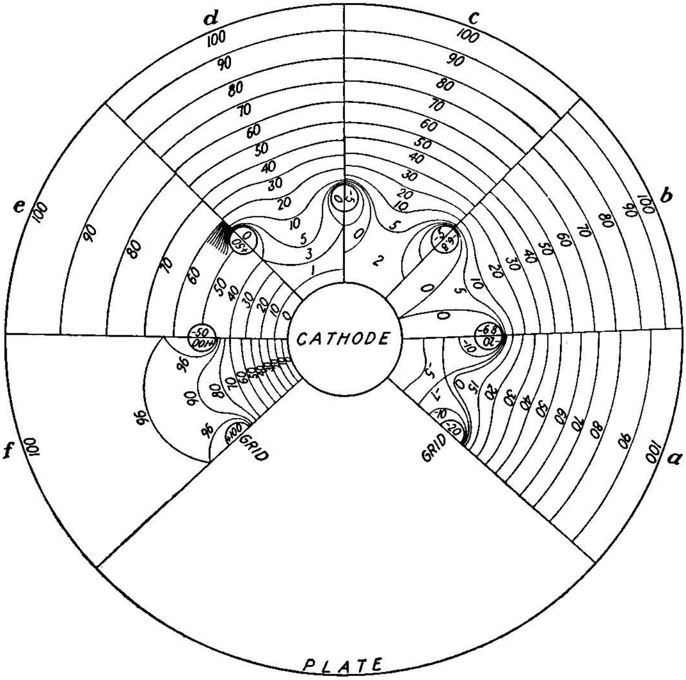


FIG. 7.6.—Equipotential contours in the cylindrical-electrode triode: (*a*) grid beyond cutoff potential; (*b*) grid at cutoff potential; (*c*) grid negative but above cutoff potential; (*d*) grid at zero potential; (*e*) grid at "natural" potential; (*f*) grid at positive plate potential.

As in the case of the parallel-electrode tube it is desirable to express the charges in terms of the electrode potentials. This is done by setting the cathode potential equal to zero and solving for q_c and q_g.

$$q_c = \frac{2\pi\varepsilon_0\left[V_g \ln\left(\frac{s_p}{s_g}\right) - \frac{V_p}{N}\ln\left(2\sin\frac{NR_g}{2s_g}\right)\right]}{N\left[\ln\left(\frac{s_p}{s_g}\right)\ln\left(\frac{s_c}{s_g}\right) + \frac{1}{N}\ln\left(\frac{s_p}{s_c}\right)\ln\left(2\sin\frac{NR_g}{2s_g}\right)\right]} \tag{7.23}$$

$$q_g = -\frac{2\pi\varepsilon_0\left[V_p \ln\left(\frac{s_c}{s_g}\right) + V_g\ln\left(\frac{s_p}{s_c}\right)\right]}{N\left[\ln\left(\frac{s_p}{s_g}\right)\ln\left(\frac{s_c}{s_g}\right) + \frac{1}{N}\ln\left(\frac{s_p}{s_c}\right)\ln\left(2\sin\frac{NR_g}{2s_g}\right)\right]} \tag{7.24}$$

Potential Contours of a Cylindrical Triode. Contour representations of potential are shown in Fig. 7.6 for various relative electrode potentials. The contours of the cylindrical triode exhibit the same general characteristics as those of the plane-electrode triode. In each case the contours near any electrode have the same shape as the electrode. This means that the contours about the cathode and just inside the plate are circles concentric about the center of the tube. There are also circles about each of the grid wires. The circles inside the plate have a nonlinear spacing in the case of the cylindrical triode. This is better understood by reference to the potential profiles.

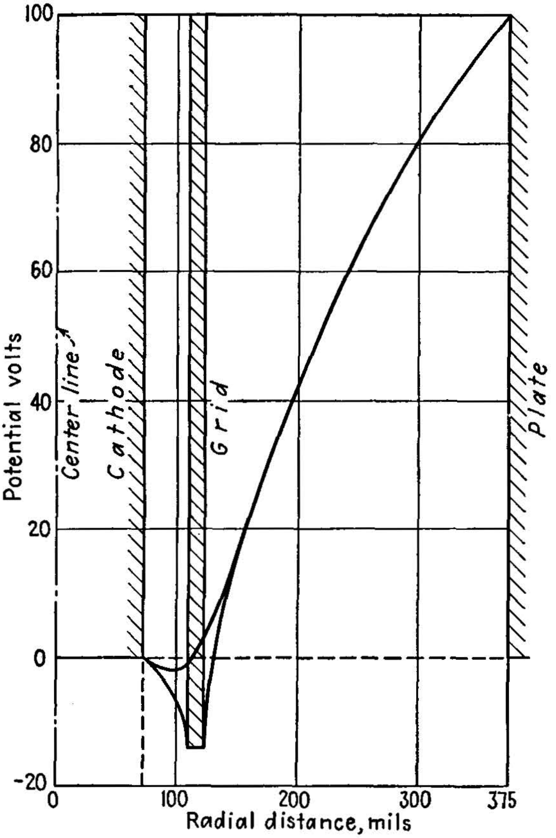


FIG. 7.7*a*.—Potential profiles in a cylindrical triode, with grid at twice the cutoff value of potential.

Potential Profiles of a Cylindrical Triode. Reference to the potential profiles of Fig. 7.7 reveals several striking differences between the plane-electrode and cylindrical-electrode cases. Although the profiles have the same general trend, they are characterized by different curvature characteristics. In the plane-electrode case the profiles through and between the grid wires coincided near the cathode and plate and were nearly straight lines there. In the case of the cylindrical triode they do again coincide but are curved instead of straight. The coincident profiles near the plate tend to be logarithmic in shape, as would be the case in a cylindrical diode. The same is true for the profiles near the cathode, though in the particular case of the relatively high-mu tube shown, the region in which the profiles coincide near the cathode is small because of the short cathode-grid distance.

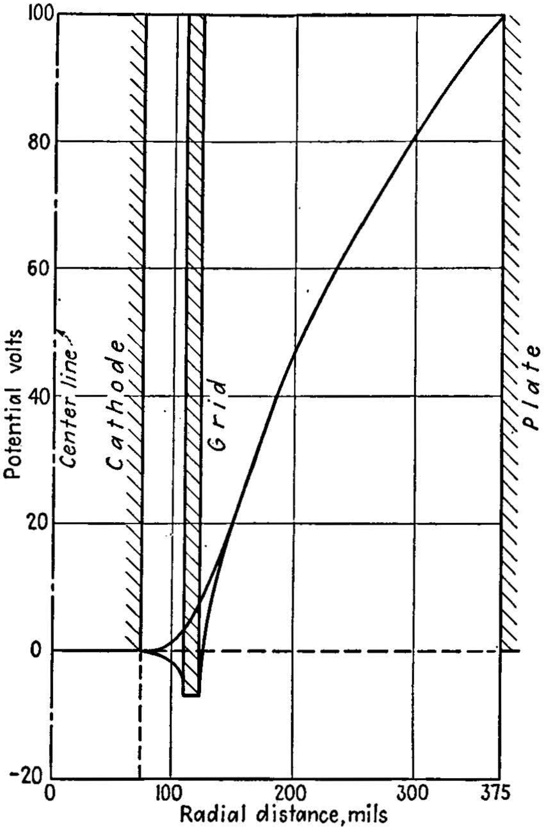


FIG. 7.7*b*.—Potential profiles in a cylindrical triode, with grid at the cutoff value of potential.

Below cutoff in Fig. 7.7*a* the cathode gradient of potential is negative. At cutoff as in Fig. 7.7*b* it is zero, and it can again be seen that this condition of zero cathode gradient is independent of the grid-cathode distance provided that the distance is not too small. In the particular figure shown any larger cathode-grid distance would not change the amplification factor of the tube. The remaining figures show the profiles for a grid negative, but above cutoff in Fig. 7.7*c*, for a grid at zero potential in Fig. 7.7*d*, and for an extreme condition of positive grid potential in Fig. 7.7*e*.

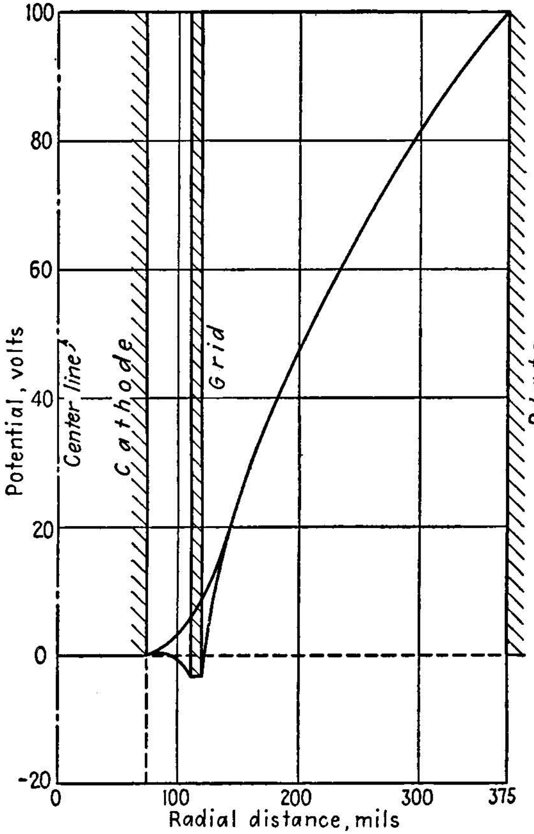


FIG. 7.7*c*.—Potential profiles in a cylindrical triode, with grid at half the cutoff value of potential.

The curvature conditions that were noted in the case of the plane-electrode triode are no longer valid in the case of the cylindrical triode. It is no longer true that if one profile is concave upward the other is concave downward. This follows from the fact that the Laplace equation for polar coordinates can no longer be interpreted so simply in terms of curvatures. If the coordinates were changed so that the profiles were plotted against the logarithm of r, then the curvature conditions that held for the plane-electrode case would be fulfilled. This follows because with this change of variables the Laplace equation becomes identical with that for Cartesian coordinates. In Fig. 7.8 are shown the potential profiles of Fig. 7.7*a* plotted with a logarithmic scale of radius. It is seen that the profiles become straight lines in the vicinity of the cathode and plate and thus resemble the plane-triode profiles in this type of plot.

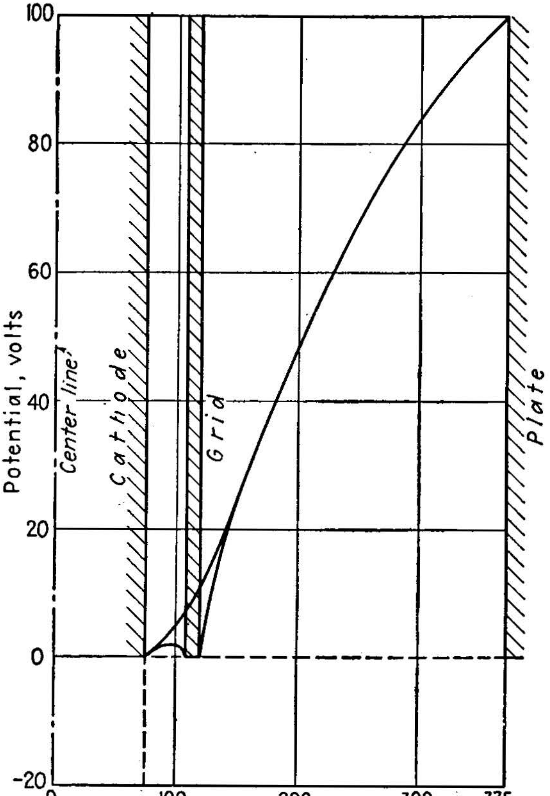


FIG. 7.7*d*.—Potential profiles in a cylindrical triode, with grid at zero potential.

7.4. Analysis of the High-mu Triode. *Potential Contours and Profiles.* The method of Maxwell discussed in the previous sections has some limitations that make the results inaccurate when the attempt is made to apply them to a high-mu triode. In the previous analysis it was assumed that the equipotentials about the grid line charge in the fundamental tube of the Z-plane representation were circles concentric about the line. This is very nearly true provided that the circles are not too large. But if we examine the shape the grid wire in the Z plane must have as determined by circles in Fig. 7.1*b* transformed back to the Z plane, it is found that departures from small circles centered about the point (1,0) are soon encountered as the grid-wire radius is increased. Consider the grid wires of Fig. 7.1*b*, and let the grid-wire radius be increased from a small value to a fairly large value without changing the other dimensions. This is equivalent to increasing the screening fraction and increases the mu of the tube. As the grid-wire radii are increased, the corresponding contours in the Z plane, which are at first small circles with centers at the point (1,0), become larger curves that are nearly circular in shape but that are shifted in position so that their centers are not at the point (1,0) but to the right of this point. This applies as the screening fraction is increased from 0.1 to 0.2. The progressive changes in shape encountered are shown in Fig. 7.9. In the analysis indicated in this section another line charge is introduced to take account of the shift in position of the circular grid-potential contour. As the screening frac-

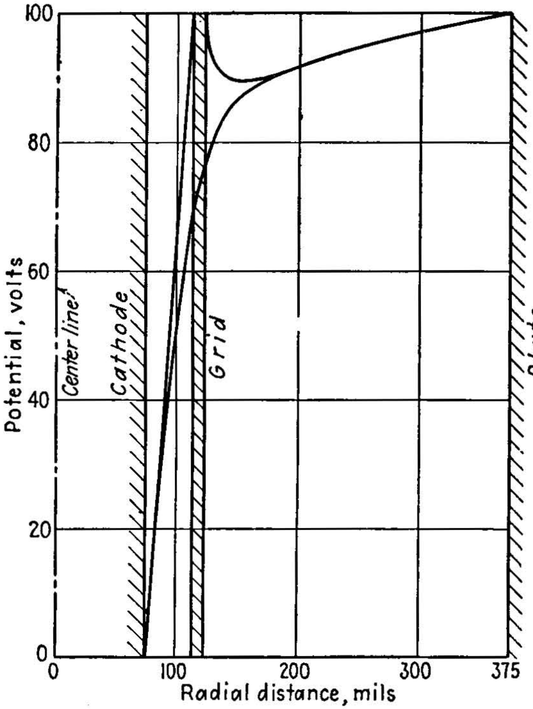


FIG. 7.7*e*.—Potential profiles in a cylindrical triode, with grid and plate at equal positive potentials.

tion is increased beyond 0.3, the transformed contour of the grid wire loses its circular shape, becoming dented on one side, and the improved analysis is no longer valid. The accuracy of the formulas developed can be extended, however, so that they may be used for tubes having screening fractions as low as $\frac{1}{6}$ instead of merely $\frac{1}{10}$. The resulting expressions are considered the most accurate simple expressions available.[1,2]

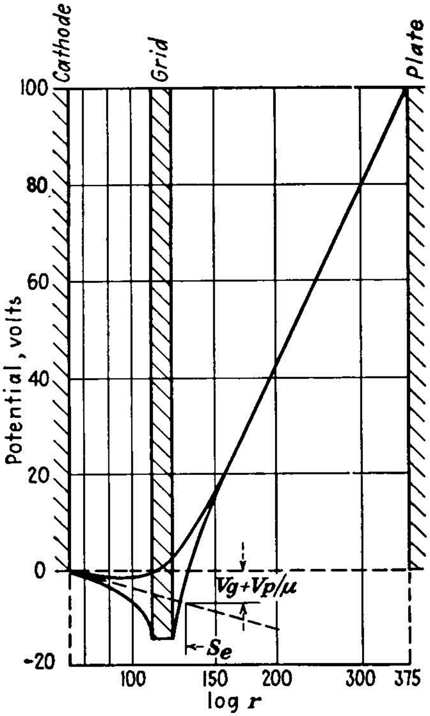


FIG. 7.8.—Potential profiles of a cylindrical triode plotted with a logarithmic scale of radius. On such a plot the profiles are straight lines in the vicinity of the cathode and plate. Also shown is the graphical construction for determining the equivalent diode radius.

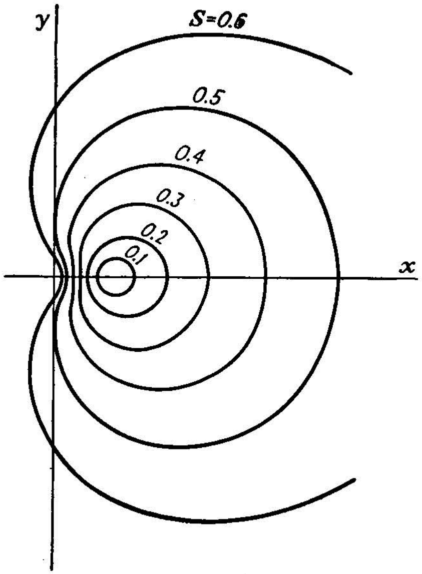


FIG. 7.9.—The shape of large transformed grid wires in the Z-plane representation of Fig. 7.1. The transformed grid wires are nearly circular for screening fractions as large as 0.2. Beyond this value, the grid contour is noncircular and cannot be represented by two line charges.

As before, use the plane-electrode transformation equations,

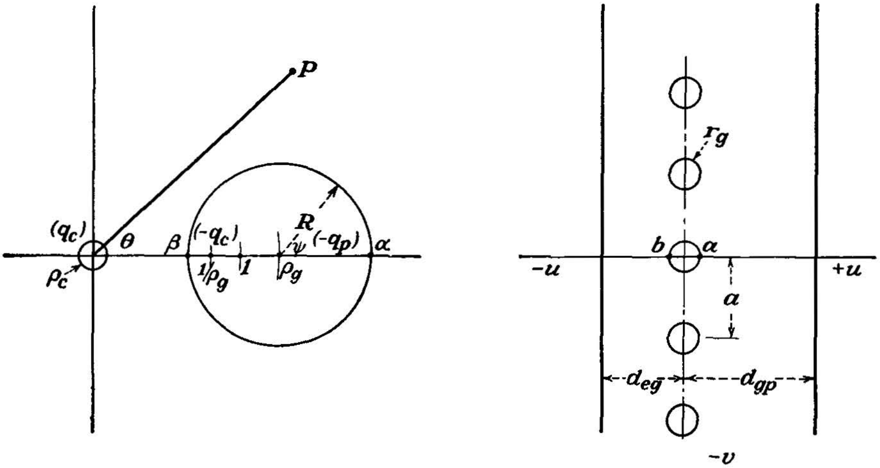
$$W = \frac{a}{2\pi} \ln Z = \frac{a}{2\pi} \ln \rho + \frac{i\theta a}{2\pi} \tag{7.3}$$

---

[1] VODGES, F. B., and F. R. ELDER, Formulas for the Amplification Constant for Three-element Tubes, *Phys. Rev.*, vol. 24, pp. 683–689, December, 1924.

[2] Dow, W. G., "Engineering Electronics," pp. 24–53, Wiley, New York, 1937.

which have the component transformation equations given by Eqs. (7.4) and (7.5), ρ being Z-plane radius. Then the points a and b in Fig. 7.10b which have the coordinates $(r_g,0)$ and $(-r_g,0)$ respectively, transform into the points α and β, which in Fig. 7.10a have the coordinates $(e^{\frac{2\pi r_g}{a}},0)$ and $(e^{-\frac{2\pi r_g}{a}},0)$. If the screening fraction is less than 0.16, the transformed grid wire is nearly a circle through these two points. The radius of the grid wire in the Z plane is half the difference of the ρ components of α and β and is given by



Z PLANE
(a)

W PLANE
(b)

FIG. 7.10.—W- and Z-plane representations of a high-mu plane-electrode triode. This is the basis of the analysis of Vodges and Elder.

$$R = \sinh \frac{2\pi r_g}{a} \tag{7.25}$$

The location of the center of the grid-wire circle is given by the average of the values of α and β and is given by

$$\rho_g = \cosh \frac{2\pi r_g}{a} \tag{7.26}$$

It is now necessary to locate line charges so that this circle will be an equipotential contour for all combinations of cathode, grid, and plate potential.

In the analysis given for the low-mu triodes a single line charge was placed within the circle at its center. When the analysis is extended so that the circle is no longer small, this is not adequate. A line charge at the center of the circle will still give an equipotential contour that will fit the circle if this charge is the only one present, which is not the case. However, it is also possible to find another position within the circle such that a line charge placed there together with a line charge at the origin with an equal charge of opposite sign will give an equipotential contour on the transformed grid-wire circle.[1] This follows from the well-known configuration of potentials about a two-wire transmission line, which is equivalent to these two line charges. Here the equipotential contours are all circles, enclosing the charges, but with their centers successively displaced.

If a line charge with a linear-charge density $-q_c$ is placed at a point $(b,0)$ within the circle having its center at $(\rho_g,0)$, then the potential at any point C on the circle due to it and to a linear charge with density q_c located at the origin is given by

$$V_c = -\frac{q_c}{4\pi\varepsilon_0} \ln \left[\frac{{\rho_g}^2 + R^2 + 2\rho R \cos \psi}{(\rho_g - b)^2 + R^2 - 2(\rho - b) R \cos \psi} \right] \tag{7.27}$$

where ψ is the angle between the line joining C and $(\rho_g,0)$ and the axis. It is required that this expression be independent of the angle ψ. It is easily shown by substitution that if b has the value $\dfrac{1}{\rho_g}$ this condition is fulfilled. The two line charges with the above positions take care of the charge on the cathode and part of the charge on the grid. If now a charge of magnitude $-q_p$ be placed at the center of the transformed grid-wire circle, it will be the source of flux lines that will extend in all directions, becoming radial at great distances and terminating on a large plate circle. Since any value of either q_c or q_p gives rise to an equipotential circle of radius R with center at ρ_g, this circle can be made the grid-wire circle for any combination of charges and hence of potentials. With this location of line charges it is easy to write the potential at any point in the tube.

At any point (ρ,θ) within the tube the potential is given by

$$V = -\frac{1}{4\pi\varepsilon_0} \left[q_c \ln \rho^2 - q_c \ln \left(\rho^2 + \frac{1}{{\rho_g}^2} - \frac{2\rho}{\rho_g} \cos \theta \right) \right.$$
$$\left. - q_p \ln (\rho^2 + {\rho_g}^2 - 2\rho\rho_g \cos \theta) + C \right] \tag{7.28}$$

in which the constant C is introduced to adjust the level of the potential. In most tubes the cathode can be fitted to a small circle about the origin.

---

[1] This attack on the problem was first successfully applied by W. G. Dow.

Let $\rho = \rho_c$, $\theta = 0$; then

$$V_c = -\frac{1}{4\pi\varepsilon_0}\left[q_c \ln \rho_c^2 - q_c \ln\left(\rho_c - \frac{1}{\rho_g}\right)^2 - q_p \ln (\rho_c - \rho_g)^2 + C\right] \tag{7.29}$$

If the cathode potential be taken as zero and the value of C then obtained from Eq. (7.29) be substituted into Eq. (7.28), the expression for the potential along the axis through the grid wire simplifies to

$$V = \frac{1}{4\pi\varepsilon_0}\left\{q_c \ln\left[\left(\frac{\rho - \dfrac{1}{\rho_g}}{\rho_c - \dfrac{1}{\rho_g}}\right)\frac{\rho_c}{\rho}\right]^2 + q_p \ln\left(\frac{\rho - \rho_g}{\rho_c - \rho_g}\right)^2\right\} \tag{7.30}$$

This has been obtained by setting θ equal to zero and gives the potential along the axis through the grid wire and also through the grid wires in the plane- and cylindrical-electrode tubes, which may be derived from the simple fundamental tube. To get the potential between the grid wires, θ is set equal to π, and the resulting expression is like Eq. (7.30) except that the negative signs within the brackets become positive.

To find the potential of the grid wire let $\rho = \rho_g - R$. Then making use of the fact that $\rho_g^2 = 1 + R^2$ from the hyperbolic relations, and that ρ_c is much smaller than ρ_g or its reciprocal,

$$V_g = \frac{1}{4\pi\varepsilon_0}\left[q_c \ln (R\rho_c)^2 + q_p \ln\left(\frac{R}{\rho_g}\right)^2\right] \tag{7.31}$$

Similarly, to find the plate potential let $\rho = \rho_p$, and make use of the fact that ρ_c is much less than 1, which in turn is much less than ρ_p. Then

$$V_p = \frac{1}{4\pi\varepsilon_0}\left[q_c \ln (\rho_c\rho_g)^2 + q_p \ln\left(\frac{\rho_p}{\rho_g}\right)^2\right] \tag{7.32}$$

Equations (7.31) and (7.32) give the grid and plate potential in terms of cathode and grid charges. In order to calculate potential profiles it is desirable to know the charges in terms of the potentials. The above equations are readily rearranged to give this form.

$$q_c = 2\pi\varepsilon_0\left[\frac{V_g \ln\left(\dfrac{\rho_p}{\rho_g}\right) - V_p \ln\left(\dfrac{R}{\rho_g}\right)}{\ln (R\rho_c) \ln\left(\dfrac{\rho_p}{\rho_g}\right) - \ln\left(\dfrac{R}{\rho_g}\right) \ln (\rho_c\rho_g)}\right] \tag{7.33}$$

$$q_p = 2\pi\varepsilon_0\left[\frac{V_p \ln (R\rho_c) - V_g \ln (\rho_c\rho_g)}{\ln (R\rho_c) \ln\left(\dfrac{\rho_p}{\rho_g}\right) - \ln\left(\dfrac{R}{\rho_g}\right) \ln (\rho_c\rho_g)}\right] \tag{7.34}$$

Amplification Factor of a High-mu Plane-electrode Triode. Again the amplification factor is given by the ratio of plate to negative grid potential for zero cathode charge. It will be noticed that the cathode charge is a linear function of plate and grid potentials. In multielectrode tubes the cathode charge is a linear function of all the electrode potentials. In the case of a pentode, for instance, the cathode charge can be written

$$q_c = \frac{V_g + \dfrac{V_1}{\mu_1} + \dfrac{V_2}{\mu_2} + \dfrac{V_3}{\mu_3}}{D} \tag{7.35}$$

in which the various μ's indicate the relative effectiveness of the control grid and the electrode in question in controlling the off-cathode potential gradient. In the case of the triode considered here, the amplification factor is

$$\mu = -\frac{\ln\left(\dfrac{\rho_p}{\rho_g}\right)}{\ln\left(\dfrac{R}{\rho_g}\right)} \tag{7.36}$$

Substitution of values from Eqs. (7.3), (7.25), and (7.26) gives the amplification factor of a plane-electrode triode as

$$\mu = \frac{\left(\dfrac{2\pi d_{gp}}{a}\right) - \ln\cosh\left(\dfrac{2\pi r_g}{a}\right)}{\ln\coth\left(\dfrac{2\pi r_g}{a}\right)} \tag{7.37}$$

Making use of the definition of the screening fraction as the percentage of the area in the grid-wire plane occupied by the grid wires, numerically equal to $\frac{2r_g}{a}$, and denoting the screening fraction by S, the expression for the amplification factor can also be written

$$\mu = \frac{\left(\dfrac{2\pi d_{gp}}{a}\right) - \ln\cosh\pi S}{\ln\coth\pi S} \tag{7.38}$$

From this it is seen that the amplification factor depends upon only two factors, the screening fraction and the ratio of grid-plate distance to grid-wire spacing. The way in which the amplification factor varies with these two factors is shown in Fig. 7.11. It is seen that the amplification factor increases with the screening fraction and increases as the ratio of grid-plate to grid-wire spacing increases. The upper solid curve represents the limit of accuracy of the formula given by Eq. (7.38).

If the screening fraction is small, then Eqs. (7.37) and (7.38) reduce to the same expression that results from Eq. (7.13) so that these two different expressions give substantially the same numerical result when the screening fraction is less than $\frac{1}{10}$.

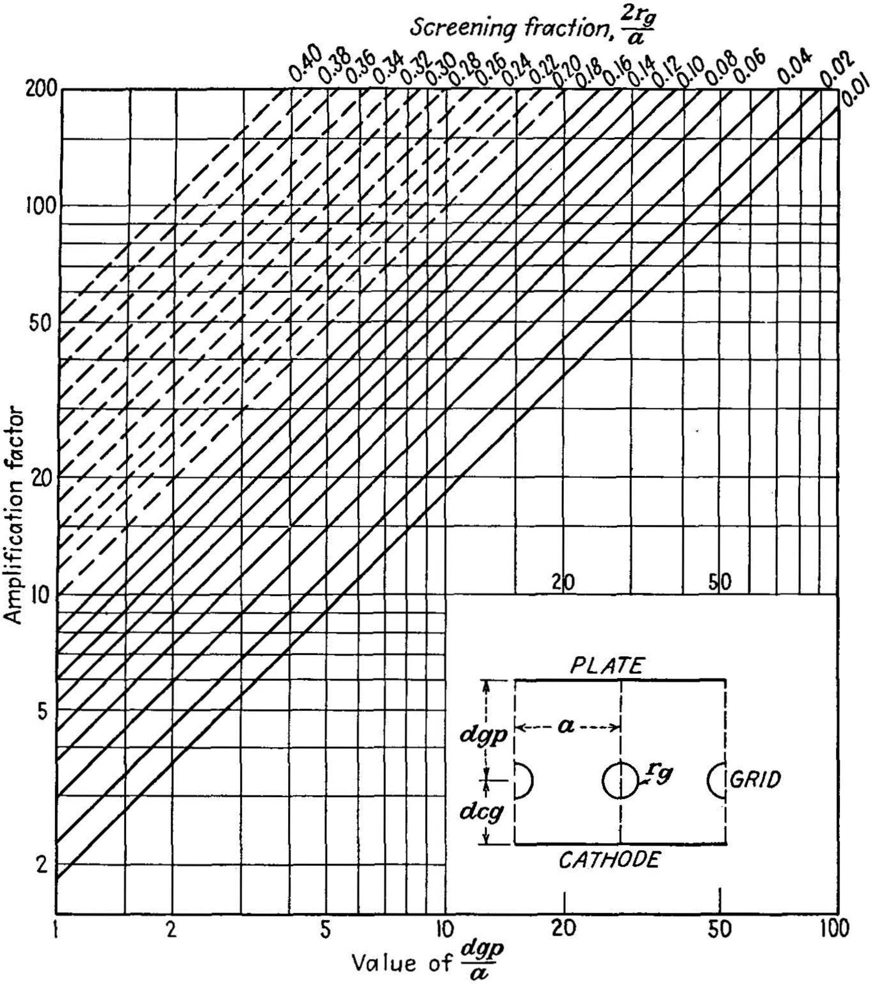


FIG. 7.11.—Chart giving the amplification factor of a plane-electrode triode. The solid curves were obtained from Eq. (7.38), which is accurate up to screening fractions of 0.16. Dotted curves were obtained from Eq. (7.71), which is accurate up to screening fractions of 0.4.

Amplification Factor of a High-mu Cylindrical Triode. When the transformation relations of Eqs. (7.16) and (7.17) are applied to Eq. (7.36), the expression for the amplification factor of a cylindrical triode is obtained. This development requires some intermediate justification because for the cylindrical triode the transformation is that of Eq. (7.15) instead of Eq. (7.3). It is readily shown, however, that expressions similar to Eqs. (7.25) and (7.26) are obtained.

In the notation of Figs. 7.10*a* and 7.5*b* the points α and β are given by $\left(1 + \frac{R_g}{s_g}\right)^N$ and $\left(1 - \frac{R_g}{s_g}\right)^N$. Usually the factor $\frac{R_g}{s_g}$ is much smaller than 1 so that series expansions for these expressions can be used to simplify the development. Great care must be used in approximating these expressions by terms of the series expansions; for the difference between the expressions is desired, and two terms of the series are not sufficient. The expressions for α and β are given very closely by exp $\left(\frac{NR_g}{s_g}\right)$ and exp $\left(\frac{-NR_g}{s_g}\right)$. The series expansions for these exponential terms are identical with those for the binomials given above for the first two terms and differ only by a factor of $1 - \frac{1}{N}$ in the third term. Using the difference of the exponential terms to get the radius of the transformed grid wire,

$$R = \sinh \frac{NR_g}{s_g} \tag{7.39}$$

and, using the average to get the location of the grid-wire center,

$$\rho_g = \cosh \frac{NR_g}{s_g} \tag{7.40}$$

It is seen that Eqs. (7.39) and (7.40) are the exact counterparts of Eqs. (7.25) and Eq. (7.26).

If now Eqs. (7.39), (7.40), and (7.15) are substituted into Eq. (7.36), the expression for the amplification factor of the cylindrical triode results.

$$\mu = \frac{N \ln \left(\frac{s_p}{s_g}\right) - \ln \cosh \left(\frac{NR_g}{s_g}\right)}{\ln \coth \left(\frac{NR_g}{s_g}\right)} \tag{7.41}$$

Since the screening fraction for the cylindrical triode is $S = \frac{NR_g}{\pi s_g}$, the above expression may be written

$$\mu = \frac{N \ln\left(\frac{s_p}{s_g}\right) - \ln \cosh \pi S}{\ln \coth \pi S} \tag{7.42}$$

The amplification factor is seen to depend upon three factors, the screening fraction, the ratio of plate to grid-wire circle radii, and the number of grid wires. This is much more difficult to plot but may be done by using a number of axes each corresponding to a different number of grid wires. Such a plot of the amplification factor of the cylindrical triode is shown in Fig. 7.12. It is seen that the amplification factor increases with the screening fraction, with the number of grid wires, and with the ratio $\frac{s_p}{s_g}$. The upper solid curve represents the limit of an accuracy of about 2 per cent. The formulas of Vodges and Elder given

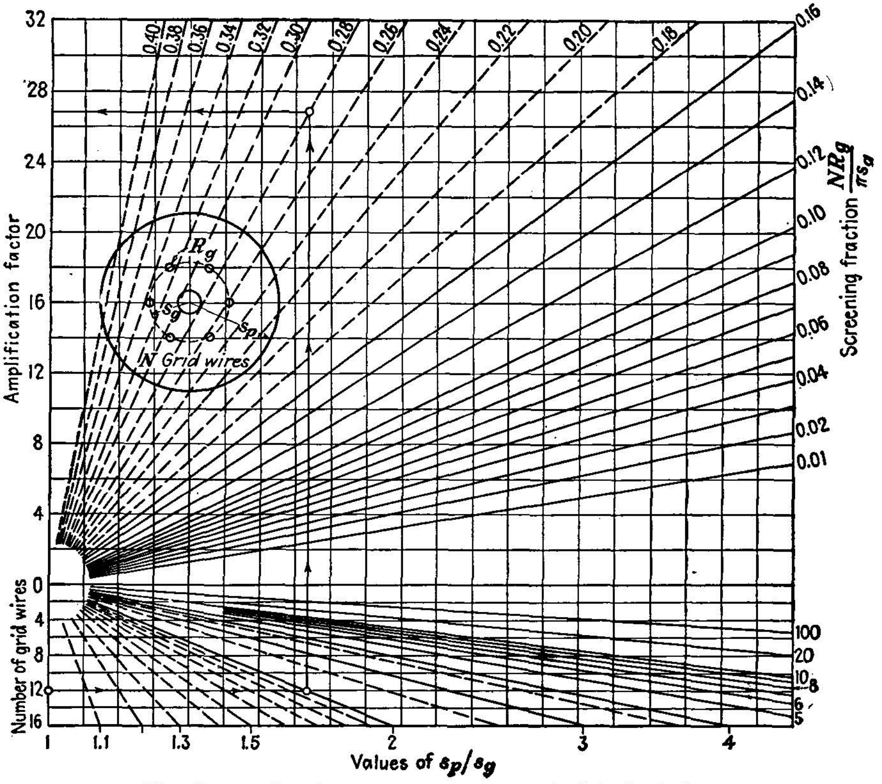


FIG. 7.12.—Amplification factor of a cylindrical triode.

in Eqs. (7.38) and (7.42) are considered the best practical formulas available for the amplification factor of triodes from the standpoint of simplicity and accuracy.

7.5. The Equivalent Electrostatic Circuit of a Triode. Examination of Eqs. (7.33) and (7.34) shows that these expressions have the form of equations for a delta of capacities. A delta of capacities such as could be used to represent a triode is shown in Fig. 7.13. If the cathode is considered to be at zero potential, then the relations between the potentials and charges are

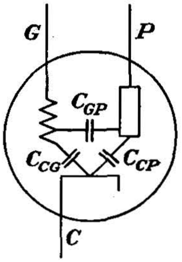


FIG. 7.13.—Delta of triode interelectrode capacities.

$$-q_c = V_gC_{cg} + V_pC_{cp} \tag{7.43}$$

and

$$q_p = (V_p - V_g)C_{gp} + V_pC_{cp} \tag{7.44}$$

which can be arranged into the simpler form

$$q_p = -V_gC_{gp} + V_p(C_{gp} + C_{cp}) \tag{7.45}$$

Equations (7.43) and (7.45) are the exact counterparts of Eqs. (7.33) and (7.34). Equation (7.43) can be rearranged to give

$$-q_c = C_{cg}\left(V_g + \frac{V_p}{\dfrac{C_{cg}}{C_{cp}}}\right) \tag{7.46}$$

from which, by comparison with Eq. (7.35), it is evident that the amplification factor is given by the ratio of the grid-cathode to plate-cathode capacity.

$$\mu = \frac{C_{cg}}{C_{cp}} \tag{7.47}$$

This is physically reasonable since the ratio of these two capacities is a measure of the extent to which the cathode is electrostatically shielded from the plate by the grid. The capacities involved can be evaluated by reference to Eqs. (7.33) and (7.34) if it is desired to know them in terms of the geometry of the tube. The capacities in the above expressions are in farads per unit length (meter) per grid-wire section of the elementary tube in the Z plane of Fig. 7.10. When the dimensions are transformed to other planes, the capacities of the corresponding tubes result.

Reference to any tube manual will show that the numerical ratio of the grid-cathode to grid-plate capacities listed there differs considerably from the amplification factor of the tube. This is because the capacities listed in the manual include the capacities of the leads and supports as well as those of the parts of the tube in which the electrons are effective. In most triodes the capacities between the leads and supports may be as large as those of the active portions of the tube or larger, the apparent discrepancy being thus accounted for.

7.6. Equivalent-diode Spacing of a Triode. In the case of the plane-electrode triode, examination of the potential profiles showed that the profiles were straight lines in the vicinity of the cathode for all conditions of potential. This means that the potential distribution in a triode as seen from the cathode is the same as that in a diode. The cathode has no data by which it can tell whether it is part of a diode or triode. It was further shown in Eq. (7.14*a*) that the off-cathode gradient of potential depended upon an equivalent voltage, $V_g + \frac{V_p}{\mu}$. It would be expected, therefore, that for every triode there would exist an equivalent diode which would have the same off-cathode gradient when the equivalent-triode voltage is applied to its plate. To find such an equivalent diode it is necessary only to find the equivalent-diode spacing.

The equivalent-diode spacing of any triode may be found graphically by extending the straight-line portion of the potential profile in the vicinity of the cathode until it reaches a potential equal to the equivalent-triode potential, $V_g + \frac{V_p}{\mu}$. The distance from the cathode at which this potential is reached is the equivalent-diode spacing. This construction is shown in Fig. 7.4*c*. Once this equivalent-diode spacing is found it can be used for all combinations of plate and grid potential.

The concept of the equivalent diode and the equivalent-diode spacing is useful in the study of the current characteristics of triodes. Since the current law for space-charge-limited diodes is known, the current law for a triode can be approximated from the equivalent diode. Triodes with equal equivalent-diode spacings and equal amplification factors have *approximately* the same mutual conductance and plate resistance. Actually, the triode and its equivalent diode are truly equivalent only for a condition of no current flow, which means cutoff or beyond for the triode since the flow of current changes the potential distribution. The concept is, however, sufficiently useful to justify its inclusion here. The subject of current flow and mutual conductance will be discussed in the chapter on Space-charge Effects.

Diode Equivalent to a Plane-electrode Triode. An analytical expression for the equivalent-diode spacing can be found from the expression for cathode charge in terms of the equivalent-triode potential and the geometry of the tube. Since the cathode charge was taken as $+q_c$ per grid-wire section, the off-cathode gradient of potential is given by

$$\left(\frac{dV}{dx}\right)_c = \frac{-q_c}{a\varepsilon_0} \tag{7.48}$$

or

$$\left(\frac{dV}{dx}\right)_c = -\left[\frac{V_g + \left(\frac{V_p}{\mu}\right)}{D}\right]\frac{1}{a\varepsilon_0} \tag{7.49}$$

in which all the symbols have the previous significance and D is the denominator of Eq. (7.33) or (7.35) when the equivalent-voltage factor is extracted.

For a plane-electrode diode the electrode spacing is the ratio of the potential difference to the gradient. For the triode the equivalent-diode spacing is the ratio of the equivalent-triode potential to the off-cathode gradient. Thus, from Eq. (7.33)

$$d_e = \frac{a}{2\pi}\left[\frac{\ln\,(\rho_c\rho_g)\ln\left(\frac{R}{\rho_g}\right)}{\ln\,(\rho_p/\rho_g)} - \ln\,(\rho_c R)\right] \tag{7.50}$$

where d_e is the equivalent-diode spacing in terms of the Z-plane dimensions. It is desirable to express d_e in terms of the cathode-grid distance, the grid-plate distance, and the amplification factor. If the expression for the amplification factor [Eq. (7.36)] is used to eliminate R, then

$$d_e = \frac{-a}{2\pi}\left[\left(\frac{1}{\mu}+1\right)\ln\,(\rho_c\rho_g) - \frac{1}{\mu}\ln\left(\frac{\rho_p}{\rho_g}\right)\right] \tag{7.51}$$

If now the plane-electrode transformation is applied to insert the triode dimensions,

$$d_e = d_{cg} + \frac{d_{gp}+d_{cg}}{\mu} - \frac{a}{2\pi}\left(1+\frac{2}{\mu}\right)\ln\cosh\pi S \tag{7.52}$$

For tubes in which the screening fraction is less than $\frac{1}{10}$, the last term of Eq. (7.52) is negligible so that the expression reduces to

$$d_e \cong d_{cg}\left[1 + \frac{d_{cg}+d_{gp}}{\mu d_{cg}}\right] \tag{7.53}$$

Examination of Eq. (7.53) shows that the plate of the equivalent diode always lies beyond the actual grid of the triode. The distance beyond is relatively less if the amplification factor and cathode-grid distance are large and relatively more if the grid-plate distance is large. The mutual conductance of a tube is an inverse function of the equivalent-diode spacing so that the influence of the various tube dimensions is readily determined from Eq. (7.53).

Diode Equivalent to a Cylindrical-electrode Triode. The procedure that was used to find the equivalent-diode spacing of a plane-electrode triode is quite general and may be applied to cylindrical triodes as well, though the form of the resulting expressions is quite different, as may be expected from the fact that the potential profile for the plane-electrode diode is a straight line, while for the cylindrical diode it is a logarithmic curve. For a cylindrical diode the relation between cathode and plate radius, potential difference, and electrode charge per unit length of outer electrode is

$$\ln\left(\frac{r_p}{r_c}\right) = \frac{2\pi\varepsilon_0 V}{q} \tag{7.54}$$

For the cylindrical triode the equivalent-diode radius is given by

$$\ln\left(\frac{s_e}{s_c}\right) = -\frac{2\pi\varepsilon_0}{q_c N}\left(V_g + \frac{V_p}{\mu}\right) \tag{7.55}$$

where s_e is the equivalent-diode radius and q_c is the cathode charge per unit length for each of N grid-wire sections. Substitution of the value of q_c from Eq. (7.33) and using Eq. (7.36) gives

$$\ln\left(\frac{s_e}{s_c}\right) = -\left[\frac{\ln\,(\rho_c\rho_g)}{\mu} + \ln\,(\rho_c R)\right]\frac{1}{N} \tag{7.56}$$

Eliminating R by means of Eq. (7.39) and making the substitutions of Eqs. (7.16), (7.40), and Eq. (7.42),

$$\ln\left(\frac{s_e}{s_c}\right) = \left[\frac{1}{\mu}\left(\ln\frac{s_p}{s_g} - \frac{1}{N}\ln\cosh\pi S\right) - \left(1+\frac{1}{\mu}\right)\left(\ln\frac{s_c}{s_g} + \frac{1}{N}\ln\cosh\pi S\right)\right] \tag{7.57}$$

For tubes with screening fractions less than 1/10 the terms involving the hyperbolic cosine can be neglected so that the expression takes the form

$$\ln\left(\frac{s_e}{s_c}\right) = \ln\left(\frac{s_g}{s_c}\right)\left[1 + \frac{1}{\mu} + \frac{1}{\mu}\,\frac{\ln\left(\frac{s_p}{s_g}\right)}{\ln\left(\frac{s_g}{s_c}\right)}\right] \tag{7.58}$$

The same remarks that applied to the diode spacing for plane electrodes apply here. The equivalent-diode plate lies outside of the grid-wire circle. The equivalent-diode radius for the cylindrical triode is readily obtained by graphical construction. If the profiles of Fig. 7.7 are plotted against a logarithmic abscissa as in Fig. 7.8, then the profiles are straight lines in the vicinity of the cathode and plate. The equivalent-diode radius can be found by projecting the straight-line portion of the potential profile at the cathode until it reaches a potential equal to the equivalent-grid potential.

7.7. Application of Amplification-factor Formulas to Actual Triodes. The amplification-factor formulas previously developed have been derived for simple idealized structures not always encountered in actual tubes. It is, however, possible in most cases to interpret these formulas so that they will apply to tubes whose structure departs somewhat from that for which the formulas were derived.

The formula for the amplification factor of the cylindrical triode [Eq. (7.42)] was given for a grid in which the wires had a squirrel-cage structure of evenly spaced wires parallel to the axis of the tube. In this expression the quantity N is the number of grid wires and is also the active length of grid wire per unit axial length of the tube. The expression may therefore be generalized by letting

$$N = L_g \tag{7.59}$$

where L_g is the active length of grid wire per unit axial length of the tube.[1]

In case the grid structure differs from that postulated in the derivation of the amplification-factor formula, the screening fraction S may always be interpreted as the ratio of the actual area of the grid structure to the total area of the surface containing the grid.

If the cylindrical grid consists of a square mesh of fine wires, then

$$L_g = 4\pi \frac{s_g}{d} \tag{7.60}$$

where d is the spacing of the square mesh and s_g is the radius of the grid-wire circle. If the diameter of the wires in the square mesh is appreciable, then

$$L_g = \frac{4\pi s_g}{d}\left(1 - \frac{r_g}{d}\right) \tag{7.61}$$

where r_g is the radius of the grid wires as shown in Fig. 7.14*a*. For a cylindrical grid of parallel rings having supports parallel to the axis,

$$L_g = \frac{2\pi s_g}{s} + N_s - \frac{2N_s r_s}{s} \tag{7.62}$$

where N_s is the number of supports and r_s is the radius of the support wires and s is the spacing of the grid wires as shown in Fig. 7.14*b*.

If the grid is a helix of diameter $2s_g$ and of pitch d as in Fig. 7.14*c*,

---

[1] KUSUNOSE, Y., Design of Triodes, *Proc. I.R.E.*, vol. 17, pp. 1706–1749, October, 1929.

$$L_g = \sqrt{1 + \left(\frac{2\pi s_g}{d}\right)^2} \tag{7.63}$$

The screening fractions for the cases listed above are readily evaluated. In general if all the wires have the same radius,

$$\pi S = L_g \frac{r_g}{s_g} \tag{7.64}$$

For parallel grid rings with supports,

$$\pi S = \frac{2r_g(L_g - N_s) + 2N_s r_s}{2s_g} \tag{7.65}$$

in which the symbols have the same significance as in Eq. (7.62). Equation (7.64) also holds for a helical grid. Equation (7.65) reduces to Eq. (7.64) for $r_g = r_s$.

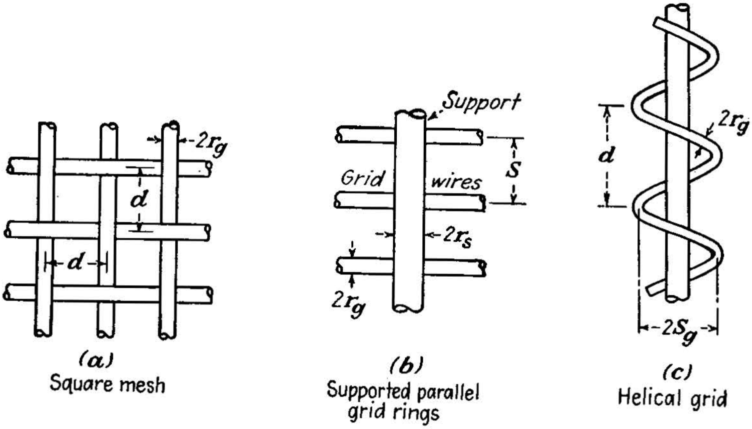


FIG. 7.14.—Practical grid structures: (*a*) square mesh; (*b*) grid rings with supports parallel to the axis; (*c*) helical grid.

The generalized amplification-factor formula for the cylindrical-electrode triode is

$$\mu_c = \frac{L_g \ln \frac{s_p}{s_g} - \ln \cosh \pi S}{\ln \coth \pi S} \tag{7.66}$$

The generalized amplification-factor formula for the plane-electrode triode is

$$\mu_p = \frac{2\pi d_{gp} L_g' - \ln \cosh \pi S}{\ln \coth \pi S} \tag{7.67}$$

The relation of this to Eq. (7.38) is evident. L_g' is the length of grid wire per unit area of the grid plane. For parallel grid wires, $L_g' = \frac{1}{a}$.

In many tubes the structure is neither plane nor cylindrical but something intermediate. In such cases it has been found empirically that a combination of the plane and cylindrical amplification-factor formulas gives very nearly the amplification factor of the tube.[1] The combination formula is

$$\mu = \mu_p - K(\mu_p - \mu_c) \tag{7.68}$$

where μ_p is the amplification factor as calculated by the plane-electrode formula, μ_c is the amplification factor as calculated by the cylindrical-electrode formula, K is a constant depending upon the tube structure, and μ is the amplification factor of the actual tube. The constant K for a number of tube types ranging progressively from a plane to a cylindrical structure is given by the following table:

TABLE IV

CONSTANT OF EQ. (7.68) FOR THE CALCULATION OF THE AMPLIFICATION FACTOR OF TUBES OF NONIDEAL FORM

Tube Type	K
....	0.00 plane electrode
2A3	0.11
....	0.22
....	0.33
26	0.44
76	0.55
75	0.66
6K5	0.77
6B5	0.88
100TL	1.00 cylindrical electrode

The results of Table IV are shown graphically in Fig. 7.15. The empirical constant K given here includes the effect of the grid supports.

7.8. More Accurate Amplification-factor Formulas. The amplification-factor formula of Vodges and Elder given in Eq. (7.38) is accurate only for certain ranges of electrode dimensions. Specifically, the formula breaks down if

$$(a)\ d_{gp} < a$$
$$(b)\ S > 0.16$$
$$(c)\ d_{cg} < a$$

In general, the formula breaks down if any of the electrodes are too close together. Since many modern tubes are built with very close-

---

[1] JERVIS, E. R., Amplification Factor Chart, *Electronics*, vol. 12, p. 45, June, 1939.

spaced electrodes, it is desirable to extend the range of the above relations. Various formulas have been worked out that extend the range of any of the three limitations listed above, but as yet no formula has appeared that is valid over the complete range of all variables.

Formula for Small Grid-plate Spacings. When the grid-plate spacing is small, an improved amplification-factor formula may be worked out by placing an image set of grids outside of the plate position and then fitting the plate to the equipotential curve midway between.[1] The resulting expression is

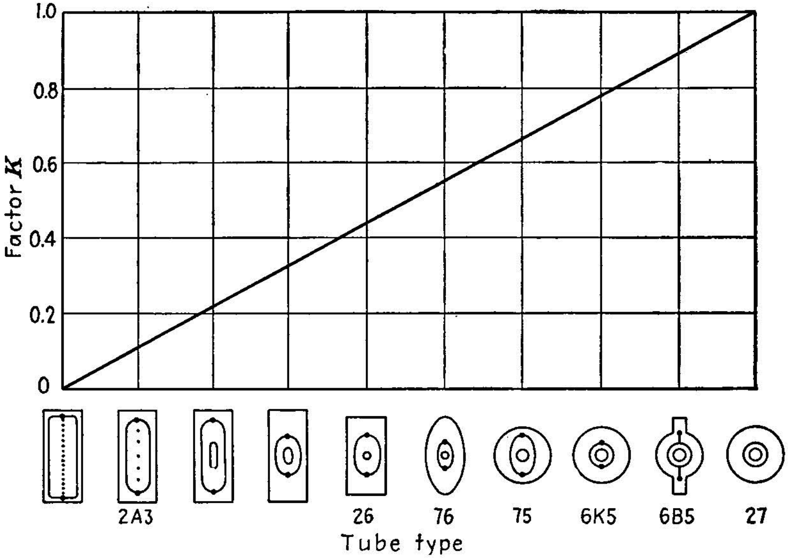


FIG. 7.15.—Value of the constant K of Eq. (7.68) for calculating the amplification factor of tubes with geometries that are intermediate between plane and cylindrical geometries.

$$\mu = \frac{\ln \cosh \pi S - 2\pi \dfrac{d_{gp}}{a}}{\ln \tanh \pi S - \ln \left[1 - \cosh^2 \pi S \exp\left(\dfrac{-4\pi d_{gp}}{a}\right)\right]} \tag{7.69}$$

This expression is valid for grid-plate spacings as low as one-quarter of the spacing between grid wires but still assumes that the cathode-grid spacing and the screening fraction are large.

---

[1] SALZBERG, BERNARD, Formulas for the Amplification Factor of Triodes, *Proc. I.R.E.*, vol. 30, pp. 134–138, March, 1942.

Formulas for Large Screening Fraction. Perhaps the most serious limitation to the amplification formula of Vodges and Elder is that it begins to be in error for a screening fraction of $\frac{1}{6}$ and is 10 per cent low for a screening fraction of $\frac{1}{3}$. Many modern tubes are built with very large screening fractions, and it therefore is desirable to have a formula valid in this range.

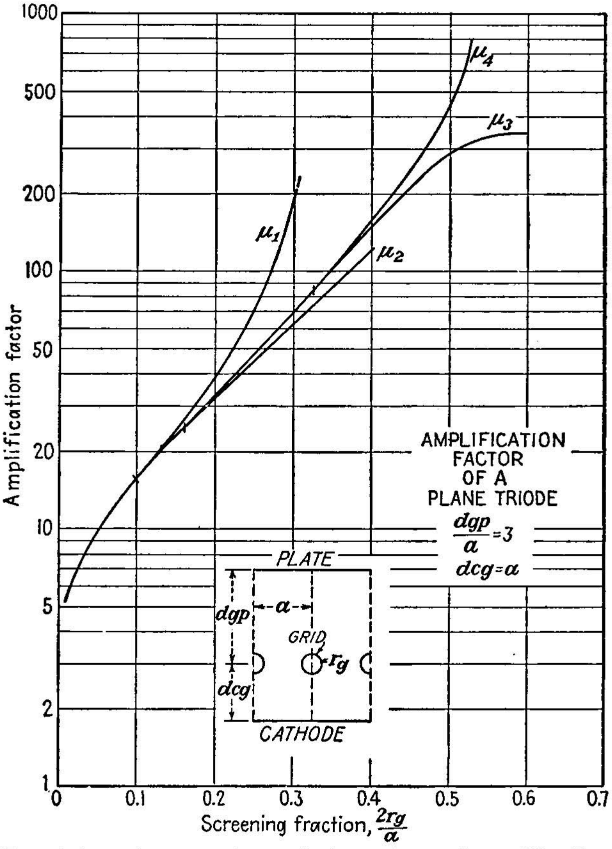


FIG. 7.16.—A comparison of the values of amplification factor of a plane-electrode triode as a function of screening fraction as indicated by various formulas. The extension of the region of validity by the successive refinements in formulas is evident. (See discussion on page 161.)

Such a formula has been evaluated[1,2] by an analysis based upon the

---

[1] OLLENDORF, FRANZ, Berechnung des Durchgriffes durch enge Steggitter, *Elektrotech. u. Maschinenbau*, vol. 52, pp. 585–591, Dec. 16, 1934.

[2] See also HERNE, H., Valve Amplification Factors, *Wireless Engr.*, vol. 21, pp. 59–64, February, 1944.

expression $W = \ln \sin Z$, which gives the potential due to a row of evenly spaced line charges positioned along a straight line. Derivatives of this expression give the potential fields due to a row of dipoles, quadripoles, and so on. By combining in series form such a succession of fields and fitting the resultant field to the circular grid wires, some highly accurate formulas for amplification factor are obtained.

The first approximation using the expression for the row of line charges alone gives a formula which is virtually the same as that of Eq. (7.13) based upon Maxwell's grating theory. The *second approximation* is obtained by using the field due to a row of line charges and a row of dipole line charges. This gives

$$\mu = \frac{\dfrac{2\pi d_{gp}}{a} - \dfrac{\frac{1}{2}(\pi S)^2}{1 + \frac{1}{12}(\pi S)^2}}{-\ln(\pi S) + \dfrac{\frac{1}{4}(\pi S)^2}{1 + \frac{1}{12}(\pi S)^2}} \tag{7.70}$$

A *third approximation* may be obtained by including the field corresponding to the next derivative of the field of a row of line charges. This yields

$$\mu = \frac{\dfrac{2\pi d_{gp}}{a} - \dfrac{\frac{1}{2}(\pi S)^2}{1 + \frac{1}{12}(\pi S)^2}}{-\ln(\pi S) + \dfrac{\frac{1}{4}(\pi S)^2}{1 + \frac{1}{12}(\pi S)^2} - \dfrac{\frac{1}{288}(\pi S)^4}{1 + \frac{1}{240}(\pi S)^4}} \tag{7.71}$$

Equations (7.70) and (7.71) apply to plane-electrode triodes. The corresponding expressions for cylindrical electrodes may be obtained by substituting $s_g \ln\left(\dfrac{s_p}{s_g}\right)$ for d_{gp} provided that the grid-wire radius is small compared with the radius of the grid-wire circle.

A comparison of the values of amplification factor given by the various formulas presented is shown in Fig. 7.16. In this figure the notation is as follows:

μ_1, Eq. (7.13), Maxwell.
μ_2, Eq. (7.38), Vodges and Elder.
μ_3, Eq. (7.70), Ollendorf second approximation.
μ_4, Eq. (7.71) Ollendorf third approximation.

From Fig. 7.16 it is seen that the range of validity of the various formulas within a 2 per cent error is

μ_1, Eq. (7.13), S less than 0.1.
μ_2, Eq. (7.38), S less than 0.16.
μ_3, Eq. (7.70), S less than 0.325.
μ_4, Eq. (7.71), S less than 0.4.

In the chart of Fig. 7.11 giving the amplification factor of a plane triode, the solid lines were obtained from the formula of Vodges and Elder, while the dotted lines were obtained from the Ollendorf approximations.

Formula for Small Cathode-grid Spacings. All the amplification-factor formulas previously given are restricted to electrode configurations in which the cathode-grid spacing is equal to or larger than the grid-wire spacing. Many modern tubes have grid-cathode spacings that are less than the grid-wire spacing. When this is true, it is really no longer proper to speak of the amplification factor, for the gradient of potential is not constant along the cathode but varies with position relative to the grid wires. Such a tube exhibits no true cutoff condition since as the grid is made more and more negative the cathode gradient opposite the grid wires will become negative while the cathode gradient between the wires is yet positive. This gives rise to a condition, sometimes referred to as *Insel Bildung,* in which little island strips of the cathode are emitting while other parts are not. Such a tube acts as a variable-mu tube in that every part of the cathode surface has a different amplification factor.

It is possible to find the effective amplification factor of a tube with small cathode-grid spacing if this effective amplification factor is understood to be dependent upon position on the cathode. When the cathode gradient is not uniform, as is the case with small cathode-grid spacings, the field's configuration can still be found if use is made of the theory of images.[1] The true field that exists is the same that would result from a line of grid wires of one charge and an image line of wires located as though mirrored in the cathode but having charge of opposite sign. The field can therefore be obtained by studying the potential within a fictitious tube consisting of two parallel plates of opposite charge and potential, between which there are two sets of identical grids of opposite charge and potential, symmetrically disposed with respect to the center plane. Under the conditions stated the mid-plane will be a surface of zero potential and can be identified as the cathode.

The potential due to a grid of equally charged parallel wires, equally spaced, a distance a along the y axis of the Z plane is

$$V(x,y) = -\frac{q}{4\pi\varepsilon_0} \ln \left[2 \left(\cosh \frac{2\pi x}{a} - \cos \frac{2\pi y}{a} \right) \right] + C \qquad (7.72)$$

where q is the charge per unit length of each grid wire. This expression may be obtained by considering the potential due to a line charge of density q located at the point (1,0) and one of density $-\frac{q}{2}$ located at the

---

[1] FREMLIN, J. H., Calculation of Triode Constants, *Phil. Mag.*, vol. 27, pp. 709–741, June, 1939. Also published in *Elec. Commun.*, July, 1939.

point (0,0) in the Z plane and then transforming this potential to the W plane by means of the transformation $W = \frac{a}{2\pi} \ln Z$. Equation (7.72) can also be shown to be the real part of the expression $\ln \left(\sin \frac{i\pi Z}{a} \right)$. A plot of Eq. (7.72) is given in Fig. 7.17.

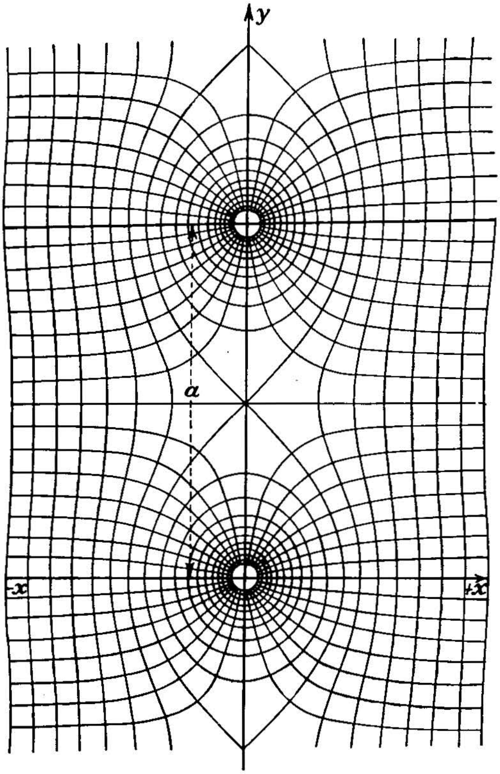


FIG. 7.17.—The potential field in the vicinity of a row of line charges as given by Eq. (7.72).

Upon using Eq. (7.72) to obtain the field due to grids of opposite charge located a distance d_{cg} from the cathode and adding a linear component of field to account for the effect of the plate the expression for the potential within the tube is obtained. It is

$$V = + \frac{q}{4\pi\varepsilon_0} \ln \left[\frac{\cosh \frac{2\pi}{a} (x + d_{cg}) - \cos \frac{2\pi}{a} y}{\cosh \frac{2\pi}{a} (x - d_{cg}) - \cos \frac{2\pi}{a} y} \right] + Bx \qquad (7.73)$$

in which q and B are related to the electrode potentials by

$$V_g = +\frac{q}{4\pi\varepsilon_0}\ln\left[1+\frac{\sinh^2\left(\frac{2\pi d_{cg}}{a}\right)}{\sin^2\left(\frac{\pi r_g}{a}\right)}\right] + Bd_{cg} \tag{7.74}$$

and

$$V_p = \frac{d_{cg}q}{a\epsilon_0} + Bd_{cp}, \tag{7.75}$$

If $-E_c(0,y)$ is the gradient of potential at the cathode—a quantity that varies with y—then the effective amplification factor may be defined as

$$\frac{1}{\mu} = +\frac{\dfrac{\partial E_c}{\partial V_p}}{\dfrac{\partial E_c}{\partial V_g}} \tag{7.76}$$

The resultant expression for amplification factor is given by

$$\frac{1}{\mu} = \frac{\dfrac{a}{4\pi d_{cp}}\ln\left[1+\dfrac{\sinh^2\left(\frac{2\pi d_{cg}}{a}\right)}{\sin^2\left(\frac{\pi r_g}{a}\right)}\right] - \dfrac{d_{cg}{}^2}{d_{cp}{}^2}}{\dfrac{\sinh\left(\frac{2\pi d_{cg}}{a}\right)}{\cosh\left(\frac{2\pi d_{cg}}{a}\right) - \cos\left(\frac{2\pi y}{a}\right)} - \dfrac{d_{cg}}{d_{cp}}} - \frac{d_{cg}}{d_{cp}} \tag{7.77}$$

This expression is properly independent of grid and plate potential and reduces to the low-mu amplification-factor formula for large values of $\frac{d_{cg}}{a}$ and $\frac{d_{cp}}{d_{cg}}$. This formula is reasonably accurate for values of $\frac{d_{cg}}{a}$ as low as 0.4. For small values of the ratio $\frac{d_{cg}}{a}$ the reciprocal of the amplification factor, sometimes referred to as the penetration factor or *Durchgriff* since it is a measure of the shielding effect of the grid, exhibits what is nearly a sinusoidal variation with distance parallel to the grid. For the case in which the grid-cathode distance is 0.4 of the grid-wire spacing, the relative amplification factor may vary between 0.7 and 1.4 times the average value, the average value being very nearly equal to that given by Eq. (7.13).

Expressions for the amplification factor of a plane triode have now been given that cover nearly the entire range of practical electrode dimensions. All the expressions given are limited, however, to some definite range of electrode dimensions. The expressions given in this subsection, for instance, are limited to screening fractions of 0.1 or less. The region of validity of the various amplification-factor formulas is shown in Fig. 7.18. Here it is seen that formulas are good for either large screening fractions or large ratio of grid-wire spacing to cathode-grid distance but not both.[1]

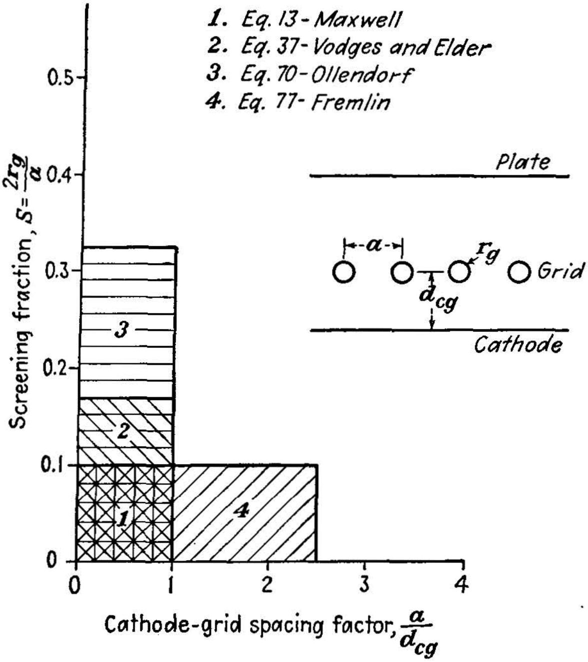


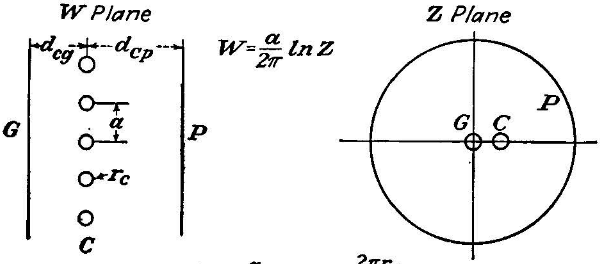
FIG. 7.18.—The region of validity of various amplification-factor formulas.

7.9. Amplification Factor of Unconventional Tubes. The methods that have been studied in this chapter may be applied to numerous structures other than the idealized plane and cylindrical structures so far treated. Where simple geometries are involved, it is usually possible to find a correspondingly simple arrangement of line charges that can be transformed into the desired structure. In Fig. 7.19 are shown some sample unconventional tube structures along with their elementary forms and the corresponding amplification-factor formula.

---

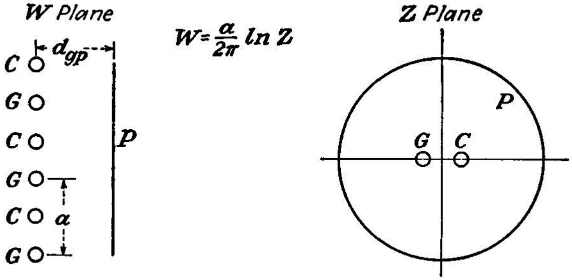
[1] See FREMLIN, J. H., R. N. HALL, and P. A. SHAFFORD, Triode Amplification Factors, *Elec. Commun.*, vol. 23, no. 4, pp. 426–435, 1946, for a semiempirical formula good for both small screening fraction and small cathode-grid distance.

In Fig. 7.19*a* is shown the famous gammatron "gridless wonder." This consists of a row of filament wires between grid and plate planes. The tube has an inherently low amplification factor whose value is approximately equal to $\frac{d_{cp}}{d_{cg}}$ as may be seen from consideration of interelectrode capacities. The formula given is valid for screening fractions of 0.1 or less and is based upon the same sort of analysis as was used to treat the low-mu triode. It is seen that the amplification factor has a second-order dependence upon the cathode radius.



$$\mu=\frac{d_{cp}-\frac{a}{2\pi}\ln\cosh\frac{2\pi r_c}{a}}{d_{cg}-\frac{a}{2\pi}\ln\cosh\frac{2\pi r_c}{a}}$$

FIG. 7.19*a*.—The amplification factor of a triode with a cathode in the form of a row of filament wires located between a grid and a plate plane.

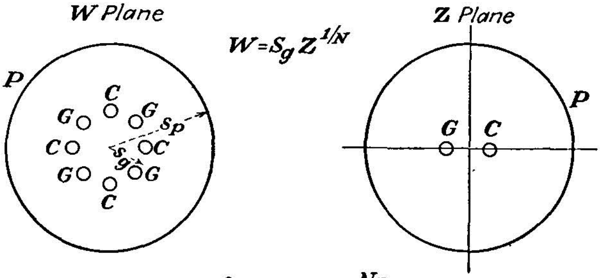


$$\mu=\frac{\frac{2\pi}{a}d_{gp}-\ln\left(2\cos\frac{\pi r_c}{a}\right)}{\ln\cos\frac{\pi r_g}{a}-\ln\sin\frac{\pi r_c}{a}}$$

FIG. 7.19*b*.—The amplification factor of a plane-electrode triode with a cathode consisting of a row of filament wires spaced midway between the grid wires and having a single plate.

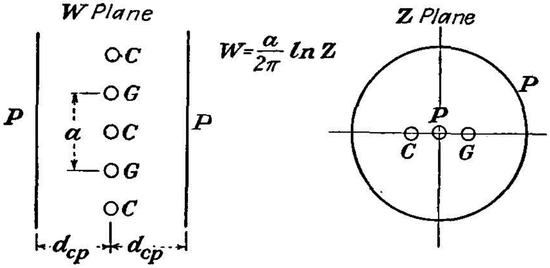
In Figs. 7.19*b*, *c*, and *d* are shown tubes in which filament wires are placed between the grid wires. The formulas are not greatly different from those for tubes which have a grid between cathode and plate except that they exhibit a second-order dependence upon cathode radius.

The question arises as to what the effect of having a filamentary emitter rather than a solid cathode would be. In general, if the filament-grid spacing is large compared with the grid-wire spacing, the amplification factor will be the same as for the case of a solid cathode. The equipotential lines around the filament will be circles very near to it but become nearly straight lines, the same as for the solid cathode, between the filament and grid. The current-voltage characteristics may, however, be considerably different.



$$\mu=\frac{N\,ln\frac{s_p}{s_g}-ln\left(2\cos\frac{Nr_c}{2s_g}\right)}{ln\left(2\cos\frac{Nr_g}{2s_g}\right)-ln\left(2\sin\frac{Nr_c}{2s_g}\right)}$$

FIG. 7.19c.—The amplification factor of a cylindrical triode whose cathode is a group of wires located between the wires of a squirrel-cage grid.



$$\mu=\frac{\frac{\pi d_{gp}}{a}-ln\left(2\cos\frac{\pi r_c}{a}\right)}{ln\left(2\cos\frac{\pi r_g}{a}\right)-ln\left(2\sin\frac{\pi r_g}{a}\right)}$$

FIG. 7.19d.—The amplification factor of a plane-electrode triode whose grid and cathode wires are alternate and equally spaced in a plane between two equidistant plate planes.

CHAPTER 8

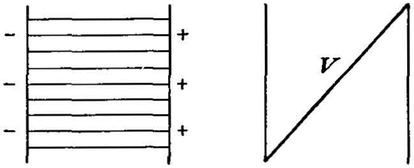
SPACE-CHARGE EFFECTS

8.1. Effects of Current Flow. In the previous chapter a study was made of the potential fields inside of tubes in the absence of current flow. Such studies can give only a partial picture of the true condition within a vacuum tube; for ordinarily currents will flow, and the presence of the electrons constituting the current introduces a distribution of charge known as "space charge," which changes the tube behavior. The previous studies are perfectly valid in determining such things as the amplification factor of tubes, for they can be applied to a condition of cutoff at which no current flows and yet at which the relative influence of the various electrodes is the same as for current flow. For studies of such subjects as the variation of current with potential and the determination of mutual conductance it is necessary to take into account the effect of currents and the corresponding space charge.

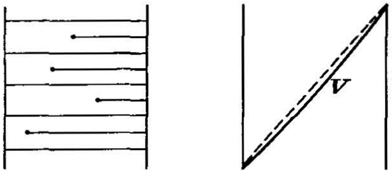
The effect of space charge is most readily studied in the case of the diode, and the results obtained from this study can then be extended to give the relations existing in triodes and multielectrode tubes. Actually, this extension can be made only approximately, but enough information can be obtained to answer most purposes. The most striking effects of space charge in a diode are to limit the current to a value determined by the three-halves power of plate voltage and to cause the potential distribution within the tube to be nonlinear.

In the plane-electrode diode the potential distribution in the absence of space charge is a straight line from one electrode to the other as shown in Fig. 8.1*a*. In this case all the flux lines emanating from positive charges on the positive plate terminate on negative charges on the other plate, as shown in the same figure.

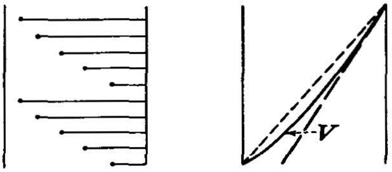
If now one of the electrodes is a cathode capable of emitting a small number of electrons and the other electrode is positive with respect to this, there will be some electrons in the field between the two electrodes moving toward the plate and some of the flux lines will terminate on electrons in the field as shown in Fig. 8.1*b*. The drawing has been conventionalized and the unit of flux density chosen so that there is only one flux line to each electron. Because of the negative charge in the field, the potential at any point will be less than in the previous case. Accordingly the potential profile will be pulled down and will be a curve concave upward at every point below the straight line, with the same initial and final values. If instead of negative electrons there had been positive ions in the field, the profile would have been moved up and would have formed a continuous curve concave downward at every point above the straight-line charge-free potential profile.



NO CURRENT
(a)



SMALL CURRENT
(b)



CURRENT LIMITED BY SPACE CHARGE
(c)

FIG. 8.1.—Electric flux lines and potential distribution in a plane-electrode diode for various degrees of space charge.

If the cathode is capable of emitting an unlimited number of electrons, the current will limit itself to a definite value because of the mutual repulsion between the electrons and because of the fact that the potential contour can be depressed only until its slope at the cathode is zero. There is an equilibrium here; for, as shown in Fig. 8.1*c*, if in some manner the slope could be made less than zero at the cathode, the electrons starting out would be forced to return to the cathode, no current would flow, the space charge would be reduced, and the potential contour would lift until its slope was zero at the cathode, at which point an equilibrium would be reached. If in some manner the slope were greater than zero at the cathode, more electrons would be encouraged to leave the vicinity of the cathode, the space charge would be increased, the potential-distribution curve would be depressed, and this action would continue until the slope at the cathode again became zero. The zero slope at the cathode indicates that the charge on the cathode is zero, which means that the flux lines emanating from charges on the positive plate all terminate on electrons in the field and that none of them get through to the cathode. In the case of the current limited by space charge, the potential distribution is a four-thirds-power law, as will be shown presently. In this case also, the current varies as the three-halves power of the potential on the plate.

The reason for the three-halves-power variation of current with potential can be shown from simple theoretical considerations. In Fig. 8.1c it is seen that the space-charge-limited diode is roughly equivalent to a condenser. As is the case in a condenser, charge on the plate and hence also in the interelectrode space is proportional to the potential. The current density is the product of the space-charge density and the velocity of the electrons at that point. The velocity of the electrons is proportional to the square root of potential so that the current density, which is the product of charge and velocity, is proportional to the three-halves power of potential. Although the above is not a very rigorous demonstration of the validity of the three-halves-power law of current, it will at least make the relation seem reasonable. A more rigorous demonstration will be undertaken in the next section.

8.2. Plane-electrode Space-charge Flow. The relations between potential, distance, and current in the plane-electrode case can be obtained from Poisson's equation, the energy equation, and the relation between current, charge, and velocity.

Poisson's equation in the one-dimensional case reduces to

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon_0} \tag{8.1}$$

where V is potential, ρ is volumetric space-charge density, and ε_0 is the dielectric constant of free space in rationalized mks units.

The energy equation has the form

$$\tfrac{1}{2}mv^2 = Ve \tag{8.2}$$

where m and e are the mass and charge of the electron and v is the velocity at any potential V. Electrical quantities are in practical units, and physical quantities are in mks units. This assumes that the electron has started from rest at a point of zero potential.

The relation between current density, charge, and velocity is

$$J = \rho v \tag{8.3}$$

The three equations above suffice for a determination of all the relations involved in a parallel-electrode space-charge flow. If ρ is expressed in terms of J and V from Eqs. (8.2) and (8.3) and the resulting expression substituted into Eq. (8.1),

$$\frac{d^2V}{dx^2} = \frac{J}{\varepsilon_0}\sqrt{\frac{m}{2e}}\,V^{-\frac{1}{2}} \tag{8.4}$$

where J is now the magnitude of the current density, actually electronic flow in the positive x direction is negative. A first integration is achieved by multiplying both sides of Eq. (8.4) by $2\,\frac{dV}{dx}$ and integrating,

$$\left(\frac{dV}{dx}\right)^2 = \frac{4J}{\varepsilon_0}\sqrt{\frac{m}{2e}}\,V^{1/2} + C_1 \tag{8.5}$$

The constant of integration is zero because the gradient is zero when the potential is zero.

A second integration gives

$$\frac{4V^{3/4}}{3} = \sqrt{\frac{4J}{\varepsilon_0}\sqrt{\frac{m}{2e}}}\,x + C_2 \tag{8.6}$$

in which the constant is again zero because the potential is taken as zero when the distance is zero. Solving for current density,

$$J = \frac{4\varepsilon_0}{9\sqrt{\frac{m}{2e}}}\frac{V^{3/2}}{x^2} \qquad \text{amperes per unit area} \tag{8.7}$$

Numerically this is equal to

$$J = \frac{2.335 \times 10^{-6} V^{3/2}}{x^2} \qquad \text{amperes per unit area} \tag{8.8}$$

If x is in centimeters, the current density is in amperes per square centimeter.

From the above equations it is seen that the current varies as the three-halves power of potential and inversely as the square of the distance. The last two equations constitute the Child-Langmuir space-charge law. It has been verified experimentally.[1-3] If the equations be solved for potential, there results

$$V = 5{,}680 J^{2/3} x^{4/3} \tag{8.9}$$

showing that the potential varies as the four-thirds power of distance between cathode and plate.

Values of current density in terms of distance and potential are given in the curves of Fig. 8.2.

It is of interest to note how various other factors vary with distance. The gradient of potential is given by the derivative of potential with distance. Hence

$$E = k_1 x^{1/3} \tag{8.10}$$

---

[1] CHILD, D. C., Discharge from Hot CaO, *Phys. Rev.*, vol. 32, pp. 492–511, May, 1911.

[2] LANGMUIR, I., The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum, *Phys. Rev.*, Ser. 2, vol. 2, pp. 450–486, December, 1913.

[3] LANGMUIR, I., and K. B. BLODGETT, Currents Limited by Space Charge between Coaxial Cylinders, *Phys. Rev.*, Ser. 2, vol. 22, pp. 347–357, October, 1923.

Velocity varies as the square root of potential, and therefore

$$v = k_2 x^{2/3} \tag{8.11}$$

Space-charge density varies inversely as velocity, from Eq. (8.3), since the current density is constant; thus

$$\rho = k_3 x^{-2/3} \tag{8.12}$$

Curves showing the variation of these various factors are shown in Figs. 8.3*a* and *b*. Here it is seen that the potential and gradient are zero at the cathode. Since the velocity was assumed zero at this point, the space-charge density is theoretically infinite here. Actually, the electrons start with a small finite velocity rather than with zero velocity so that the gradient is initially negative for a small distance, passes through a

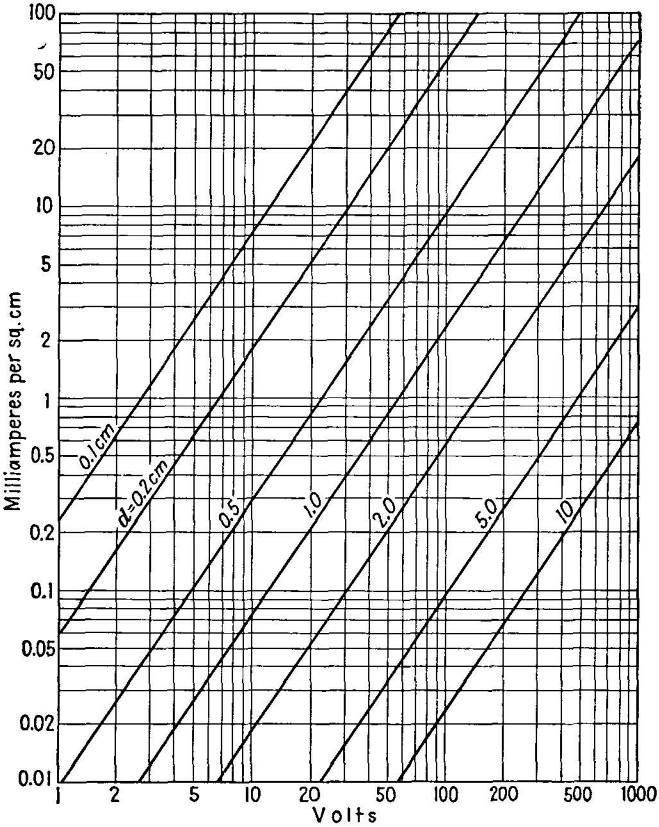


Fig. 8.2.—Current density in a plane-electrode diode as a function of voltage and electrode spacing, Eq. (8.8).

minimum at a small negative potential, and then increases. In spite of this difference, the Child-Langmuir law is quite accurate except for very low values of potential corresponding to the average velocity of emission, which is of the order of a few volts.

8.3. Cylindrical-electrode Space-charge Flow. When the electrode structure consists of two concentric cylinders the inner of which is capable of emitting electrons, the space-charge-limited current still varies with the three-halves power of the voltage but the effect of the electrode dimensions is a little more complex.

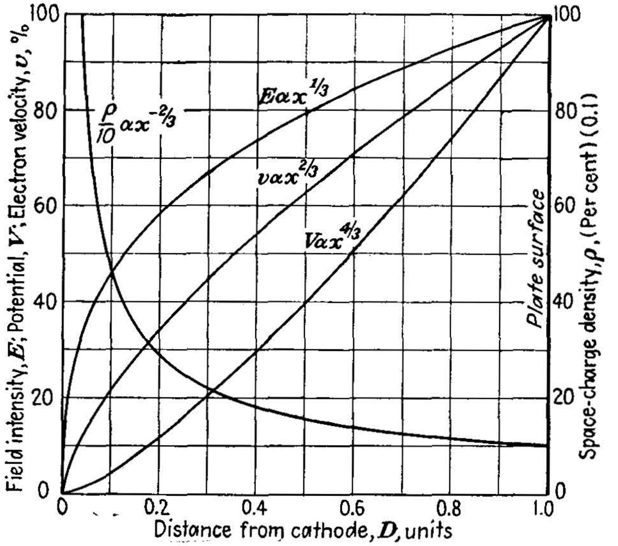


FIG. 8.3*a*.—Curves of potential, gradient, space-charge density, and velocity as a function of interelectrode distance in a space-charge-saturated plane diode, linear scales.

The behavior of the cylindrical diode can be studied in just the same way as the plane-electrode diode, but in this case it is necessary to use cylindrical coordinates. Actually, this does not complicate the problem too much, for the conditions of symmetry are such that, at any fixed radial distance, conditions are the same regardless of angle. The problem is thus still a one-dimensional one.

For this case Poisson's equation reduces to

$$\frac{1}{r}\frac{d}{dr}\left(\frac{r\,dV}{dr}\right) = -\frac{\rho}{\varepsilon_0} \tag{8.13}$$

The same energy equation as for the plane-electrode case holds,

$$\tfrac{1}{2}mv^2 = eV \tag{8.2}$$

Current density at any radial distance r is given by

$$J(r) = \rho v \tag{8.14}$$

Since the current density varies with radial distance, it is more convenient to express the space-charge density in terms of total current per unit length of axis.

$$\frac{I}{l} = 2\pi r J(r) \tag{8.15}$$

$$= 2\pi r \rho v \tag{8.16}$$

where I is the total current passing at right angles through an imaginary cylinder of radius r and length l.

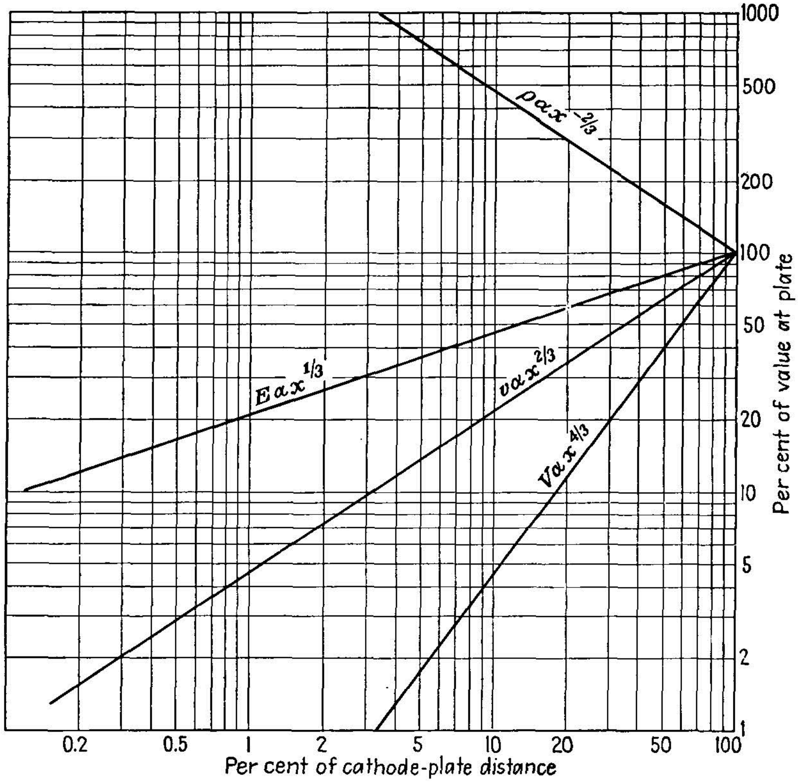


FIG. 8.3*b*.—Same as Fig. 8.3*a*, plotted with logarithmic scales.

Expressing the right side of Eq. (8.13) in terms of I and V from Eqs. (8.2) and (8.16),

$$\frac{d}{dr}\left(r\,\frac{dV}{dr}\right) = \frac{I}{2\pi\varepsilon_0 l}\sqrt{\frac{m}{2eV}} \tag{8.17}$$

which may also be written as

$$\frac{d^2V}{dr^2} + \frac{1}{r}\frac{dV}{dr} = \frac{J}{\varepsilon_0}\sqrt{\frac{m}{2eV}} \tag{8.18}$$

An approximate solution of Eq. (8.18) can be obtained by observing that for large values of r the second term of the left member of the equation is negligible. If this term is dropped, the resulting equation is the same as Eq. (8.4) for the plane-electrode case so that the solution for large values of r would be expected to be the same in both cases. Thus

$$J = \frac{2.335 \times 10^{-6}V^{3/2}}{r^2} \qquad \text{amperes per unit area} \tag{8.19}$$

or

$$\frac{I}{l} = \frac{2\pi \times 2.335 \times 10^{-6}V^{3/2}}{r} \qquad \text{amperes per unit length of axis} \tag{8.20}$$

These approximate equations hold within 10 per cent for values of r greater than ten times the cathode radius.

A more exact solution is obtained by assuming that the expression for current is of the form

$$\frac{I}{l} = \frac{KV^{3/2}}{r\beta^2} \tag{8.21}$$

where β^2 is a function of the ratio of the radius at any point to the cathode radius.[1] Substituting Eq. (8.21) into Eq. (8.17) gives

$$3\beta r^2\frac{d^2\beta}{dr^2} + r^2\left(\frac{d\beta}{dr}\right)^2 + 7\beta r\frac{d\beta}{dr} + \beta^2 - 1 = 0 \tag{8.22}$$

This can be simplified slightly by letting $u = \ln\left(\frac{r}{r_c}\right)$, a logical substitution because the space-charge-free potential is expressible in this form. With this change of variable,

$$3\beta\frac{d^2\beta}{du^2} + \left(\frac{d\beta}{du}\right)^2 + 4\beta\frac{d\beta}{du} + \beta^2 - 1 = 0 \tag{8.23}$$

This may be solved by series, the solution being

$$\beta = u - \frac{2u^2}{5} + \frac{11u^3}{120} - \frac{47u^4}{3,300} + \cdots \tag{8.24}$$

This expression is valid for either an internal or an external cathode.

---

[1] LANGMUIR, *op. cit.*

Values of β^2 have been calculated by Langmuir by means of this series and other equivalent expressions which are more convenient for large values of u. In Table V are tabulated the values of β^2 and in Fig. 8.4 is shown a curve of β^2 as a function of $\frac{r}{r_c}$. It is seen that for ratios of $\frac{r}{r_c}$ greater than 7 the value of β^2 differs from 1 by less than 10 per cent, thus substantiating the approximate form of the equation for current, previously given. In the chart of Fig. 8.5 is given the current per unit length of axis for various voltages and various values of plate radius for the case in which β^2 has the value of 1. For other values of β^2 the current obtained from this chart must be divided by the value of β^2 as obtained from Fig. 8.4.

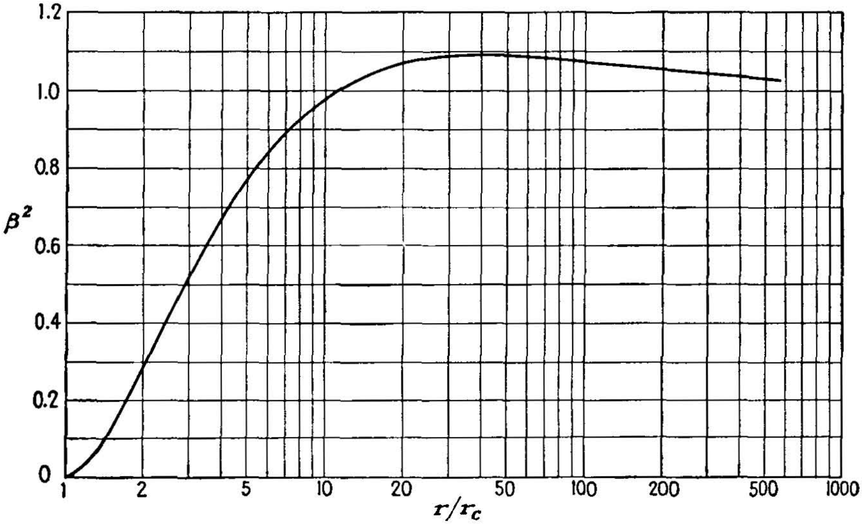


FIG. 8.4.—Curve of β^2 as a function of $\left(\frac{r}{r_c}\right)$.

In practical units the expression for current is

$$\frac{I}{l} = \frac{14.66 \times 10^{-6} V^{3/2}}{r\beta^2} \quad \text{amperes per unit length of axis} \qquad (8.25)$$

Since the current per unit length of axis is a constant, for a constant plate voltage, the variation of potential with radial distance r is given by

$$V = k_1(r\beta^2)^{2/3} \qquad (8.26)$$

in which it must be recognized that β^2 is not a constant but a function of the radial distance r. Velocity of the electron is given by the square

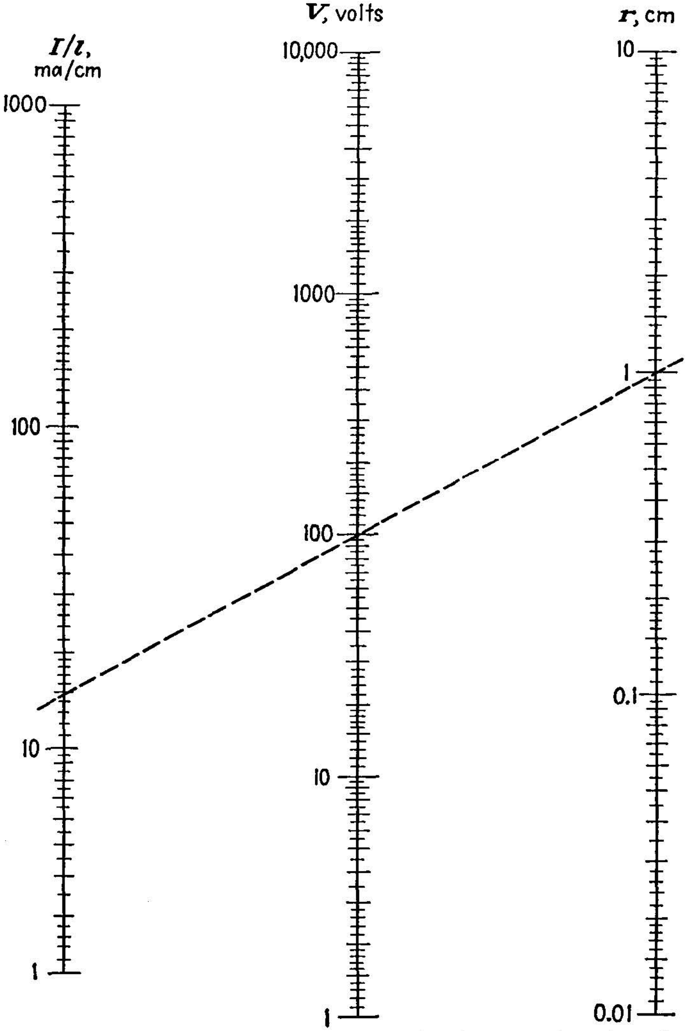


FIG. 8.5.—Nomographic chart of linear current density as a function of potential and plate radius in a cylindrical diode ($\beta^2 = 1$).

root of potential, and thus

$$v = k_2(r\beta^2)^{1/3} \tag{8.27}$$

From Eq. (8.16), space-charge density is the reciprocal of the product of velocity and radial distance r,

$$\rho = \frac{k_3}{r(r\beta^2)^{1/3}} \tag{8.28}$$

TABLE V

VALUES OF β^2 AS A FUNCTION OF $\frac{r}{r_c}$ AS GIVEN BY EQ. (8.24)

(β^2 applies where $r > r_c$; $-\beta^2$ applies where $r < r_c$)

$\frac{r}{r_c}$ or $\frac{r_c}{r}$	β^2	$-\beta^2$	$\frac{r}{r_c}$ or $\frac{r_c}{r}$	β^2	$-\beta^2$
1.00	0.0000	0.0000	3.8	0.6420	5.3795
1.01	0.00010	0.00010	4.0	0.6671	6.0601
1.02	0.00039	0.00040	4.2	0.6902	6.7705
1.04	0.00149	0.00159	4.4	0.7115	7.5096
1.06	0.00324	0.00356	4.6	0.7313	8.2763
1.08	0.00557	0.00630	4.8	0.7496	9.0696
1.10	0.00842	0.00980	5.0	0.7666	9.887
1.15	0.01747	0.02186	5.2	0.7825	10.733
1.20	0.02815	0.03849	5.4	0.7973	11.601
1.30	0.05589	0.08504	5.6	0.8111	12.493
1.40	0.08672	0.14856	5.8	0.8241	13.407
1.50	0.11934	0.2282	6.0	0.8362	14.343
1.60	0.1525	0.3233	6.5	0.8635	16.777
1.70	0.1854	0.4332	7.0	0.8870	19.337
1.80	0.2177	0.5572	7.5	0.9074	22.015
1.90	0.2491	0.6947	8.0	0.9253	24.805
2.0	0.2793	0.8454	8.5	0.9410	27.701
2.1	0.3083	1.0086	9.0	0 9548	30.698
2.2	0.3361	1.1840	9.5	0.9672	33.791
2.3	0.3626	1.3712	10.0	0.9782	36.976
2.4	0.3879	1.5697	12.0	1.0122	50.559
2.5	0.4121	1.7792	16.0	1.0513	81.203
2.6	0.4351	1.9995	20.0	1.0715	115.64
2.7	0.4571	2.2301	40.0	1.0946	327.01
2.8	0.4780	2.4708	80.0	1.0845	867.11
2.9	0.4980	2.7214	100.0	1.0782	1174.9
3.0	0.5170	2.9814	200.0	1.0562	2946.1
3.2	0.5526	3.5293	500.0	1.0307	9502.2
3.4	0.5851	4.1126	∞	1.000	∞
3.6	0.6148	4.7298			

For small u

$$\beta^2 = u^2(1 - 0.8u + 0.344u^2 + \cdots)$$

$$\frac{d\beta^2}{du} = 2u - 2.4u^2 + 1.374u^3 - 0.509u^4 + \cdots$$

where $u = \ln \frac{r}{r_c}$.

Gradient of potential is given by the negative derivative of potential;

$$E = \frac{-k_4 \dfrac{d(r\beta^2)}{dr}}{(r\beta^2)^{1/3}} \tag{8.29}$$

For radial distance r equal to r_c both the numerator and the denominator of the expression for E are zero. However, the numerator is a zero of higher order, as may be checked by referring to the series for β, and therefore the gradient is zero for r equal to r_c. In plotting a curve of the gradient the derivative of $r\beta^2$ must be evaluated numerically from the table of values of β^2. Curves of V, E, v, and ρ for a typical cylindrical triode are shown in Fig. 8.6 for two ratios of plate to cathode radii. Curves for other radii have forms similar to those shown. Although not evident from the appearance of the curves, the potential profile leaves the cathode with zero slope, having a considerable change of slope in a short distance. This can also be seen from the series expansion of the expression for potential near the cathode, the first term of which is

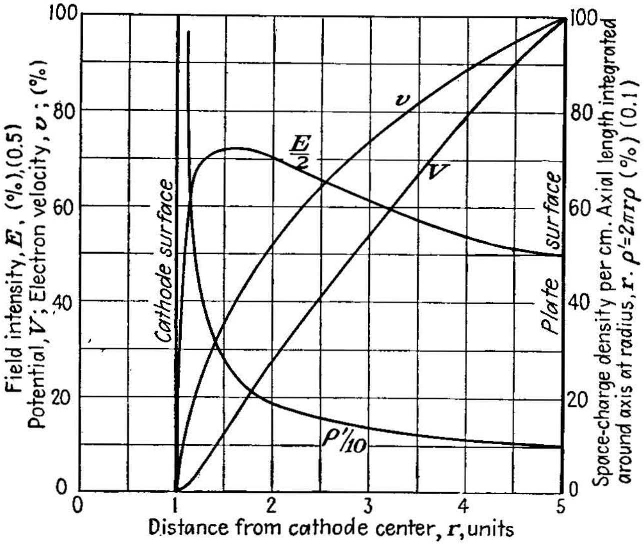


FIG. 8.6*a*.—Curves of potential, gradient, space-charge density, and velocity as a function of radial distance in a space-charge-saturated cylindrical diode, $\left(\frac{r_p}{r_c}\right) = 5$.

$$V = k_1 r_c^{2/3} \left(\frac{r - r_c}{r_c}\right)^{4/3} \tag{8.30}$$

From this it can also be seen that the gradient has the value of zero at the cathode.

In all the foregoing it has been assumed that the cathode cylinder is smaller than the plate cylinder. The formulas are also valid if the cathode is the outer cylinder, though tubes are rarely built this way. It might be thought that if the cathode were the outer cylinder a greater current would flow for a given voltage than if it were the inner cylinder because the same current would be distributed through a greater volume where its velocity is lowest. Examination of the numerical values shows that this is not so, however. When the cathode is the outer cylinder, the functon β^2 increases very rapidly as the ratio of plate to cathode radius decreases. As a result, the current is actually less when the cathode is the outer cylinder; for although the factor r in Eq. (8.21) is decreased, the factor β^2 has increased more than enough to offset this.

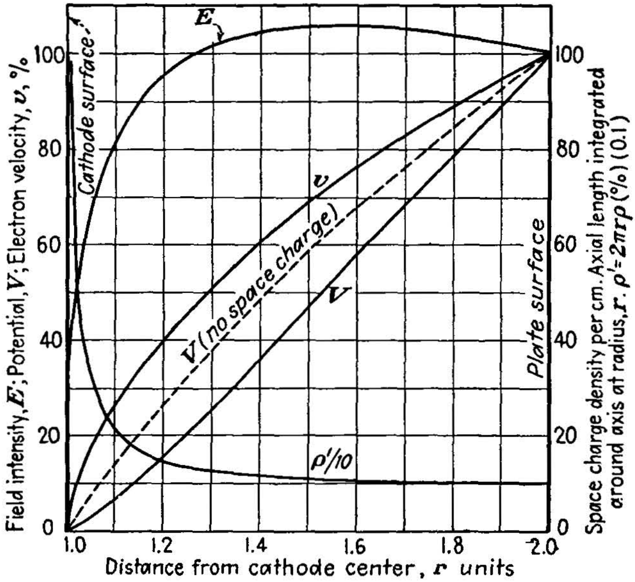


FIG. 8.6b.—Same as Fig. 8.6a, $\left(\frac{r_p}{r_c}\right) = 2.$

It is of interest to compare Eqs. (8.8) and (8.25) for the plane and cylinder case. If Eq. (8.25) is divided by $2\pi r$, it then resembles Eq. (8.8) except that x is replaced by $r\beta$. When r is very large, β approaches unity and the expressions become identical. When r is only slightly larger than r_c, β is approximately equal to $\ln\left(\frac{r}{r_c}\right)$ and hence to $\frac{r - r_c}{r_c}$, with the result that the factor $r\beta$ nearly equals $r - r_c$ and the expressions for current density again approach identity with x replaced by $r - r_c$. This identity is expected for values of $\frac{r}{r_c}$ near unity; for here the electrode spacing is small compared with the radii, and a plane structure is approximated. The fact that Eqs. (8.8) and (8.25) are identical for limiting values of the ratio $\frac{r}{r_c}$ does not mean that they are nearly equal for all intermediate values. For values of $\frac{r}{r_c}$ between 4 and 20 the value of current density as calculated from the plane-electrode formula exceeds that obtained from the cylindrical-electrode formula by nearly 20 per cent. This is the maximum discrepancy that can occur.

8.4. Space-charge Flow for Other Geometries. *Spherical Electrodes.* The equations for space-charge flow of current can also be derived for concentric spherical electrodes.[1] For this case it is found that the current varies as the three-halves power of potential and is inversely proportional to a dimensionless function of the ratio of plate and cathode radius. The total current is given by

$$I = \frac{29.34 V^{3/2}}{\alpha^2} \times 10^{-6} \qquad \text{amperes} \tag{8.31}$$

where V is the potential difference between the spherical electrodes in volts and α is a function of $u = \ln\left(\frac{r}{r_c}\right)$, given by

$$\alpha = u - 0.3u^2 + 0.075u^3 - 0.00143u^4 + 0.00216u^5 - \cdots \tag{8.32}$$

This expression for α is valid whether the cathode is external or internal.

Values of α for the spherical case and of β for the cylindrical case are equal within 2 per cent for values of $\frac{r}{r_c}$ between 0.65 and 1.35. For values of $\frac{r}{r_c}$ less than 1, α differs not more than 10 per cent from the larger value of β.[2] For values of $\frac{r}{r_c}$ very nearly equal to 1 the current density approaches that for plane electrodes.

The General Case. The observation that the current varies with the three-halves power of potential for plane, cylindrical, and spherical electrodes leads one to believe that this is the case for electrodes of any shape. Actually this is so, but the conclusion must be examined carefully, for the three cases enumerated are special cases in which the elec-

---

[1] LANGMUIR, I., and K. T. COMPTON, Electrical Discharges in Gases, Part II, *Rev. Modern Phys.*, vol. 13, pp. 191–257, April, 1931.

[2] See Appendix VII for values of α^2.

trons move in straight lines perpendicular to the equipotential surfaces. For other geometries this is not necessarily so. For electrodes of other shapes the electrons will in general cut across the lines of electric force and move in curved paths.

The validity of the three-halves-power law of potential can however be shown by a simple dimensional analysis of the basic equations from which the current laws were developed. These are Poisson's equations,

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0} \tag{8.33}$$

the energy equation

$$\tfrac{1}{2}mv^2 = eV \tag{8.34}$$

and the current-density expression

$$\mathbf{J} = \rho\mathbf{v} \tag{8.35}$$

In previous cases the quantities in the last equation above have been treated as scalar quantities, but in the general case current density and velocity must be treated as vector quantities because they do not necessarily have the direction of the gradient of potential.

Consider now what happens if the potential is increased by a factor k. If the electrons move in a curved path, their centrifugal force $\frac{mv^2}{R}$, where R is the instantaneous radius of curvature, must equal the force due to the component of the gradient of potential normal to the path, $e\nabla_n V$. From Eq. (8.34) the centrifugal force will have increased by a factor k, and likewise the gradient will have increased by the same factor, so that the shape of the electron paths will be unchanged. This is the same conclusion that was reached in the case of the space-charge-free fields. Once this is established, the final conclusion follows immediately. From Eq. (8.33) the space-charge density is increased by a factor k, and from Eq. (8.34) the velocity is increased by a factor $k^{1/2}$. Hence, by Eq. (8.35), the current density is increased by a factor $k^{3/2}$, and the general validity of the three-halves-power law of potential for electrons starting from rest is established.

It can also be shown by the same type of reasoning that, if an electrode structure is enlarged by a factor n maintaining geometrical similarity and voltages are kept unchanged, then if the current is space-charge-limited, the current is also unchanged. From Eq. (8.33) space-charge density is decreased by a factor n^2, and hence current density is decreased by a factor of n^2 from Eq. (8.35), velocity at corresponding points in the original and enlarged structure being the same. However, since the total area over which the current density is summed is increased by a factor of

n^2, it is seen that the total current is unchanged if the voltages are unchanged.

8.5. Current Law for Plane Triodes. It is found experimentally in triodes that the total current released from the emitter is very nearly proportional to the three-halves power of the equivalent voltage, $V_g + \frac{V_p}{\mu}$. This is most readily shown by plotting curves of constant space current, *i.e.*, the sum of plate and grid current, against equivalent voltage on logarithmic paper. All points and curves so plotted tend to fall on the same straight line, which has a slope of nearly $\frac{3}{2}$. Slight departures from the slope of $\frac{3}{2}$ are sometimes encountered because of initial electron velocity, the Schottky effect, and because of potential drop along the emitter. The relation holds whether the grid is negative or positive even though the space current in the first case is all plate current and in the

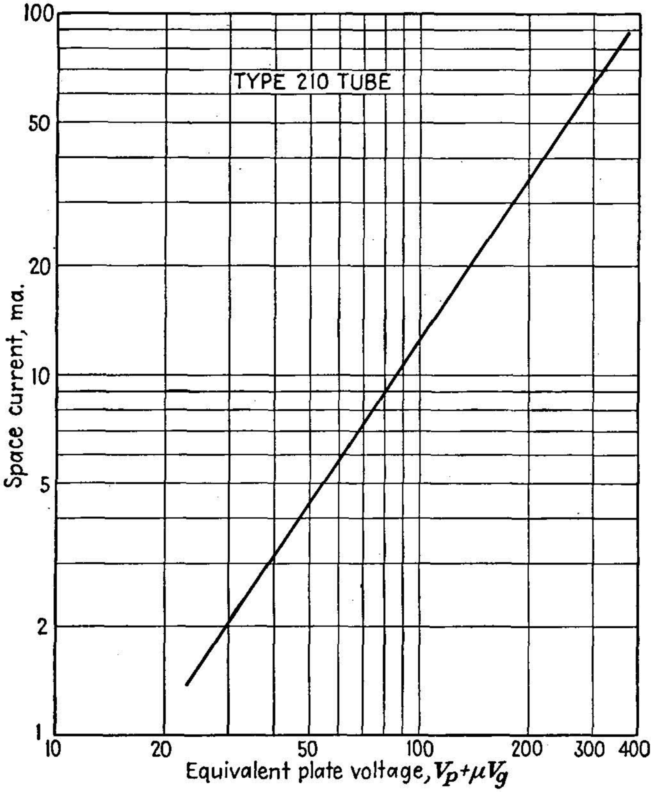


FIG. 8.7.—Current as a function of equivalent voltage in a 210 triode.

second case is the sum of grid and plate current. This is in accord with the reasoning presented in the previous section.

The correlation between the theory and experiment is sufficiently good so that in general it is a very good approximation to write

$$J = J_p + J_g = k\left(V_g + \frac{V_p}{\mu}\right)^{3/2} \tag{8.36}$$

in which k is a constant to be determined later.

The equivalent voltage referred to here is the same as the equivalent voltage encountered in the study of the space-charge-free potential distributions in tubes. There it was found that the off-cathode gradient was a linear function of the equivalent voltage just as the cathode gradient of the space-charge-free diode is directly proportional to the plate potential. Here the space current depends upon the three-halves power of the equivalent potential just as the saturated diode current is proportional to the three-halves power of the plate potential. This is strictly true only when the space charge in the grid-plate region is negligible. A curve showing how the space current varies with equivalent voltage in a typical triode is given in Fig. 8.7.

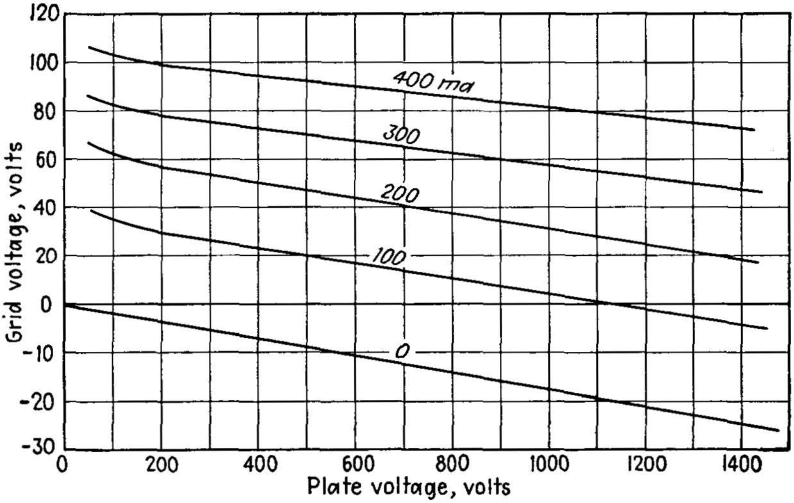


FIG. 8.8.—Constant-current representation of triode characteristics.

The fact that the relative effectiveness of the plate and grid in controlling current flow is the same for a great range of current is shown by plotting contours of constant space current against axes of grid and plate voltage. Such a set of curves is shown in Fig. 8.8. It is seen that the curves are substantially straight lines whose slope by definition is $-\frac{1}{\mu}$. The constancy of the slope of these straight-line curves attests to the constancy of the amplification factor. It will be observed that for very small currents the slope decreases, indicating a lower amplification factor. This is explained by the fact that the amplification factor of all parts of the electrode structure is not the same, with the result that the parts with the highest μ cut off first, leaving current to pass through the parts of lower μ. Variations in the magnitude of the amplification factor, which actually is an equivalent amplification factor of a number of areas with slightly different factors connected in parallel, also account for slight departures of the current law from a strict three-halves power law.

Current Law in Terms of Electrode Dimensions. The coefficient of the current law given in Eq. (8.36) can be evaluated by fitting the triode electrode potentials to the diode law for a particular combination of potentials and then assuming that the relation which holds for this particular case holds for all combinations by virtue of the experimental observations.[1]

Consider a plane-electrode triode, and imagine first that the grid is not present and that there is a space-charge-limited current flow from cathode to plate. Then if the grid were inserted at a positive potential corresponding to that which existed at its location before its insertion, its presence would not disturb the existing potential distribution and would not change the magnitude of the plate current. Since the equivalent voltage can now be determined for a given current, the constant of Eq. (8.36) can be evaluated. When this constant is known for one combination of potentials, our experimental observations show that it is the same for all combinations of potentials and thus the current law for triodes is determined in terms of the electrode dimensions.

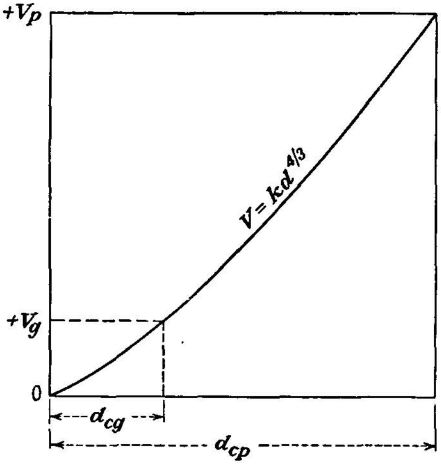


FIG. 8.9.—Potential distribution in a positive-grid triode with grid at its natural potential.

The distribution of potentials referred to above is shown in Fig. 8.9. Here it is seen that the potential distribution from cathode to plate is a

---

[1] FREMLIN, J. H., Calculation of Triode Constants, *Elec. Commun.*, July, 1939.

four-thirds-power-law curve as in the case of the diode. On this basis, the relation between the current density, the diode plate potential V_v, and the cathode-plate distance is

$$J = \frac{KV_p^{3/2}}{d_{cp}^2} \quad \text{amperes per unit area} \tag{8.37}$$

where K is 2.335×10^{-6}. Let this be written in the form

$$V_p = \frac{J^{2/3} d_{cp}^{4/3}}{K^{2/3}} \tag{8.38}$$

The grid is inserted with the potential that would exist in the diode at the location of the grid plane as shown in the figure. The relation for current density, positive grid potential, and cathode-grid distance is

$$V_g = \frac{J^{2/3} d_{cg}^{4/3}}{K^{2/3}} \tag{8.39}$$

But the experimental observation in keeping with theoretical considerations is that

$$J = k\left(V_g + \frac{V_p}{\mu}\right)^{3/2} \tag{8.40}$$

Substitutions from Eqs. (8.38) and (8.39) give

$$J = \frac{k\left(J^{2/3} d_{cg}^{4/3} + \frac{J^{2/3} d_{cp}^{4/3}}{\mu}\right)^{3/2}}{K} \tag{8.41}$$

so that

$$k = \frac{K}{\left(d_{cg}^{4/3} + \frac{d_{cp}^{4/3}}{\mu}\right)^{3/2}} \tag{8.42}$$

or, numerically,

$$k = \frac{2.335 \times 10^{-6}}{\left(d_{cg}^{4/3} + \frac{d_{cp}^{4/3}}{\mu}\right)^{3/2}} \tag{8.43}$$

As a result, the expression for current is

$$J = \frac{2.335 \times 10^{-6}\left(V_g + \frac{V_p}{\mu}\right)^{3/2}}{d_{cg}^2\left[1 + \frac{1}{\mu}\left(\frac{d_{cp}}{d_{cg}}\right)^{4/3}\right]^{3/2}} \quad \begin{array}{c}\text{amperes per}\\ \text{unit area}\end{array} \tag{8.44}$$

This is the expression that has been sought and that has been the object of the above development. It is seen to be of the same form as the expression for the diode current density of Eq. (8.8). There is good experimental verification of Eq. (8.44).

It may also be seen that the space-charge-saturated equivalent-diode spacing of the triode is

$$d_{se} = d_{cg}\left[1 + \frac{1}{\mu}\left(\frac{d_{cp}}{d_{cg}}\right)^{4/3}\right]^{3/4} \tag{8.45}$$

A nomograph giving the space-charge-saturated equivalent-diode spacing of a triode in terms of the cathode-grid distance, the cathode-plate distance, and the amplification factor is given in Fig. 8.10. It should be noted that the space-charge-saturated equivalent-diode spacing of a triode given here is somewhat different from the space-charge-free equivalent-diode spacing given in Eq. (7.53) of the chapter on Electrostatic Field of a Triode. Each is slightly greater than the cathode-grid distance. For a μ of 10 and a ratio of cathode-plate to cathode-grid spacing of 5 the value of diode spacing from Eq. (8.45) is only 6 per cent greater than that from Eq. (7.53). For large values of μ the difference is even less. Various attempts have been made to devise expressions for the triode current and mutual conductance in terms of the space-charge-

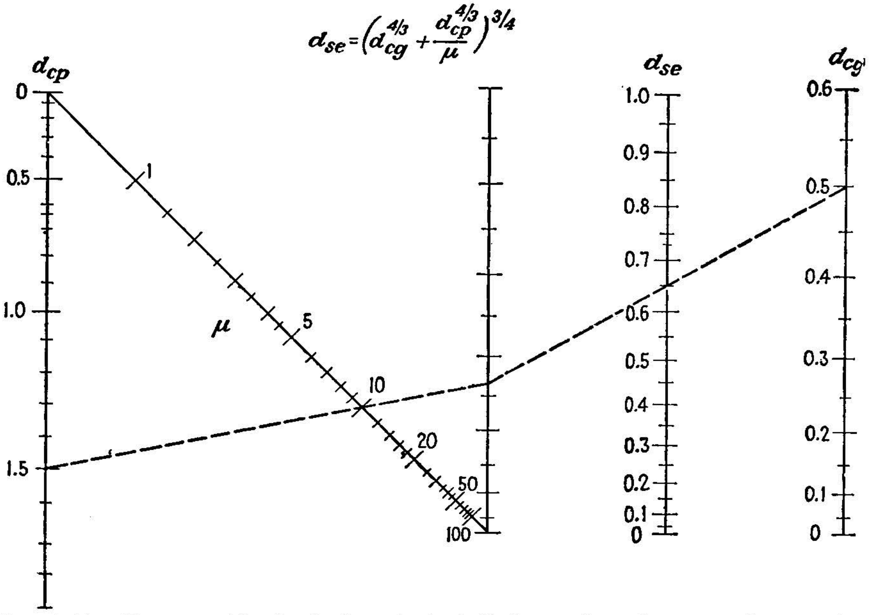


FIG. 8.10.—Nomographic chart of equivalent-diode spacing of a space-charge-saturated plane-electrode triode.

free equivalent-diode spacing, but all these are subject to an inescapable error.[1]

8.6. Mutual Conductance of a Plane Triode. The mutual conductance is easily obtained from the expression for current density of Eq. (8.44). By definition,

$$g_m = \frac{\partial I}{\partial V_g} = \left(\frac{dI_p}{dV_g}\right)_{V_p=k} \tag{8.46}$$

Performing this operation on Eq. (8.44),

$$g_m = \frac{3.51 \times 10^{-6}\left(V_g + \dfrac{V_p}{\mu}\right)^{1/2}}{d_{cg}{}^2\left[1 + \dfrac{1}{\mu}\left(\dfrac{d_{cp}}{d_{cg}}\right)^{4/3}\right]^{3/2}} \quad \text{amperes per volt per unit area} \tag{8.47}$$

This can also be written in the form

$$g_m = \frac{2.64 \times 10^{-4} J^{1/3}}{d_{cg}{}^{4/3}\left[1 + \dfrac{1}{\mu}\left(\dfrac{d_{cp}}{d_{cg}}\right)^{4/3}\right]} \quad \text{amperes per volt per unit area} \tag{8.48}$$

As far as tube geometry is concerned, the mutual conductance of a triode depends primarily upon the cathode-grid spacing. The smaller the cathode-grid spacing, the larger the mutual conductance. The mutual conductance also increases, though to a smaller extent, as the ratio of cathode-plate to cathode-grid spacing is decreased and as the amplification factor is increased.

It will be observed further that the mutual conductance increases as both the equivalent voltage and the current are increased. A specification of mutual conductance is really meaningless unless the corresponding voltages are also indicated. The variation of mutual conductance with the one-third power of current is a general law that holds well for all types of tubes, including pentodes as well as triodes.

8.7. Mutual Conductance of a Cylindrical Triode. The current law and mutual conductance of cylindrical triodes are readily evaluated by an analysis similar to that used for the plane-electrode triode. The current is given by

$$\frac{I}{l} = \frac{14.66 \times 10^{-6}\left(V_g + \dfrac{V_p}{\mu}\right)^{3/2}}{r_g\beta_{cg}{}^2\left[1 + \dfrac{1}{\mu}\left(\dfrac{r_p\beta_{cp}{}^2}{r_g\beta_{cg}{}^2}\right)^{2/3}\right]^{3/2}} \quad \text{amperes per unit length} \tag{8.49}$$

---

[1] WALKER, G. B., Theory of the Equivalent Diode, *Wireless Engr.*, vol. 24, pp. 5–7, January, 1947.

where r_g and r_p are grid and plate radii and the combination subscripts indicate that β^2 is to be determined by the ratio of the radii of the electrodes indicated by the subscripts.

The mutual conductance is

$$g_m = \frac{22.0 \times 10^{-6}\left(V_g + \frac{V_p}{\mu}\right)^{1/2}}{r_g\beta_{cg}{}^2\left[1 + \frac{1}{\mu}\left(\frac{r_p\beta_{cp}{}^2}{r_g\beta_{cg}{}^2}\right)^{2/3}\right]^{3/2}} \quad \text{amperes per volt per unit length} \qquad (8.50)$$

8.8. Effect of Filamentary Emitters. When tubes have filaments instead of solid cathodes, a number of effects contribute to making the behavior different from that of tubes with solid cathodes. Foremost among these is the voltage drop along the filament, which may cause considerable divergence from the simple three-halves power of current with voltage. If the plate current is returned to the negative filament lead, the current is at all times less than that for a unipotential cathode but becomes nearly equal to that value for large values of the ratio of plate to filament voltage. The ratio of the current without to that with a unipotential filament is shown in Fig. 8.11. This is essentially a correction factor for the fact that the filament potential is not uniform. The ratio of currents is 0.4 for a plate- to filament-voltage ratio of 1, dropping linearly to 0 with this ratio for plate voltages less than the filament voltage. When the plate voltage exceeds the filament voltage by more than a factor of 15, the current ratio is within 5 per cent of unity.

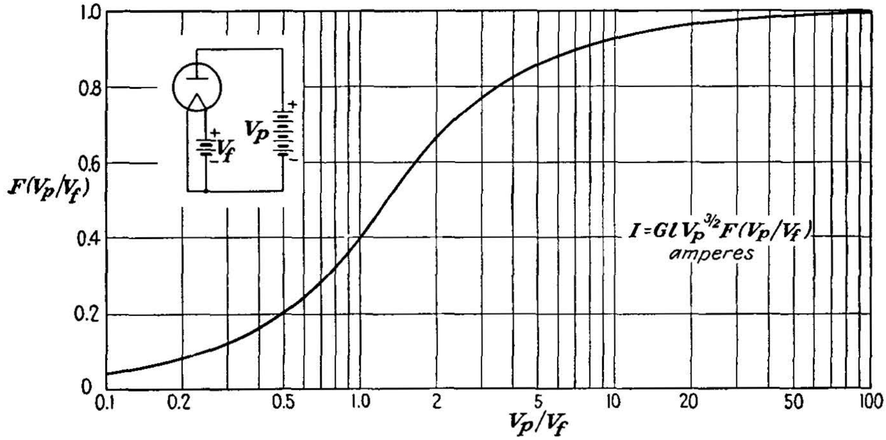


FIG. 8.11.—Filament voltage-drop emission correction factor.

The above results are arrived at by integrating the emission effect along the filament, taking into account the different potential differences to the plate at each point. Every element of length of the filament contributes a current

$$dI = \frac{14.66 \times 10^{-6}[V(x)]^{3/2}\,dx}{r_p\beta^2} \tag{8.51}$$

which may be written more simply as

$$dI = G[V(x)]^{3/2}\,dx \tag{8.52}$$

in which G is the so-called "perveance" of the tube. The perveance is simply the coefficient of the voltage factor in the Child-Langmuir law. It contains the geometrical factors of the tube and has dimensions of current per unit length per volt$^{3/2}$.

It is necessary to introduce the potential difference between the filament and plate as a function of distance along the filament. This is

$$V(x) = V_p - \frac{x}{l}V_f \tag{8.53}$$

where V_p is the plate potential relative to the negative filament lead, x is the distance from the negative end of the filament, l is the length of the filament, and V_f is the potential drop along the filament. With this substitution, Eq. (8.52) must be integrated from 0 to l. The results of this operation fall into two parts.

Case I. $V_p < V_f$ leads to

$$I = \frac{2}{5}GV_p^{3/2}\left(\frac{V_p}{V_f}\right)l. \tag{8.54}$$

Case II. $V_p > V_f$ leads to

$$I = \frac{2}{5}\frac{Gl}{V_f}[V_p^{5/2} - (V_p - V_f)^{5/2}] \tag{8.55}$$

For purposes of computation Eq. (8.55) is best put in the form of the series

$$I = GlV_p^{3/2}\left[1 - \frac{3V_f}{4V_p} + \frac{3}{24}\left(\frac{V_f}{V_p}\right)^2 + \frac{3}{192}\left(\frac{V_f}{V_p}\right)^3 + \cdots\right] \tag{8.56}$$

All the above equations for current can be put into the form

$$I = GlV_p^{3/2}F\left(\frac{V_p}{V_f}\right) \tag{8.57}$$

in which $F\left(\frac{V_p}{V_f}\right)$ is the current ratio plotted in Fig. 8.11.

Some difficulty is occasionally encountered in calculating currents in tubes with filaments and plane electrodes. Neither the plane- nor the cylindrical-electrode formulas will fit directly here. Experimentally it is found that there is an equivalent filament area which may be applied. This equivalent area is obtained by projecting the filament onto the filament plane and surrounding it by a band twice as wide as the distance from the filament to the nearest electrode. This equivalent area may be used either to obtain current in diodes or to obtain current or mutual conductance in triodes. The same concept may be applied to helical filaments in cylindrical-electrode tubes.

8.9. Effect of Initial Electron Velocity. In all the foregoing analyses it has been assumed that the electrons start with zero velocity from a point of zero potential. This is not quite correct because of the mechanism of electron emission. Actually, electrons having zero velocity would never get started from a cathode in the presence of space-charge saturation. The electrons come off from the emitter with a Maxwellian distribution of velocities ranging from zero to infinity. The distribution is such that 90 per cent of the electrons have velocities below ½ volt at usual cathode temperatures, and fewer and fewer have successively higher velocities.

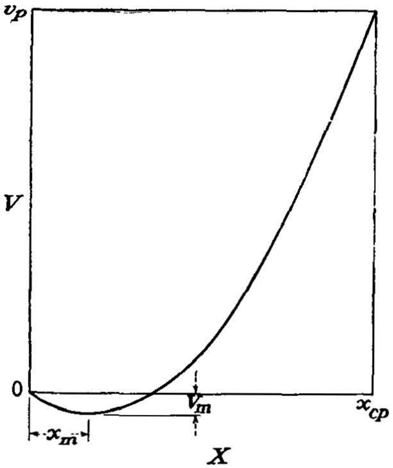


FIG. 8.12.—Potential distribution in a plane-electrode diode for the case of uniform initial velocity of emission.

A good idea of what the actual potential distribution is when the initial velocities are considered may be obtained by assuming that all electrons leave the cathode with the same normal velocity. Because of the initial velocity, the gradient at the cathode may and does become negative, and the potential curve moves down until it has a minimum somewhere close to the cathode at which the potential is negative and corresponds to the velocity of emission. For this condition the electrons are slowed down until they all come to rest at the potential minimum. From this point, which acts like an ideal cathode and which is called a "virtual" cathode, the electrons may start in either direction, either being returned to the cathode or going to the plate.

For a condition of space-charge saturation the potential distribution on either side of the virtual cathode will follow the four-thirds-power law. The location of the potential minimum will be determined by the fraction of the current that goes on to the plate. A typical potential distribution is shown in Fig. 8.12.

Let it be assumed that the fraction of the emitted current which continues on to the plate is given by P. Then if the emitted current per unit area is J_e, the current to the right of the potential minimum is PJ_e and that to the left is $(2 - P)J_e$. The current to the left of the minimum is made up of the emitted current J_e going in one direction and the returning current $(1 - P)J_e$ going in the other. As far as the space-charge effects are concerned, the directions of these current components are immaterial because the charge densities add regardless of sign.

The relation between current, potential, and distance to the right of the minimum is given by

$$PJ_e = \frac{2.335 \times 10^{-6}(V + V_m)^{3/2}}{(x - x_m)^2} \quad \text{amperes per unit area} \tag{8.58}$$

where V is potential with cathode potential taken as zero, V_m is the magnitude of the negative potential at the minimum, x is the distance measured from the cathode, and x_m is the distance from the cathode to the minimum.

Similarly the relation to the left of the minimum is

$$(2 - P)J_e = \frac{2.335 \times 10^{-6}(V_m + V)^{3/2}}{(x_m - x)^2} \quad \text{amperes per unit area} \tag{8.59}$$

where V and V_m are magnitudes of potential. Equations (8.58) and (8.59) give the potential distribution if the magnitude and the location of the potential minimum and the fraction of the transmitted current are known.

To determine the factors in terms of which Eqs. (8.58) and (8.59) are expressed let these two equations be evaluated at the plate and cathode, respectively. Then

$$PJ_e = \frac{2.335 \times 10^{-6}(V_p + V_m)^{3/2}}{(x_{cp} - x_m)^2} \quad \text{amperes per unit area} \tag{8.60}$$

and

$$(2 - P)J_e = \frac{2.335 \times 10^{-6}(V_m)^{3/2}}{x_m^2} \quad \text{amperes per unit area} \tag{8.61}$$

Taking the ratio of these two expressions and solving for x_m,

$$x_m = \frac{x_{cp}}{1 + \left(\frac{V_p}{V_m} + 1\right)^{3/4} \left(\frac{2}{P} - 1\right)^{1/2}} \tag{8.62}$$

from which the location of the potential minimum can be determined for any assumed fraction of transmitted current. The magnitude of the potential at the minimum is known from the initial velocity.

The results of the above analysis are not exact because in the actual case there is a distribution of velocities, with the result that the position and magnitude of the potential minimum and also the fraction of the transmitted current are uniquely determined from the potential and cathode temperature. An exact analysis considering the velocity distribution is given by Langmuir.[1] The exact expressions are quite involved, but some approximate expressions which are accurate to within about 2 per cent take the following form:

$$PJ_e = \frac{2.335 \times 10^{-6}(V_p - V_m)^{3/2}}{(x_{cp} - x_m)^2}\left[1 - \frac{0.0247T^{1/2}}{(V_p - V_m)^{1/2}}\right] \tag{8.63}$$

where J_e is emitted current per unit area
PJ_e is current per unit area passing potential minimum
V_p is plate potential
V_m is magnitude of minimum potential relative to cathode
x_{cp} is cathode-plate distance
x_m is cathode–potential-minimum distance
T is cathode temperature, °K

The location of the potential minimum is given by

$$x_m = 0.0156(1{,}000J)^{-1/2}\left(\frac{T}{1{,}000}\right)^{3/4} \tag{8.64}$$

The corresponding magnitude of potential minimum is

$$V_m = \left(\frac{T}{5{,}040}\right)\log_{10}\left(\frac{1}{P}\right) \tag{8.65}$$

where P is the fraction of the emitted current transmitted to the plate.

The exact relations for space-charge-saturated flow with initial electron velocity are given by Langmuir in the form of the universal curve I of Fig. 8.13. This curve gives potential as a function of distance with the origin arbitrarily taken at the potential minimum. For comparison there are also shown the potential-distribution curves of Eqs. (8.63) and (8.8) as curves II and III, respectively. It is seen that the actual potential distribution is considerably different from that of the Child-Langmuir law beyond the potential minimum and is totally different from a four-thirds-power law to the left of the potential minimum except in its immediate vicinity.

A study of the approximate relations given above and of the universal potential-distribution curve reveals the following effects of initial velocity

---

[1] LANGMUIR and COMPTON, *op. cit.*, Part II, p. 241.

upon space-charge flow: The larger the emitted current and the cathode temperature, the greater the magnitude of the potential minimum. The lower the cathode temperature and the larger the plate potential, the closer the potential minimum approaches the cathode and the lower it becomes. The plate current considering initial velocity is larger than that obtained from the Child-Langmuir law because the potential difference between the virtual cathode at the potential minimum and the plate is greater than the actual plate potential, because the distance from the virtual cathode to the plate is less than the actual cathode-plate distance, and because the electrons leave the virtual cathode with an average velocity that is greater than zero.

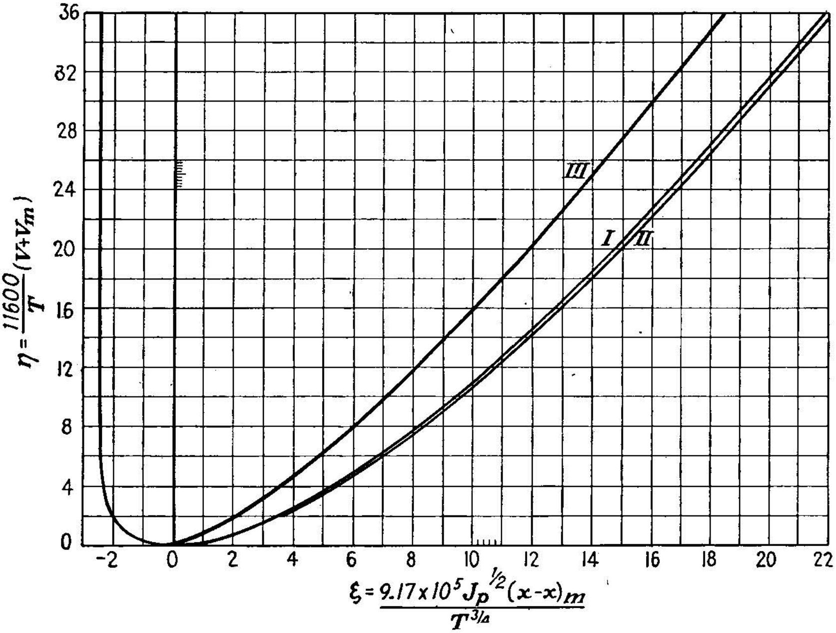


FIG. 8.13.—Universal diode potential-distribution curve including effect of Maxwellian distribution of emission velocity (Langmuir). (See discussion on page 193.)

The distance of the virtual cathode from the actual cathode may be appreciable. For a cathode temperature of 1000°K and a transmitted current density of 1 ma per cm^2, the distance from the cathode to the virtual cathode is approximately 0.006 in. In modern close-spaced electrode tubes, this distance is by no means inappreciable.

When the fraction of the emitted current transmitted beyond the virtual cathode is not known, it is necessary to solve Eqs. (8.63) to (8.65) by trial to determine the transmitted current for a given electrode spacing, plate potential, and cathode temperature.

8.10. Effect of Space Charge upon Transit Time in Diodes. In general, the transit time in an electric field is given by the integral of the reciprocal of velocity with respect to distance.

$$T = \int_{x_1}^{x_2} \frac{dx}{v} \tag{8.66}$$

For the plane-electrode diode the transit time with and without space charge is easily determined. Without space charge the potential profile is a straight line so that

$$V(x) = \frac{x}{d_{cp}} V_p \tag{8.67}$$

where $V(x)$ is the potential at any distance x from the cathode, V_p is the plate potential, and d_{cp} is the cathode-plate distance. The velocity at any point, *assuming zero initial velocity*, is then given by

$$v(x) = \left(\frac{x}{d_{cp}}\right)^{1/2} v_p \tag{8.68}$$

so that the transit time is

$$T = \frac{d_{cp}{}^{1/2}}{v_p} \int_0^{d_{cp}} \frac{dx}{x^{1/2}} \tag{8.69}$$

with the result that

$$T = \frac{2d_{cp}}{v_p} \tag{8.70}$$

When space charge is present in the plane-electrode diode, then the potential follows a four-thirds-power law so that

$$V(x) = \left(\frac{x}{d_{cp}}\right)^{4/3} V_p \tag{8.71}$$

The velocity at any point is then given by

$$v(x) = \left(\frac{x}{d_{cp}}\right)^{2/3} v_p \tag{8.72}$$

so that the transit time is

$$T = \frac{d_{cp}{}^{2/3}}{v_p} \int_0^{d_{cp}} x^{-2/3}\, dx \tag{8.73}$$

with the result that

$$T = \frac{3d_{cp}}{v_p} \tag{8.74}$$

This is seen to be of the same form as for the space-charge-free case, the only difference being that the time is 50 per cent greater.

For the cylindrical-diode case in the absence of space charge the potential profile is a logarithmic function.

$$V = \frac{\ln\left(\frac{r}{r_c}\right)}{\ln\left(\frac{r_p}{r_c}\right)} V_p \tag{8.75}$$

where r is the radial distance to any point and r_c and r_p are cathode and plate radius, respectively. The velocity at any point is

$$v = \sqrt{\frac{\ln\left(\frac{r}{r_c}\right)}{\ln\left(\frac{r_p}{r_c}\right)}}\, v_p \tag{8.76}$$

The transit time for this case is

$$T = \frac{\sqrt{\ln\left(\frac{r_p}{r_c}\right)}}{v_p} \int_{r_c}^{r_p} \frac{dr}{\sqrt{\ln\left(\frac{r}{r_c}\right)}} \tag{8.77}$$

If the substitution

$$u = \ln\left(\frac{r}{r_c}\right) \tag{8.78}$$

be made, then the transit time is

$$T = r_c \frac{\sqrt{\ln\left(\frac{r_p}{r_c}\right)}}{v_p} \int_0^{\ln\left(\frac{r_p}{r_c}\right)} \frac{\epsilon^u}{\sqrt{u}}\, du \tag{8.79}$$

This is now in a form which can readily be evaluated by series integration and in which it is apparent that the integral is not infinite. The results may be expressed in the form

$$T = \frac{d}{v_p} B\left(\frac{r_p}{r_c}\right) \tag{8.80}$$

where d is the distance between plate and cathode and $B\left(\frac{r_p}{r_c}\right)$ is a function of the ratio of plate to cathode radius given in Fig. 8.14. It is seen that when the cathode is inside the transit time is less than that in the plane-electrode case for the same distance and potential.

When there is space charge present in the cylindrical diode, the potential profile is given by

$$V = \left(\frac{r\beta_{rc}{}^2}{r_p\beta_{cp}{}^2}\right)^{2/3} V_p \tag{8.81}$$

where r is the radius to any point and β^2 is the function given in Fig. 8.4, the subscripts indicating the distances determining the ratio for which the function is evaluated. The velocity is

$$v = \left(\frac{r\beta_{rc}{}^2}{r_p\beta_{cp}{}^2}\right)^{1/3} v_p \tag{8.82}$$

The integral for the transit time is now

$$T = \frac{(r_p\beta_{cp}{}^2)^{1/3}}{v_p} \int_{r_c}^{r} (r\beta_{rc}{}^2)^{-1/3}\, dr \tag{8.83}$$

If again the substitution of Eq. (8.78) be made,

$$T = \frac{(r_p\beta_{cp}{}^2)^{1/3} r_c{}^{2/3}}{v_p} \int_0^{\ln\left(\frac{r}{r_c}\right)} \left(\frac{\epsilon^u}{\beta_{rc}}\right)^{2/3} du \tag{8.84}$$

This is readily evaluated numerically for small values of $\frac{r}{r_c}$. For large values of $\frac{r}{r_c}$ the form

$$T = \frac{3}{2}\left(\frac{r - r_c}{v}\right) \frac{\frac{d}{du}(\epsilon^u\beta_{rc}{}^2)}{\epsilon^u - 1} \tag{8.85}$$

is more suitable for computation. The results of the computation can be put into the form[1]

$$T = \frac{d}{v_p} A\left(\frac{r_p}{r_c}\right) \tag{8.86}$$

in which the function $A\left(\frac{r_p}{r_c}\right)$ is that shown in Fig. 8.14. In this figure it is seen that in the cylindrical diode with space charge and an internal cathode the transit time is less than for the corresponding plane electrode but more than for the same case not space-charge-limited. In the

[1] FERRIS, W. R., Input Resistance of Vacuum Tubes as Ultra-high-frequency Amplifiers, *Proc. I.R.E.*, vol. 24, pp. 82–107, January, 1936.

curves of Fig. 8.14 are also included the values for cylindrical diodes with the cathode outside. These are seen to have larger transit times than the plane-electrode diode, which in turn has larger transit times than the cylindrical diode with the cathode inside.

8.11. Summary. The primary effect of space charge in a tube is to make the transmitted current follow a three-halves-power law of plate voltage. In addition, it makes the plate current virtually independent of the filament voltage. Modern tubes are designed so that the emission at rated voltages is more than sufficient to supply the current required by the Child-Langmuir law. Under these conditions the emission is said to be *space-charge-limited.* The nature of this saturation is shown in Fig. 8.15. Here is shown the variation of the plate current in a diode with cathode heating power for different voltages. If the cathode emission is not very great, a departure from the three-halves-power law of voltage occurs at relatively low voltage. This occurs when the plate is collecting all the current emitted from the cathode and gives rise to what is known as *temperature saturation.* This effect is shown in Fig. 8.16. Here is shown the plate current in a diode as a function of plate voltage for various cathode powers. For low voltages the curve follows the three-halves-power law, and then at some point determined

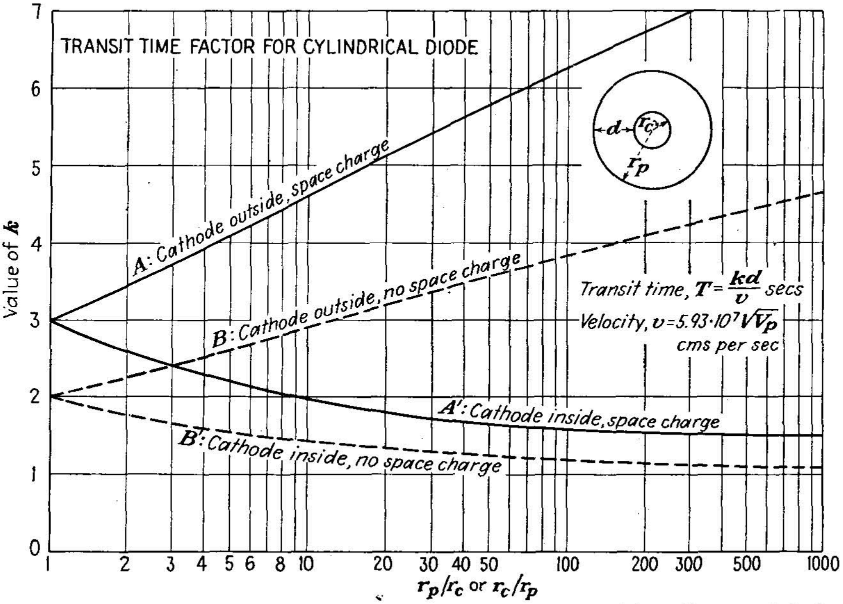


FIG. 8.14.—Transit-time curves of a cylindrical diode (Ferris). The parallel-plane case is given by $r_p/r_c = 1$.

by the cathode emission the current becomes nearly independent of the plate voltage. The nature of the saturation in this case depends upon the type of emitter. With tungsten as an emitter, the emission is very nearly independent of plate voltage in the saturation range. With thoriated tungsten and even more so with oxide emitters the emission increases slowly with plate voltage in the saturation range. This is because the emission depends upon the gradient at the cathode (Schottky effect). In the case of oxide emitters the increasing gradient through the emitter increases the liberation of emitting material in addition. In a well-designed tube, temperature-saturation effects will not occur at rated voltages.

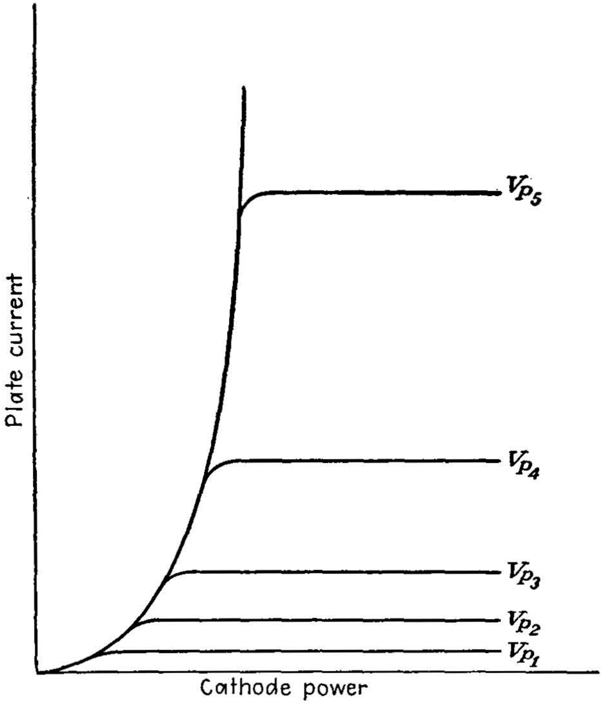


FIG. 8.15.—Diode plate current as a function of cathode power.

In addition to the three-halves-power law of voltage and the saturation effects mentioned above, space charge has the effect of reducing the capacity between electrodes. The capacity between the cathode and plate of a space-charge-saturated diode is three-fifths of that of the cold diode.[1] Further, transit times are in general increased over the diode

---

[1] LLEWELLYN, F. B., "Electron Inertia Effects," p. 50, Cambridge, London, 1941.

without space charge, being 50 per cent greater in the space-charge-saturated diode with plane electrodes.

Power relations are unchanged. Although the voltage and gradient distribution are different in the presence of space charge and in its absence, the velocity of an electron is always the same relative to the potential, as is required by the energy equation (6.4). The power put into the tube, as evidenced by the product of the voltage across it and the current through it, appears in the form of heat at the plate. The energy of impact of each electron is $\frac{1}{2}mv^2$, and the number of electrons striking the plate per unit area per second is $\frac{\rho v}{e}$. The product of these is recognized as JV per unit area.

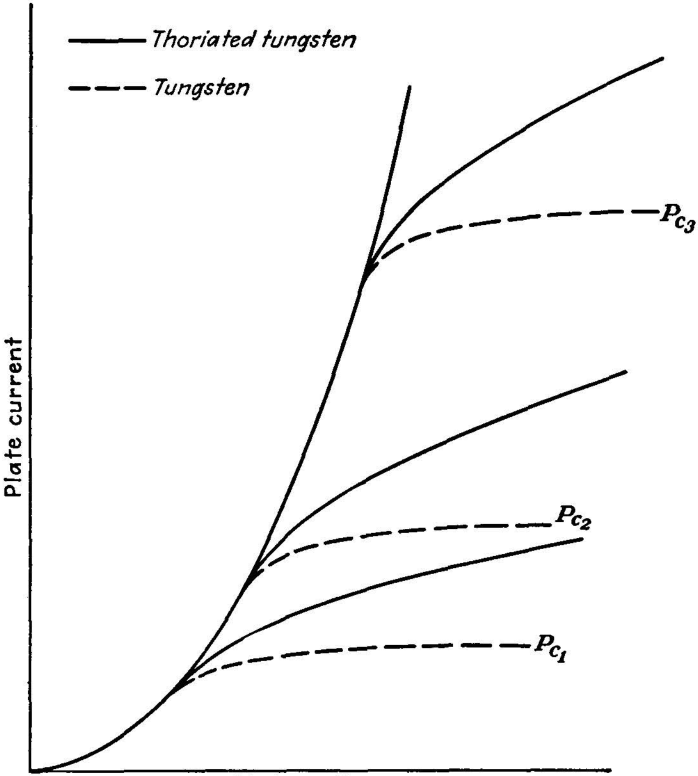


FIG. 8.16.—Diode plate current as a function of plate voltage.

CHAPTER 9

TRIODE CHARACTERISTICS

9.1. Control Action of the Grid. The triode in its commonest form consists of an emitter surrounded by a grid, which in turn is surrounded by a plate. The grid is usually a mesh of fine wires supported quite close to the cathode. The plate is spaced several times as far away. The entire structure is supported in an evacuated glass or metal envelope with leads to the electrodes coming out through glass on the bottom of the tube.

The outstanding feature of the triode is the ability of the grid to control the flow of current to the plate without itself drawing any current. As a result of this property, a small voltage on the grid is capable of producing a large voltage drop in the plate circuit. Because of the fact that the grid draws no current, the triode, at all but very high frequencies, is a voltage-operated device in that virtually no power is required to operate the tube. The term "electric valve" for a vacuum tube is particularly expressive because the grid has an electrical valve action. Vacuum tubes do not really amplify power. Actually, the grid controls the flow of power from the plate power supply.

In the chapter on Electrostatic Field of a Triode it was shown that both the plate and the grid electrodes were able to control the gradient of potential in front of the cathode. It was also shown that the grid was much more effective in so controlling the off-cathode gradient, in fact, μ times as effective. When a tube is conducting, the negative charge of the electrons passing through the tube produces a space charge that alters the potential distribution in the tube, particularly in the vicinity of the cathode, but the control property of the grid is not impaired. The potential distribution between the cathode and grid, for usual combinations of potentials, is now a curve that is concave upward (for plane electrodes) instead of being nearly a straight line. The positive plate potential reaches through the grid and causes the potential on the cathode side of the grid to be positive. Electrons are drawn off the cathode into this region of positive potential and are drawn to the positive plate between the negative-grid wires from whose *immediate* vicinity, however, they are strongly repelled. The negative control grid easily regulates the degree of positiveness of the potential before the cathode. This the grid is able to do primarily because of its greater proximity to the cathode.

9.2. Current-voltage Characteristics of the Triode. The plate current in a triode depends upon both the plate and the grid voltage. It also depends upon the filament voltage, but this is usually held at some suitable fixed value. Hence it is usual to describe the current characteristics of a triode in terms of grid and plate potential alone.

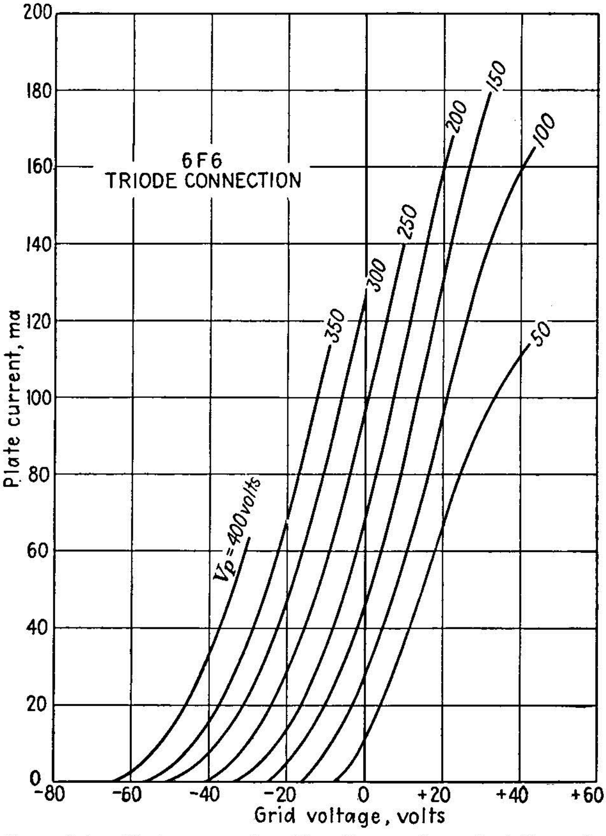


FIG. 9.1.—Plate-current–grid-voltage characteristics of a triode.

Plate-current–Grid-voltage Characteristics. As mentioned in the previous chapter, the space current in a triode is given approximately by

$$I_s = I_g + I_p = G\left(V_g + \frac{V_p}{\mu}\right)^a \tag{9.1}$$

where I_s is space current
I_p is plate current
I_g is grid current

G is perveance
V_g is grid voltage
V_p is plate voltage
μ is amplification factor
a is a constant, approximately $\frac{3}{2}$

When the grid voltage is negative, the entire space current goes to the plate. Even when the grid is positive, the fraction of the space current going to the grid is small so that Eq. (9.1) is a reasonably good approximation for plate current under all conditions except the combination of very small plate voltage and rather large positive grid voltage. An actual plot of plate current as a function of grid voltage is shown in Fig. 9.1. The principal characteristics are quite evident and consistent with expectations. The plate current is seen to increase with both grid and plate voltage but more rapidly with grid voltage. Considering the variation with grid voltage alone, the current increases slowly at first and then more rapidly. In the region of negative grid voltages the curves for the different plate voltages are seen to be of nearly the same shape but merely displaced horizontally. This is consistent with the form of Eq. (9.1). For positive grid voltages the rate of increase of current with grid voltage shows a slight decrease. This is due to two factors. (1) The grid is beginning to draw a fraction of the total, or space, current. (2) There is a tendency for the current to saturate at large plate voltages.

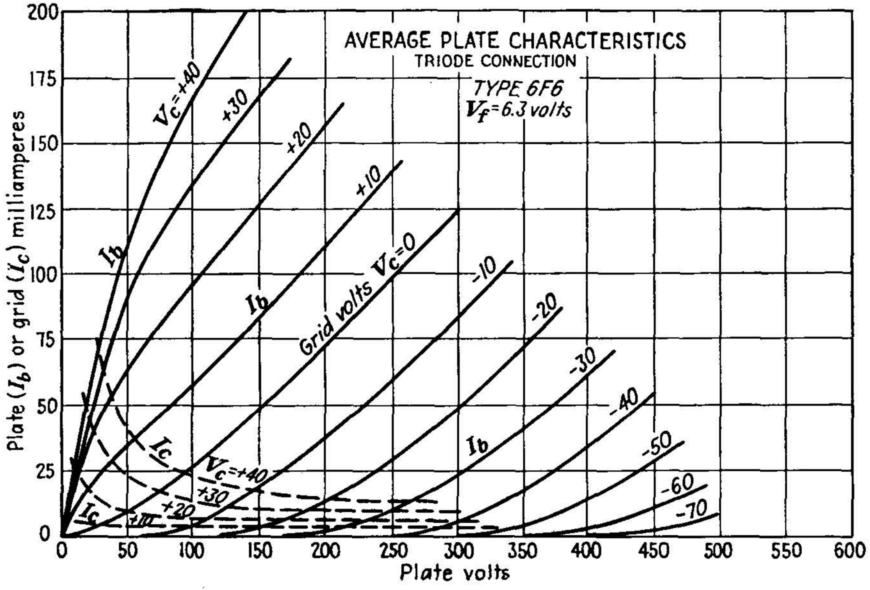


FIG. 9.2.—Plate-current–plate-voltage characteristics of a triode.

Plate-current–Plate-voltage Characteristics. A typical set of plate-current–plate-voltage characteristics of a triode is shown in Fig. 9.2. The same general properties observed in the plate-current–grid-voltage characteristics are observable here. For negative grid voltages, however, the curves are not similar in this representation. This is because for large negative grid voltages the tube is operating near cutoff and here the amplification factor of the tube is appreciably lower than for grid voltages near zero. The change in shape of the curves for positive grid voltages from concave upward to concave downward is due to the diversion of part of the space current to the grid. The actual space current still has an upward curvature with plate voltage, but the fraction of space current taken by the grid decreases as the plate voltage increases.

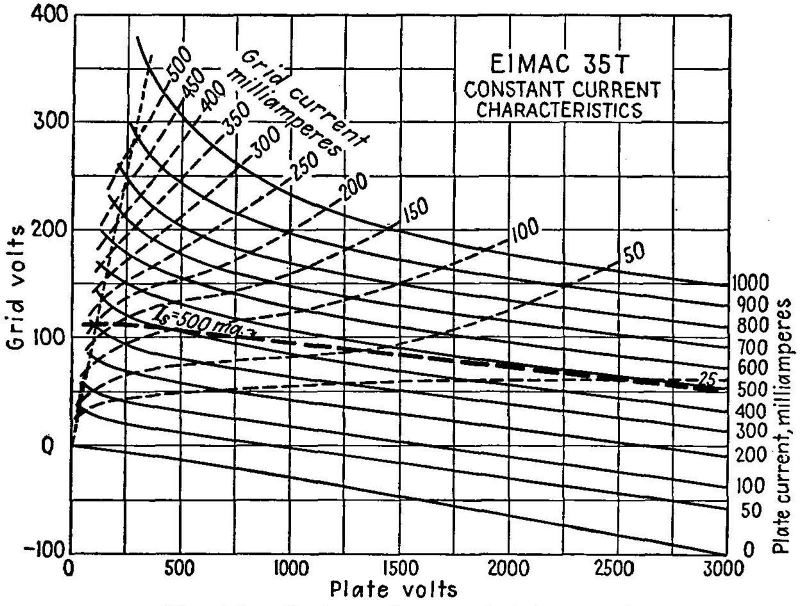


FIG. 9.3.—Contours of constant plate current.

Contours of Constant Plate Current. In Fig. 9.3 are shown contours of constant plate current plotted along axes of plate voltage and grid voltage. These curves show the combinations of grid and plate potential for which the plate current is constant. Over a large part of their range these contours are parallel straight lines. The significance of this is that the amplification factor of the tube is very nearly constant. For small values of plate voltage and positive grid voltages the curves are curved strongly upward. This is due to the diversion of part of the space current to the grid. If contours of constant space current are plotted (one is shown dashed in the figure), they are found to be nearly straight lines.

The Plate-current Surface. Inasmuch as plate current is a function of two variables, it may be represented as a surface. The height of this surface above a reference plane is given by the magnitude of the plate current. Position on the reference plane and on the corresponding point on the surface above is given by the plate and grid voltage. A sketch of the plate-current surface is shown in Fig. 9.4. The relation of the surface to the three representations of plate current previously given is evident from the figure. The plate-current–grid-voltage curves are the intersections of planes parallel to the plate-current and grid-voltage axes with the surface.

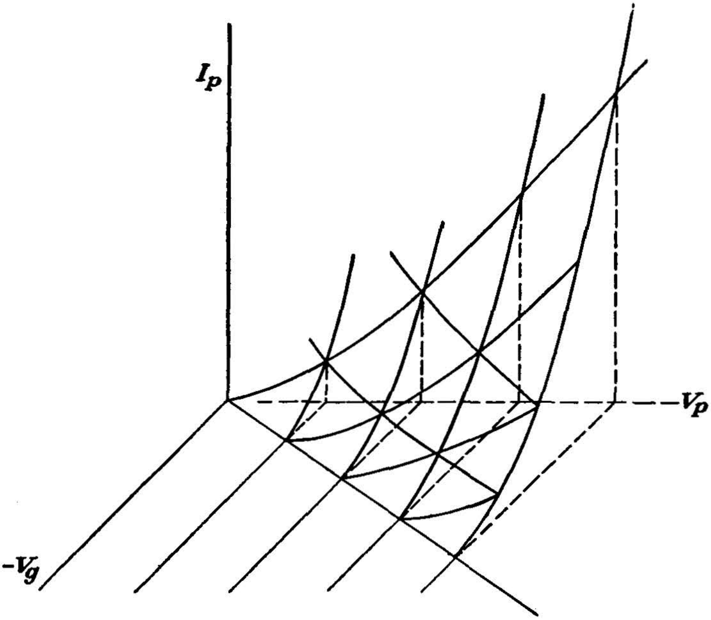


FIG. 9.4.—The plate-current surface.

Only a part of the surface is shown, to avoid confusion due to too many lines. The surface becomes a horizontal plateau for large values of current due to voltage saturation, *i.e.*, insufficient emission. Since there can be no plate current for negative plate voltages, the surface turns a corner as it approaches zero values of plate voltage.

9.3. Definition of Triode Constants. *Amplification Factor.* Although the characteristics of a triode are completely specified only by a set of voltage-current curves, an index of the tube's operation is ordinarily given in terms of the so-called tube "constants." These so-called constants suffice to describe the operation of a tube in the vicinity of a given set of electrode potentials. The three so-called tube constants are the amplification factor, the mutual conductance, and the plate resistance.

The amplification factor is a function of the electrode geometry and has already been given in terms of the dimensions. It has previously been given for a cold tube as the relative effectiveness of the plate and grid potentials in controlling the off-cathode gradient of potential. Another definition and the one generally accepted is that *amplification factor is the relative effectiveness of the plate and grid potentials in controlling the plate current.* Mathematically this is given by

$$\mu = -\frac{\dfrac{\partial I_p}{\partial V_g}}{\dfrac{\partial I_p}{\partial V_p}} = -\left(\frac{dV_p}{dV_g}\right)_{I_p \text{ const}} \tag{9.2}$$

If a tube is conducting current under a given condition of potentials and the plate voltage is then increased by a small amount, the plate current will increase by a small amount. If then the grid voltage is made more negative by the proper amount, the current will be restored to its original value. The limit of the ratio of the change in plate voltage to the change in grid voltage necessary to keep the plate current constant as these changes are made vanishingly small is the amplification factor of the tube. This is the significance of Eq. (9.2). The amplification factor of a tube whether a triode or multielectrode tube is always taken with respect to the control-grid voltage unless otherwise specified. In triodes the amplification factor is a measure of the voltage-amplifying capabilities of the tube. In multielectrode tubes it has no great significance and is usually not even listed. Amplification factor is a dimensionless constant. Practical values of amplification factor run from 2.5 to 200 in ordinary triodes.

Mutual Conductance. The mutual conductance (sometimes called the "transconductance") of a triode has already been referred to in the chapter on Space-charge Effects. *The mutual conductance of a tube is the rate of change of plate current with control-grid voltage.* Mathematically this is given by

$$G_m = \frac{\partial I_p}{\partial V_g} = \left(\frac{dI_p}{dV_g}\right)_{V_p \text{ const}} \tag{9.3}$$

An increase in the grid voltage of a tube effects an increase in the plate current. The limit of the ratio of the changes as the change in grid voltage is made vanishingly small is the mutual conductance. It will be remembered that the mutual conductance of a triode is directly proportional to the square root of the equivalent voltage of the tube. It is also proportional to the cube root of the space current. As the name implies, the dimensions of this constant are those of conductance. Mutual conductance is usually expressed in units of micromhos or microamperes per volt. Practical values of mutual conductance are between 100 and 10,000 micromhos.

Plate Resistance. Another tube constant that is commonly used is the plate resistance, also known as the "variational plate resistance" or the "dynamic plate resistance." *The plate resistance of a tube is the reciprocal of the rate of change of plate current with plate voltage.* Mathematically it is given by

$$R = \frac{1}{\dfrac{\partial I_p}{\partial V_p}} = \frac{\partial V_p}{\partial I_p} = \left(\frac{dV_p}{dI_p}\right)_{V_g \text{ const}} \tag{9.4}$$

The plate resistance of a tube is the a-c resistance of the plate circuit to a small alternating voltage superimposed upon the direct voltage. The dimensions of this constant are those of resistance, and the magnitude is usually expressed in ohms. The plate resistance of a triode may vary from 1,000 to 50,000 ohms. A typical value is of the order of 5,000 ohms.

Relation between Tube Constants. The three tube constants express relations between the quantities that determine triode operation, *viz.*, plate current, plate voltage, and grid voltage. Since the three constants are expressed in terms of only three variables, it is expected that there is a relation between them. This is the case. If plate and grid voltage are changed by small amounts, the corresponding change in plate current is

$$dI_p = \frac{\partial I_p}{\partial V_g}\, dV_g + \frac{\partial I_p}{\partial V_p}\, dV_p \tag{9.5}$$

or

$$dI_p = G_m\, dV_g + \frac{1}{R_p}\, dV_p \tag{9.6}$$

If the change in plate current is held to zero, then

$$G_m R_p = -\left(\frac{dV_p}{dV_g}\right)_{I_p \text{ const}} = \mu \tag{9.7}$$

This relation, *viz.*, that the product of the mutual conductance and the plate resistance is equal to the amplification factor, holds exactly for any tube for any combination of electrode potentials.

Variation of Tube "Constants." The rather paradoxical heading of this subsection is justified by the fact that the so-called tube constants are not constants at all except approximately so in the vicinity of some operating condition. Actually, the constants may vary considerably over the entire range of voltages and currents in a tube. Of the so-called tube constants, the amplification factor varies the least. This is because it is basically dependent upon the geometrical structure of the tube. If the tube were perfect in that there were no end effects and no asymmetries and if no other electrode but the plate drew current, the amplification factor would not vary at all.

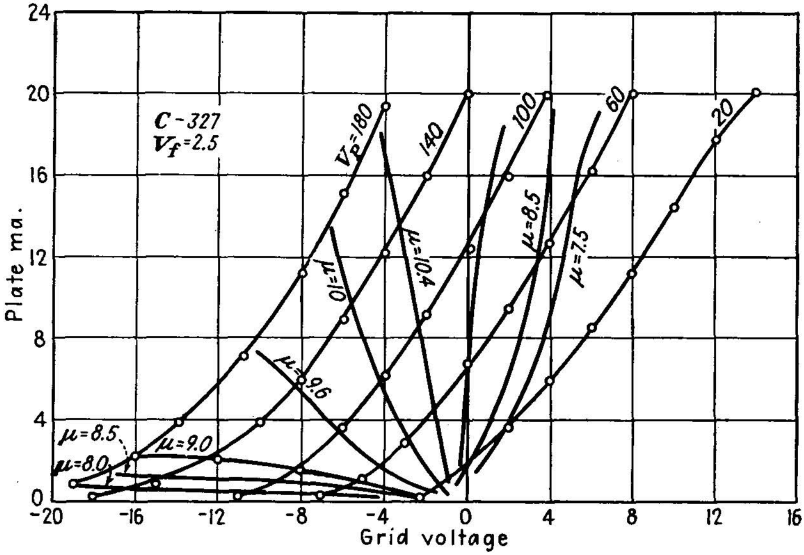


FIG. 9.5.—Contours of constant amplification factor of a triode.

From the definition of amplification factor it is seen that it is given by the negative reciprocal of the slope of the contours of constant plate current in Fig. 9.3. Inspection of the plate-current contours with this in mind reveals that for the most part the amplification factor is fairly constant. It tends to be somewhat low in the vicinity of very low plate currents and even more so in the vicinity of very low plate voltages. A better idea of the nature of the variation of the amplification factor is given in Fig. 9.5, in which there are shown contours of constant amplification factor superimposed upon the plate-current–grid-voltage curves of a triode.[1] The amplification factor is seen to be fairly constant over the entire working range of the tube. Variations in magnitude do not exceed

---

[1] TERMAN, F. E., and A. L. COOK, Variation in the Amplification Factor of Triodes, *Proc. I.R.E.*, vol. 18, pp. 1044–1047, June, 1930

15 per cent of the mean value. The drop in amplification factor near cutoff is due to end effects in the tube. Because of stray electrostatic fields near the edges of the electrodes, these edges constitute a region of low amplification factor. Thus the actual tube consists of a large electrode area of constant mu in parallel with a small area of much smaller mu. As cutoff is approached, the low-mu portion of the tube cuts off last, giving the effect of a lower amplification factor. The reason why the amplification factor is lower for low plate voltages in the positive-grid region is that here the grid takes a very large portion of the space current. Since, as will be shown later in this chapter, the fraction of the space current going to the plate increases with plate voltage, a smaller increase in plate voltage relative to a decrease in grid voltage is needed to maintain the plate current constant when the grid voltage is positive than when it is not. This means that the amplification factor is lower for the conditions stated above.

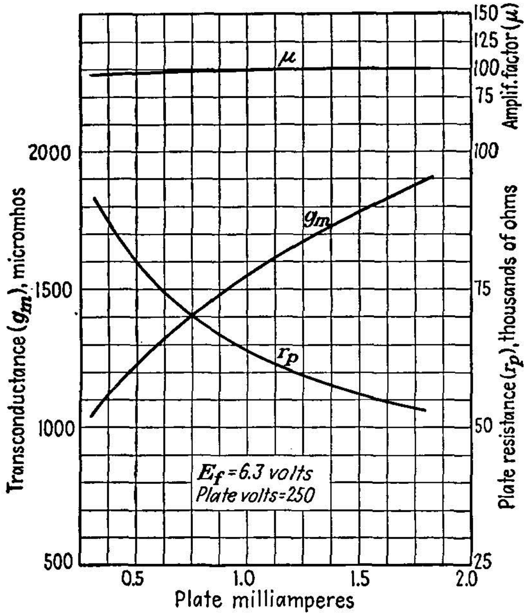


FIG. 9.6.—Variation with plate current of G_m, R_p, and μ of a Type 6F6 as triode.

From the definition, mutual conductance is seen to be equal to the slope of the plate-current–grid-voltage curves. Reference to these curves confirms that the mutual conductance is an increasing function of plate current in the region of negative grid voltages.

The mutual conductance of a triode has been shown in the chapter on Space-charge Effects to increase with the equivalent voltage and with the plate current. The nature of the variation is shown in Fig. 9.6. The variation is quite in accord with expectations as may be shown by plotting the variation of mutual conductance with plate current on logarithmic paper. Such a plot is given in Fig. 9.7 for the same tube. Here it is seen that, since the curve of mutual conductance as a function of plate current is nearly a straight line with a slope of one-third, the

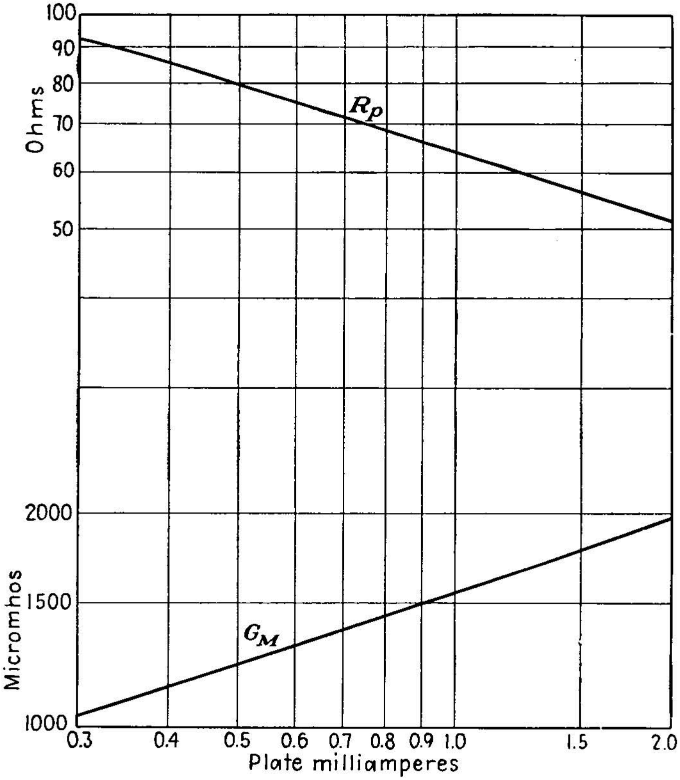


FIG. 9.7.—One-third-power-law variation of mutual conductance and plate resistance with plate current, 6F6 as triode.

mutual conductance follows very closely a one-third-power law of variation with plate current as was predicted in Eq. (8.48). The correspondence between the predicted and actual behavior is best at high currents. At low currents the variation departs somewhat from the one-third-power law because of the reduction in amplification factor.

Since the product of the mutual conductance and the plate resistance is equal to the amplification factor and since the amplification is almost constant, the plate resistance may be expected to vary with plate current in a fashion reciprocal to that in which the mutual conductance does. This is seen to be the case in Fig. 9.6. Examining the variation more critically, Fig. 9.7 shows that the plate resistance varies nearly as the negative one-third power of plate current. The plate resistance is the reciprocal of the slope of the plate-current–plate-voltage characteristics of a tube. Reference to Fig. 9.2 shows that the plate resistance decreases with increasing plate current in the negative-grid region.

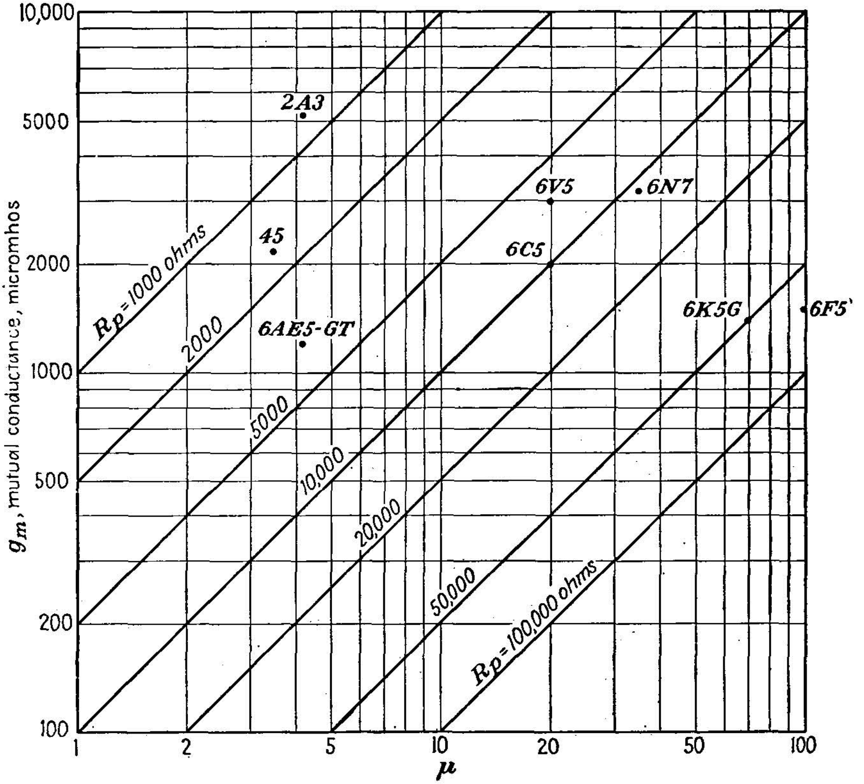


Fig. 9.8.—Constants of typical triodes.

For positive grid voltage the plate resistance may increase again. The plate resistance is lowest at high currents and low plate voltages.

The total range of all the tube constants within a single tube may be considerable. The amplification factor may vary over a range of 20 per cent. The plate resistance and the mutual conductance may vary over a range of three to one. The range of values of the constants encountered from triode to triode is even more considerable. In Fig. 9.8 are shown the tube constants for conditions of recommended operation of a number of triodes. The representation is such that the location of the point corresponding to each tube gives the three tube constants. High amplification factors, of the order of 100 or so, may be obtained, but at the expense of a low mutual conductance and a correspondingly high plate resistance.

9.4. Effective Tube Constants of Combinations of Tubes. It is of interest to consider what the effective tube constants are when two or more tubes are connected in parallel. Consider first the case of identical tubes connected in parallel. The effect of this is to double the mutual conductance, halve the plate resistance, and leave the amplification factor unchanged. This is logical, for with two tubes contributing current an increase in grid voltage produces twice as much of an increase in plate current as does a single tube. This explains the doubling of the mutual-conductance value. Since the variations in plate current for a given change in plate voltage are twice as great as for a single tube, the plate resistance is half as great. The amplification factor is unchanged because the product of mutual conductance and plate resistance is the same as for a single tube.

If tubes with different characteristics are connected in parallel, the combination characteristics are still readily determined. The effective mutual conductance is simply the sum of the individual mutual conductances since the plate currents add directly.

$$G_{m\,\mathrm{equiv}} = G_{m1} + G_{m2} + \cdots + G_{mn} \tag{9.8}$$

The equivalent plate resistance is obtained by adding the individual plate resistances as one adds resistances in parallel.

$$\frac{1}{R_{p\,\mathrm{equiv}}} = \frac{1}{R_{p1}} + \frac{1}{R_{p2}} + \cdots + \frac{1}{R_{pn}} \tag{9.9}$$

In other words, the equivalent plate conductance, reciprocal of plate resistance, is the sum of the individual plate conductances.

The equivalent amplification factor is given by the product of the equivalent mutual conductance and the equivalent plate resistance as given by Eqs. (9.8) and (9.9). For the special case of two tubes in parallel the expression for the equivalent amplification factor reduces to

$$\mu_{\mathrm{equiv}} = \frac{\mu_1 R_{p2} + \mu_2 R_{p1}}{R_{p1} + R_{p2}} \tag{9.10}$$

The equivalent amplification factor of tubes in parallel may be higher or lower than one of the individual values but will lie within the extreme values. The equivalent amplification factor will generally decrease as the grid is made more negative. This is because the high-mu tubes

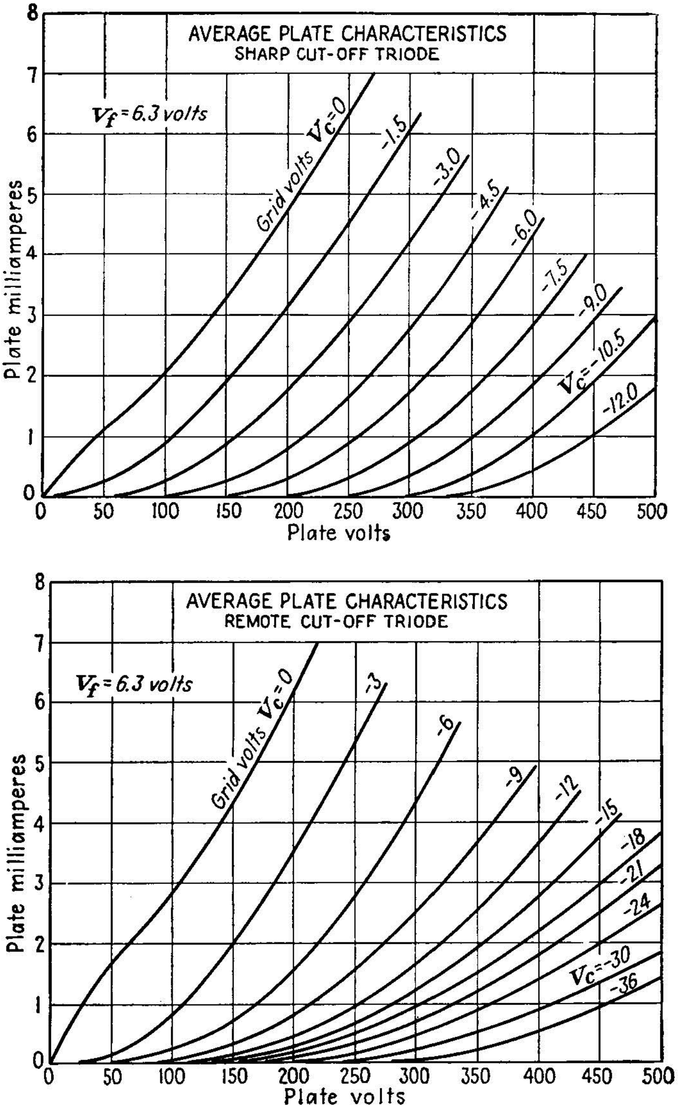


FIG. 9.9.—Comparison of plate-current–plate-voltage curves of the sharp- and remote-cutoff triode of a 6AE6G.

will cut off first, leaving the low-mu tubes to carry current alone. The action may be illustrated in the case of two tubes by studying the rate of change of the equivalent mu with respect to grid voltage. Holding the individual mu's constant in Eq. (9.10),

$$\frac{\partial \mu_{\text{equiv}}}{\partial V_g} = \frac{(\mu_2 - \mu_1)(R_{p2})^2 \dfrac{d}{dV_g}\left(\dfrac{R_{p1}}{R_{p2}}\right)}{(R_{p1} + R_{p2})^2} \tag{9.11}$$

The nature of the rate of change of the equivalent mu with respect to grid voltage is, assuming $\mu_2 > \mu_1$, determined only by the sign of $\frac{d}{dV_g}\left(\frac{R_{p1}}{R_{p2}}\right)$. If, further, the mutual conductances of tubes are the same, then the slope of $\frac{R_{p1}}{R_{p2}}$ will be positive. As a result, the equivalent amplification factor will increase as grid voltage is increased.

The above type of analysis may be extended to the case of more tubes in parallel. It may also be extended to tubes in which the amplification factor is not constant along the cathode surface. Tubes having grids wound with a variable pitch are often used to obtain an amplification factor that decreases with increasing bias. Such tubes are extensively used in r-f amplifiers for automatic volume control.[1] Such tubes, also known as "remote-cutoff tubes" or "supercontrol tubes," have the characteristics of a low-mu tube at low plate currents and of a high-mu tube at large currents.

Such tubes have a tremendous variation of mutual conductance as well as the variation in amplification factor. This large variation results from the combination of the normal increase in mutual conductance with current and the increase in amplification factor with equivalent voltage. In Fig. 9.9 are compared the plate-current–plate-voltage characteristics of triodes which are identical except for the fact that one has a constant-pitch grid, whereas the other has a variable-pitch grid. The reason for the designation "remote cutoff" as contrasted with "sharp cutoff" is apparent.

9.5. Electron Paths. In the previous discussion, tube characteristics have been studied without reference to the electron paths. This has been possible because from space-charge considerations it is possible to determine the number of electrons transmitted past the virtual cathode in front of the actual cathode. For negative grid voltages, all the electrons leaving the cathode will be transmitted to the plate. For positive grid voltages, however, part of the emitted current is intercepted by the grid, and here the actual electron paths are of interest. Electron paths are also of interest in multielectrode tubes, where they have a considerable part in determining the tube characteristics.[2]

---

[1] BALLANTINE, STUART, and H. A. SNOW, Reduction of Distortion and Cross-talk in Radio Receivers by Means of Variable-mu Tetrodes, *Proc. I.R.E.*, vol. 18, pp. 2102–2127, December, 1930.

[2] THOMPSON, A. C., Electron Beams and Their Application in Low Voltage Devices, *Proc. I.R.E.*, vol. 24, pp. 1276–1297, October, 1936.

In Fig. 9.10 are shown electron paths in a triode operating with a negative grid. These curves were obtained by photographing the motion of small balls rolled upon a suitably deformed elastic membrane. Such a model of potential takes no account of space-charge effects. It may be expected that in this case the presence or absence of space charge will

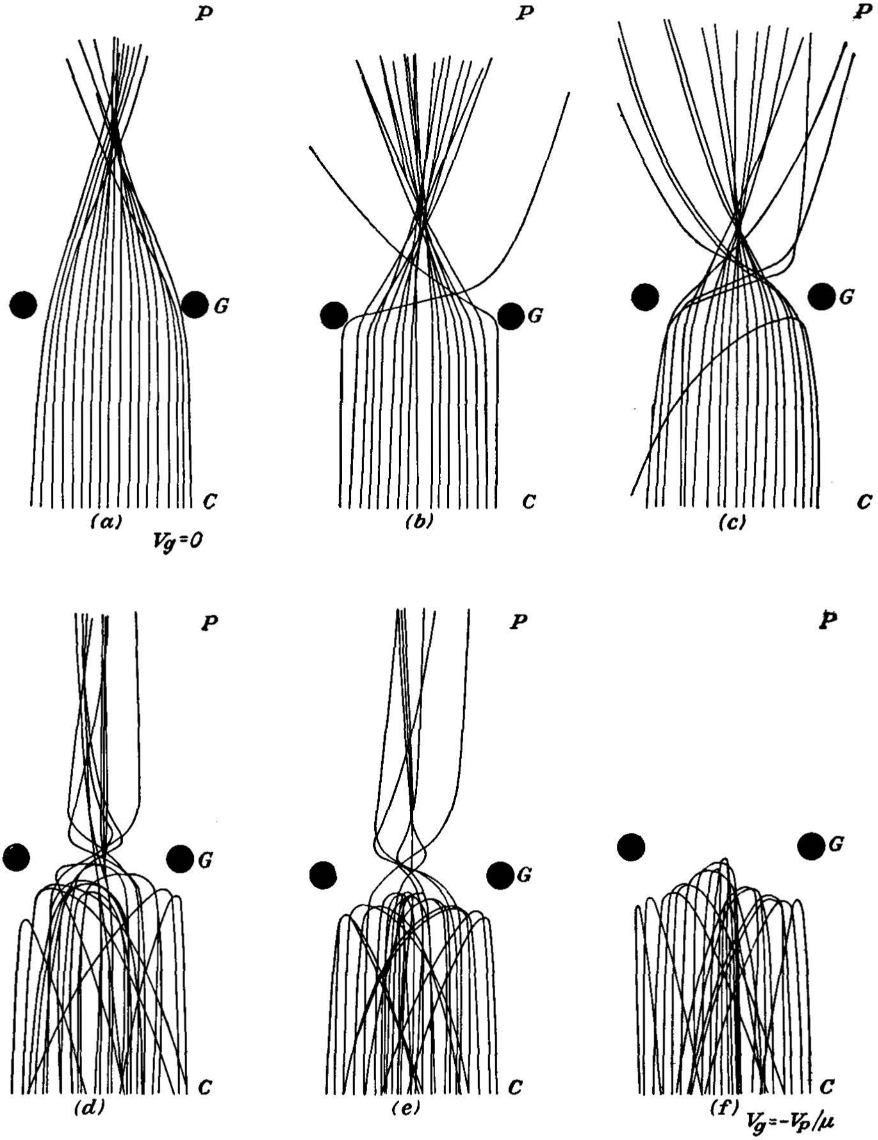


FIG. 9.10.—Electron paths in a negative-grid triode (Kleynen).

make no great difference in the electron paths. The space charge is most pronounced close to the cathode. Here the gradient of potential is normal to the cathode, and the electrons will move in straight lines away from it, it being assumed that it is plane. Deflecting components of field are not encountered by the electron until it approaches the grid plane. Here, however, the velocity of electrons passing midway between

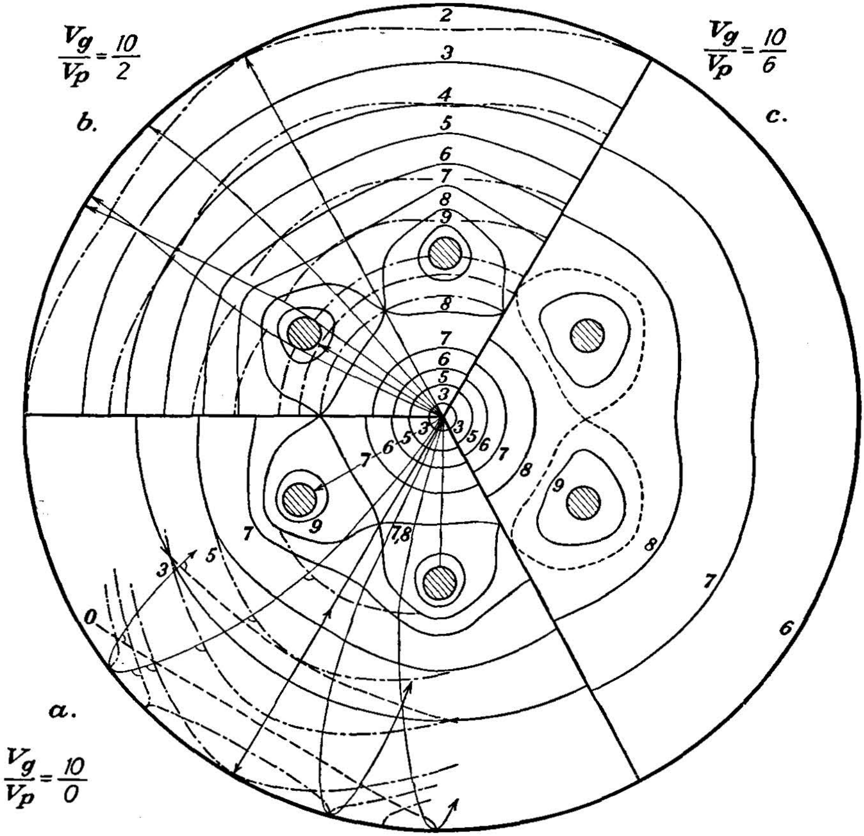
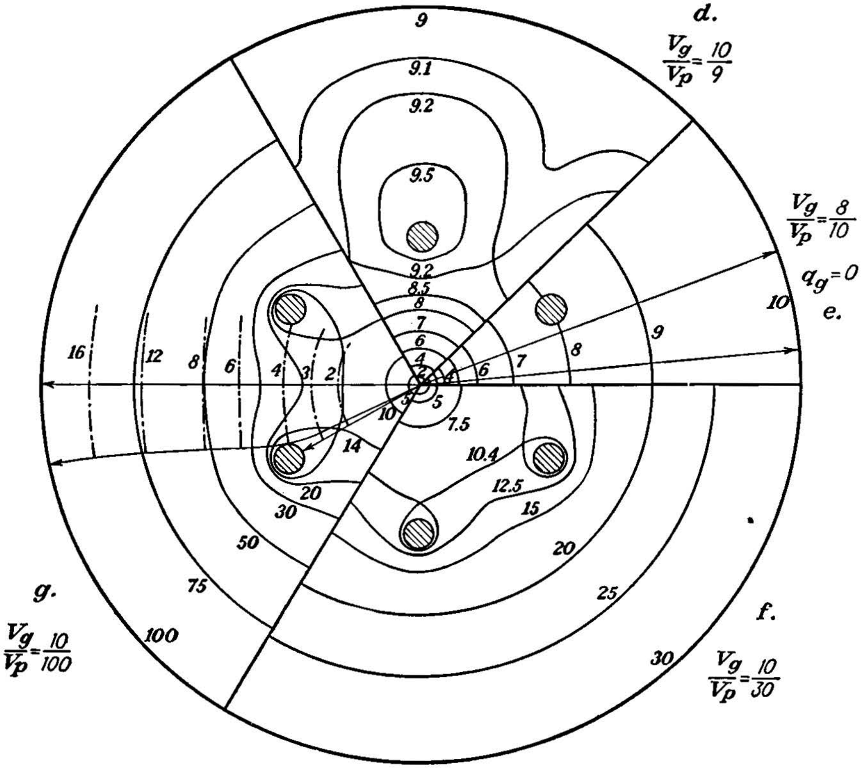


FIG. 9.11.—Electron paths in

the grid wires will be considerable and the space-charge effects will be less. In contrast, electrons approaching a grid wire directly and there turned back will be most affected by the space charge because the velocity will be low near the wires.

The successive parts of Fig. 9.10 show the effect of making the grid more and more negative until cutoff is reached. The sidewise deflecting forces become greater as the grid is made more negative until some of the electrons are turned back. Up to that condition the electrons are passed through the grid wires in a bunch that is focused beyond the grid plane. Coupled with the condition of more electrons being turned back as the grid voltage becomes more negative is the fact that fewer electrons get by the virtual cathode. Both factors contribute to the reduction in current, though the latter predominates greatly.

When the control grid is positive, it may attract electrons. The normal interception of current by the grid is roughly proportional to the projected area of the grid, though there is a strong dependence upon the relative voltages of the grid and plate. In Fig. 9.11 are shown some electron paths for a positive-grid triode.[1] These paths were calculated by the use of the action function, as described in the chapter on Laws of Electron Motion. The solid contours are equipotentials, the broken-dash contours are surfaces of constant action, and the electron paths are drawn



a positive-grid triode (Lange).

---

[1] LANGE, H., Current Division in Triodes and Its Significance in the Determination of Contact Potential, *Zeit. Hochfrequenz*, vol. 31, pp. 105–109, 133–140, 191–196, 1928.

perpendicular to these. When the grid is positive but less than its "natural potential," *i.e.*, that which would give the diode potential distribution in the tube, the action of the potential field causes these electrons, which initially miss the grids, to converge beyond, as was the case with the negative-grid triode. This situation is shown in part *g* of Fig. 9.11. When the grid is more positive than its natural potential and also more positive than the plate, as in *a* and *b* of Fig. 9.11, the potential field has a divergent action and the electrons are pulled into the grid wires. In *a* of Fig. 9.11 is shown the case of a positive grid with a plate at zero potential. Electrons that missed the grid initially will just barely graze the plate and then be pulled back toward the positive grids. Individual paths in this case will differ greatly, but in general the electrons will oscillate around the grid wires a few times before finally falling into them. This is the action encountered in a Barkhausen-Kurz oscillator.

9.6. Grid Current. Voltage amplifiers are operated with negative grid voltages, which means that grid current cannot flow. Power amplifiers of the Class B and C type are operated with the grid positive over an appreciable portion of the cycle during which grid current does flow. The grid current that does flow determines the power that is necessary to drive such amplifiers, and thus the matter of grid current is one of considerable importance.

Grid-current–Grid-voltage Characteristics. Qualitatively, the current to the grid of a triode is expected to increase as the grid voltage increases. This occurs because a more positive grid attracts electrons more strongly. Some typical grid-current–grid-voltage curves are shown in Fig. 9.12. These have the expected shape. The increase in grid current with grid voltage is more rapid than is the case for plate current. The curves for successively higher plate voltage fall below those for lower plate voltages. Thus in contrast to the grid-voltage variation, the grid current decreases with increasing plate voltage at a fixed grid voltage. This is logical, however; for as the plate is made more positive, the electrons are pulled past the grid more rapidly. They thus move in straighter lines, and therefore fewer of them are pulled into the grid. The current characteristics in the presence of secondary emission may be greatly different and will be treated separately later.

Grid-current–Plate-voltage Characteristics. Some typical grid-current–plate-voltage curves are shown in Fig. 9.13. For positive plate voltage the primary grid current decreases with increasing plate voltages as just noted. Curves for high positive grid voltage are shown above those for lower grid voltage. Grid current may flow when the plate voltage is negative, though such an operating condition is rarely encountered in practice. For this condition all the emitted current is taken by the grid. The grid current drops slightly as the negative plate voltage is made more negative. This is because the space current itself is reduced owing to the reduction in the equivalent voltage in the triode.

Constant-grid-current Contours. Contours of constant grid current are shown along with contours of constant plate current in Fig. 9.3.

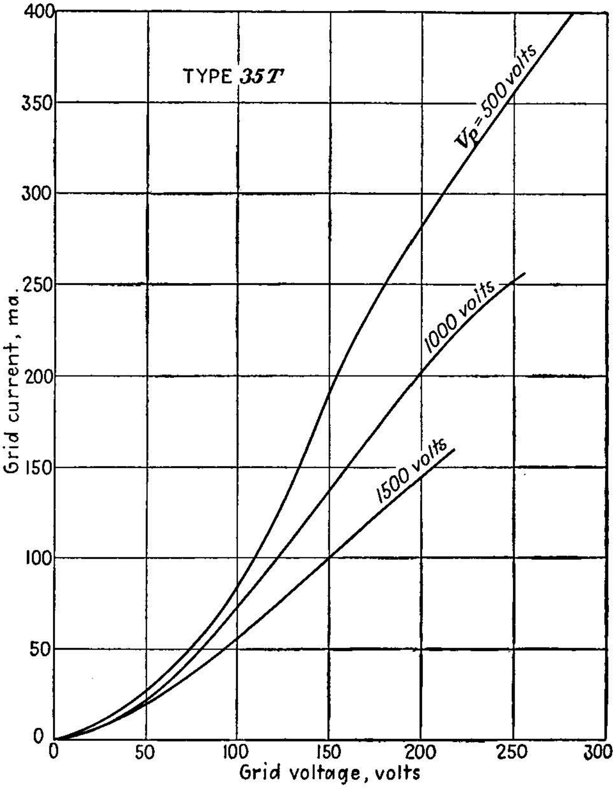


FIG. 9.12.—Grid-current–grid-voltage characteristics of a type 35T triode.

Grid current flows only when the grid is positive. The positive-grid–negative-plate quadrant is not shown because it is of little practical value. In the absence of secondary emission the contours present an orderly appearance. The contours follow no simple law as do the plate-current contours in the negative-grid region. The increase in grid current with grid voltage is much more rapid than the decrease with plate voltage.

The Grid-current Surface. Just as it was possible to draw a surface for the plate current as a function of grid and plate voltage, so is it possible to draw one for grid current. A sketch of such a surface is shown in Fig. 9.14. The previous representations of grid current will be recognized as part of this picture. The grid-current–grid-voltage curves are the intersections of the grid-current surface with a plane parallel to the grid-current and grid-voltage axes. The grid-current–plate-voltage curves are intersections of the surface with a plane parallel to the grid-current and plate-voltage axes. The constant-grid-current contours are intersections of the grid-current surface with planes parallel to the grid-voltage and plate-voltage axes.

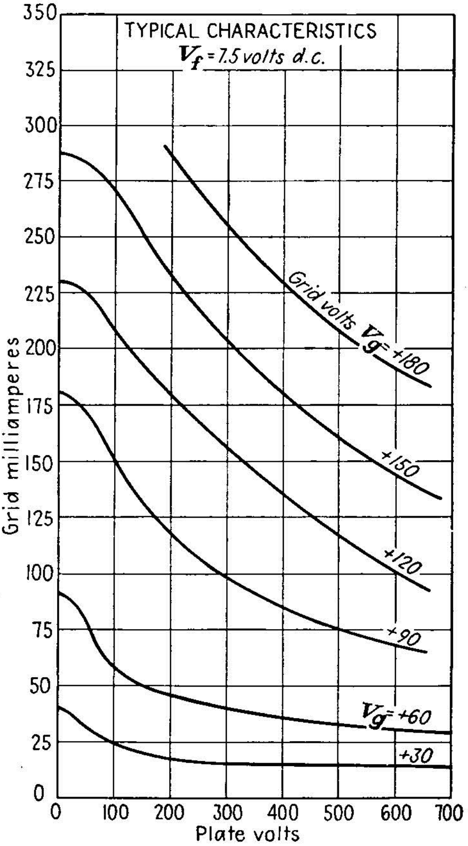


FIG. 9.13.—Grid-current–plate-voltage characteristics of a type 826 triode.

It is possible to define constants to describe the grid-current action, but this has no great value. It is of interest to note a few relations, however. The grid-current–plate-voltage transconductance is negative. The equivalent amplification factor for the inverted triode, *i.e.*, one whose grid is positive and whose plate is negative, is the reciprocal of the normal amplification factor of the tube.

Effect of Secondary Electrons. Secondary electrons are created whenever an electrode is struck with primary electrons that have been accelerated through more than a few volts. Triode characteristics are not affected much by secondary electrons as long as the grid is negative, for the secondary electrons that are formed at the plate are attracted back into the plate because there is no electrode more positive for them to go to. When the grid is positive, however, the secondary electrons formed by primaries striking the grid usually have a more positive plate to go to. As a result, the net grid current becomes the difference between the primary- and secondary-electron current. The magnitude of the secondary-electron current may be sufficient to distort the primary-grid-current curves almost beyond recognition.

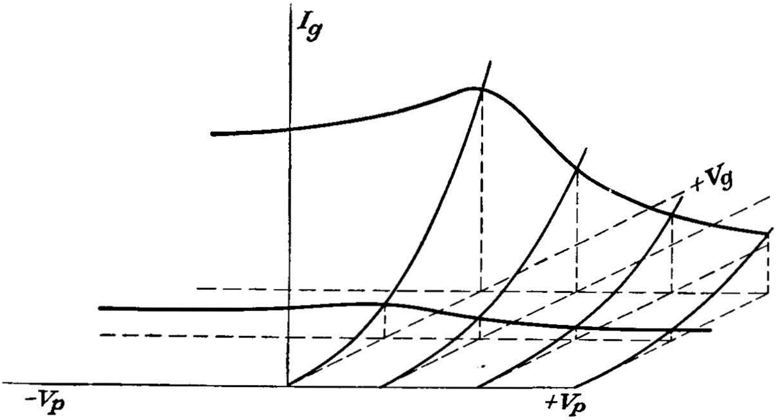


FIG. 9.14.—The grid-current surface.

When both grid and plate potentials are positive, secondary electrons are formed by primaries striking both. When the plate is more positive than the grid, the secondary electrons from the grid will be attracted to the plate but those formed at the plate will be attracted back into the plate. When the grid is more positive than the plate, the situation is reversed and secondaries from the plate will be attracted to the grid but those created at the grid will be attracted back into the grid itself. The result of this action upon the grid-current–grid-voltage characteristics is shown in Fig. 9.15. In this figure is shown a typical grid-current–grid-voltage curve in the presence of secondary emission compared with the primary-grid-current curve. For small grid voltages very few secondaries are created, and hence the currents with and without secondary emission are almost equal. As the grid voltage is increased, more secondaries are created and attracted to the plate. The grid current is therefore reduced by the amount of the secondary current to the plate. The grid current may be reduced enough to become negative. As the grid becomes more positive, more secondaries are likely to be created but the gradient of potential driving them to the plate becomes smaller, until finally it becomes negative when the grid potential exceeds the plate potential. As this occurs, the primary grid current exceeds the net grid current by less and less until when the grid potential and plate potential are equal the net grid current is nearly equal to the primary current. As the grid voltage is increased still further, the number of secondary electrons created at the grid surface becomes still greater but these electrons are confronted by a negative gradient of potential on all

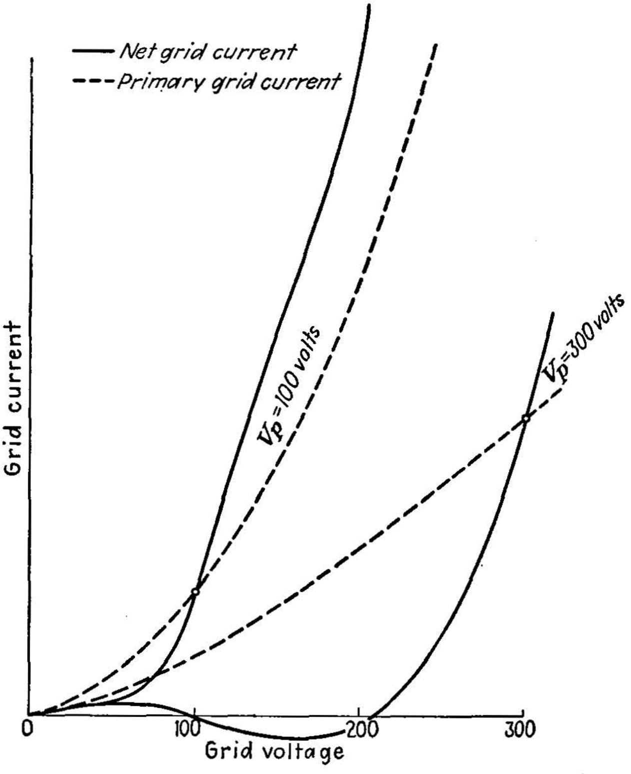


FIG. 9.15.—Influence of secondary emission upon the grid-current–grid-voltage characteristics of a triode.

sides and so are attracted back into the grid. Now, however, secondary electrons liberated from the plate are confronted by a positive gradient of potential that attracts them to the grid. The grid current is now greater than the primary grid current.

An action similar to that described above shows itself on the grid-current–plate-voltage curves. When the plate is less positive than the grid, secondary electrons from the plate are attracted to the grid and hence the actual grid current is greater than the primary grid current When the plate voltage is more positive than the grid voltage, secondary electrons from the grid are attracted to the plate and the grid current is less than the primary value. This action is shown in Fig. 9.16. Points of equal grid and plate voltage are crossover points of net and primary grid current. These points are marked by circles.

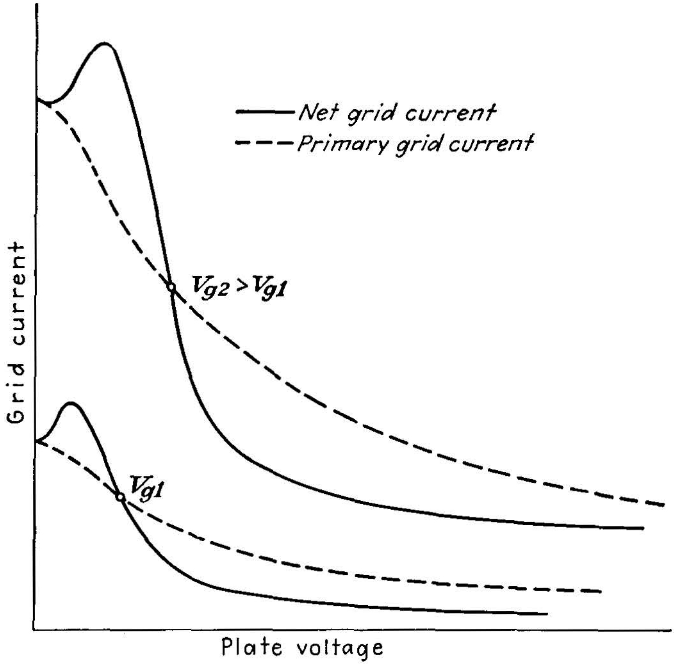


FIG. 9.16.—Influence of secondary emission upon the grid-current–plate-voltage characteristics of a triode.

The effect of secondary emission upon the contours of constant grid current may also be considerable, especially if the secondary emission is great enough to make the grid current negative. In Fig. 9.17 are shown some constant-grid-current contours of a water-cooled tube with a high degree of secondary emission. The effect of secondary emission is to raise all the large positive-current contours. It is as though a wedge of negative-current contours had been driven under the positive-current contours from the right. Contours of constant plate current are also distorted by secondary emission, though to a lesser degree.

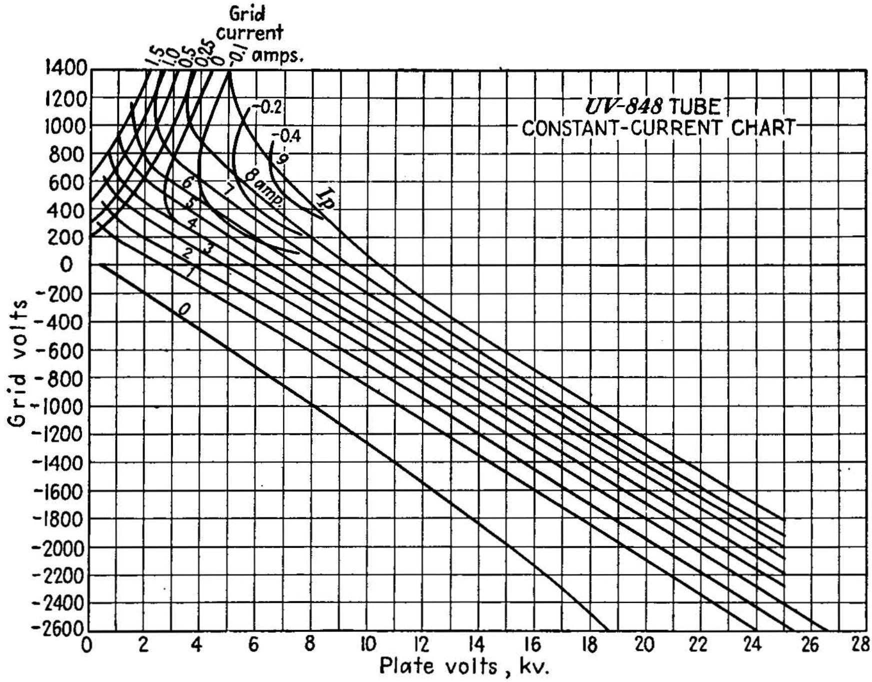


Fig. 9.17.—Effect of secondary emission upon the constant-grid-current contours of a triode.

9.7. Primary-grid-current Law. The complications introduced by secondary emission make it very difficult to treat grid current analytically. A considerable impression can, however, be made upon the subject of primary grid current. The analytical treatment of primary grid current is simplified by the observation made in Sec. 6.5 that, for a given *ratio* of plate to grid voltage, the electron paths within the tube are not altered by a change in the magnitude of these voltages. Since the electron paths are not changed, the division of current between the plate and grid is not changed and hence *the ratio of plate to grid current should be a function of the ratio of plate voltage to grid voltage alone* and be independent of the magnitude of these voltages. Were it not for secondary emission and some other effects such as the change in the position of the virtual cathode, this would be exactly true. Actually, the correspondence with expectations is quite good, as is shown in Fig. 9.18, in which there are plotted curves of the ratio of currents as a function of the ratio of voltages. The characteristics in this figure are for a small high-mu transmitting triode with tantalum electrodes. Such a tube is relatively free of secondary-emission effects. It is seen that the curves for different potentials superimpose reasonably well. If more curves were given, they would form a bundle within the limits of the curves shown.

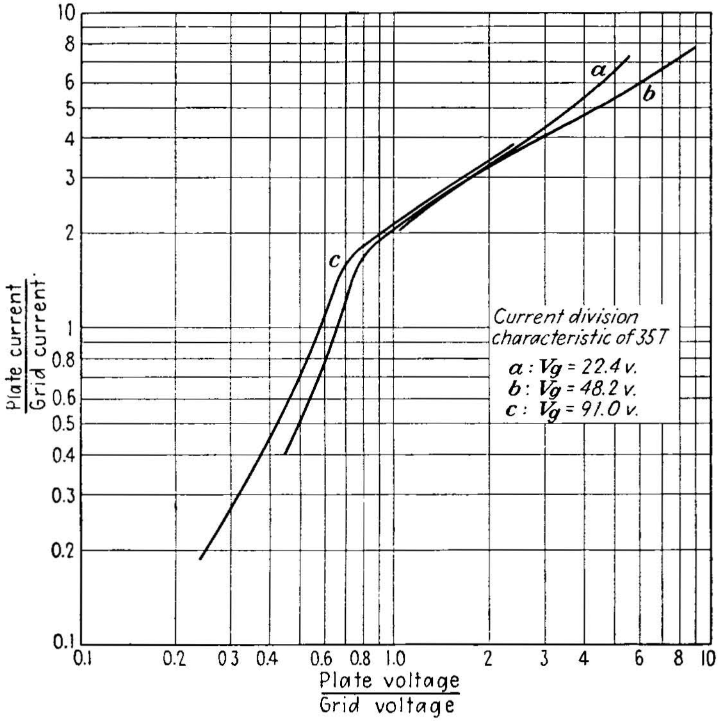


FIG. 9.18.—Electrode-current ratio in a positive-grid triode as a function of electrode-voltage ratio.

Examination of many tubes shows their primary-current-division characteristics to have the general form shown in Fig. 9.18. On such a log-log plot the curves are nearly straight lines with a slope of $\frac{1}{2}$ above a voltage ratio of 0.8 and with a slope of 2 below a voltage ratio of 0.8. Accordingly the primary current division may be expressed by

$$\frac{I_p}{I_g} = \delta \sqrt{\frac{V_p}{V_g}} \qquad \text{for } \frac{V_p}{V_g} > 0.8 \quad (9.12)$$

and

$$\frac{I_p}{I_g} = 1.392\delta \left(\frac{V_p}{V_g}\right)^2 \qquad \text{for } \frac{V_p}{V_g} < 0.8 \quad (9.13)$$

where δ is a constant known as the *current-division factor*[1-3] and defined as the ratio of plate to grid current for equal positive grid and plate voltages.

Since triodes are seldom operated with the plate less positive than the grid the form of Eq. (9.12) will be of more concern than that of Eq. (9.13). The reason for the change of slope, and hence of exponent, at the voltage ratio of 0.8 is that whereas all the electrons initially missing the grid go to the plate when the plate is more positive than the grid some of these will be returned to the grid when the latter is the more positive. This occurs because electrons that just barely miss the grid initially are strongly deflected and hence have not a sufficiently large component of velocity directed toward the plate to reach it, part of the electron energy now being in the form of a crosswise component of velocity. Thus, in addition to the grid intercepting a greater fraction of the primary space current directly as the grid voltage is made more positive relative to the plate, an increasingly greater fraction of the current that initially misses the grid returns to it.

Current-division Factor. A check upon the validity of the empirical Eq. (9.12) is given by an examination of the constancy of the coefficient of proportionality δ. This factor δ is logically called the "current-division factor" since it measures the ratio of plate to grid current for equal positive grid and plate voltage. It is a convenient reference point because it refers to a condition that is easy to measure. To measure the current-division factor it is necessary only to put current meters in the grid and plate leads of a triode and then connect the leads to a common voltage source and determine the ratio of currents. The current-division factor is also a good reference figure because it corresponds to the condition of peak current in typical Class C amplifier operation. If the ratio of plate to grid current in a triode is measured as a function of equal positive plate and grid voltages, variations of the sort shown in Fig. 9.19 result. For all the triodes shown, the current ratio rises sharply with voltage and then assumes a nearly constant value. The change in the current ratio with low voltages is caused primarily by the change in the position of the virtual cathode in front of the actual cathode. At low voltages the virtual cathode is located a considerable distance out from the actual cathode. As will be shown later, a small

---

[1] Tank, F., Zur Kentniss der Vorgänge in Elektrodenröhren, Jahr. draht. Tel. u. Tel. vol. 20, p. 80, 1922.

[2] See also LANGE, *op. cit.*

[3] EVERITT, W. L., and KARL R. SPANGENBERG, Grid-current Flow as a Factor in the Design of Vacuum-tube Power Amplifiers, *Proc. I.R.E.*, vol. 26, pp. 612–639, May, 1938.

cathode-grid distance leads to a small current-division factor. As the electrode voltages and correspondingly the current are increased, the virtual cathode moves back toward the actual cathode, causing the current ratio first to rise and then quickly to level off. The important observation about Fig. 9.19 is that the ratio of plate to grid current for equal grid and plate voltages is constant enough to make it eligible for a position as a fourth tube constant. The current-division factor in a tube free of secondary emission is as constant as the mu of the tube.

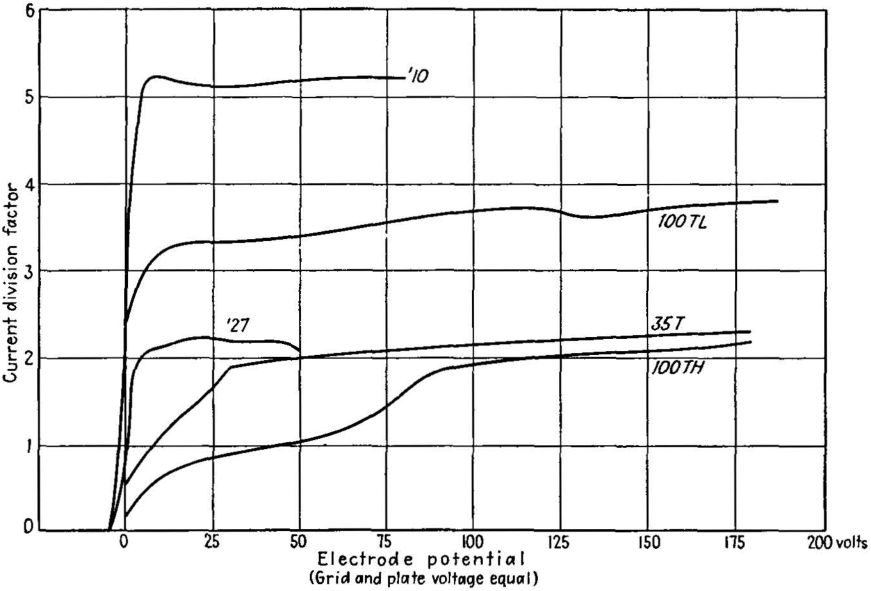


FIG. 9.19.—Plate-current–grid-current ratio as a function of equal plate and grid voltage.

Even when a triode has considerable secondary emission, the measured current ratio for equal positive grid and plate voltages is nearly equal to the primary current ratio because the interchange of secondary electrons between grid and plate is small when their voltages are equal.

Approximate Primary-grid-current Law. Since the total space current in a positive-grid triode is the sum of the grid and plate current,

$$I_s = I_p + I_g \tag{9.14}$$

this can be written

$$I_g = \frac{I_s}{1 + \dfrac{I_p}{I_g}} \tag{9.15}$$

Then, substituting the relation of Eq. (9.12),

$$I_g = \frac{I_s}{1 + \delta \sqrt{\dfrac{V_p}{V_g}}} \tag{9.16}$$

The space current itself is given by

$$I_s = G\left(V_g + \frac{V_p}{\mu}\right)^a \tag{9.17}$$

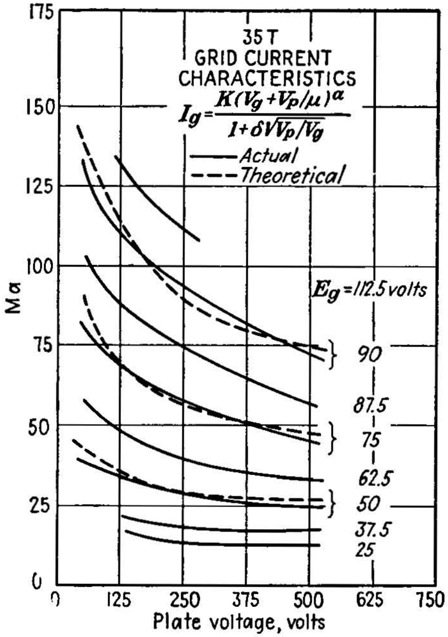


FIG. 9.20.—Comparison of actual and theoretical grid current.

where G is the perveance and a is a constant, approximately $\frac{3}{2}$. The resulting expression for primary grid current is given by

$$I_g = \frac{G\left(V_g + \dfrac{V_p}{\mu}\right)^a}{1 + \delta \sqrt{\dfrac{V_p}{V_g}}} \tag{9.18}$$

An idea of the accuracy of this approximation is given by Fig. 9.20, in which actual and theoretical grid-current curves are compared.

The expression for primary plate current for positive grid voltages corresponding to Eq. (9.18) is

$$I_p = \frac{G\left(V_g + \dfrac{V_p}{\mu}\right)^a}{1 + \dfrac{1}{\delta\sqrt{\dfrac{V_p}{V_g}}}} \tag{9.19}$$

Current-division-factor Formula. A formula for the current-division factor may be developed by solving for the point of origin on the cathode of a limiting electron that grazes the grid for a condition of equal grid and plate voltages.[1,2] The distance between the points of origin on the cathode of the two limiting electrons that strike a grid wire gives what may be called the "effective grid diameter." This is always larger than the actual grid diameter by a matter of 5 to 50 per cent, in typical cases. When the effective grid diameter or radius is known, the current-division factor δ is given by

$$\delta = \frac{a}{2r_{g\,\text{eff}}} - 1 \tag{9.20}$$

where a is grid-wire spacing and $r_{g\,\text{eff}}$ is effective grid radius.

The effective grid radius may be solved for in terms of the sidewise displacement of the electron grazing the edge of a grid wire. The component of gradient accelerating the electron toward the grid plane is virtually constant at the cathode value of

$$E_x = \frac{q_c}{\varepsilon_0} \tag{9.21}$$

The component of gradient giving the electron its sidewise deflection is

$$E_y = \frac{q_g \sin\left(\dfrac{2\pi y}{a}\right)}{2a\varepsilon_0\left[\cosh\left(\dfrac{2\pi x}{a}\right) - \cos\left(\dfrac{2\pi y}{a}\right)\right]} \tag{9.22}$$

where x and y are measured from a grid-wire center as in Fig. 7.17, q_c is given by Eq. (7.14*a*), and q_g is given by Eq. (7.14*b*). The sidewise deflection of the grazing electron is very nearly that which is obtained

---

[1] TELLEGEN, B. D. H., De Groote van der Roosterstroom in een Triode, *Physica*, vol. 6, pp. 113–116, March, 1926.

[2] SPANGENBERG, K. R., Current Division in Plane Electrode Triodes, *Proc. I.R.E.*, vol. 28, pp. 226–236, May, 1940.

by assuming that the sidewise force on the grazing electron is the same as that which exists along a line starting on the cathode at a point opposite the edge of a grid wire and passing tangent to the grid wire. This assumed force is correct at the point of contact on the grid, at which point the force is greatest.

Making the small-value approximation for x and taking y as r_g in Eq. (9.22) above,

$$E_y = \frac{q_g r_g}{2\pi\varepsilon_0(x^2 + r_g^2)} \tag{9.23}$$

(Note that this has the correct value when $x = 0$.)

Upon substituting the approximate values of E_x and E_y from the above into the acceleration equations (6.41) and (6.42), eliminating time, and equating grid and plate voltage, the expression for the sidewise displacement of the grazing electron is found to be

$$y_0 = \frac{a\mu}{2\pi(\mu + 1)} \frac{r_g}{2d_{cg}} \ln\left(\frac{4\epsilon d_{cg}}{r_g}\right) \tag{9.24}$$

The effective grid radius is equal to $y_0 + r_g$. When the expression for the effective grid radius is applied to Eq. (9.20), it is found that the current-division factor is

$$\delta = \frac{a}{\dfrac{a\mu}{\pi(\mu + 1)} \dfrac{r_g}{2d_{cg}} \ln \dfrac{4\epsilon d_{cg}}{r_g} + 2r_g} - 1 \tag{9.25}$$

in which a = distance between grid wires
μ = amplification factor
r_g = grid-wire radius
d_{cg} = cathode-grid distance
ϵ = Napierian base, 2.718

The magnitude of the current-division factor is given by the nomographs of Fig. 9.21 and 9.22. In Fig. 9.21 is a nomographic chart from which the effective grid radius is given in terms of the grid-wire spacing, the cathode-grid distance, and the amplification factor. This chart is read by means of two perpendicular lines ruled upon a transparent sheet. The construction cross shown on the chart gives the effective radius of a type 210 tube. The nomograph of Fig. 9.22 is a graphical representation of Eq. (9.20) and gives the current-division factor from the effective grid-wire radius and grid-wire spacing. Examination of Figs. 9.21 and 9.22 shows that the current-division factor increases with both grid-wire spacing and grid-cathode spacing. The current-division factor also increases with amplification factor, but only slightly. Typical values of the effective grid-wire radius will be 105 to 150 per cent of the actual grid-wire radius.

Current-division Law in the Presence of Secondary Emission. When the analysis that led to the current-division factor is generalized by allowing the grid and plate voltage to assume general values, the sidewise deflection of the electron grazing the grid is found to be

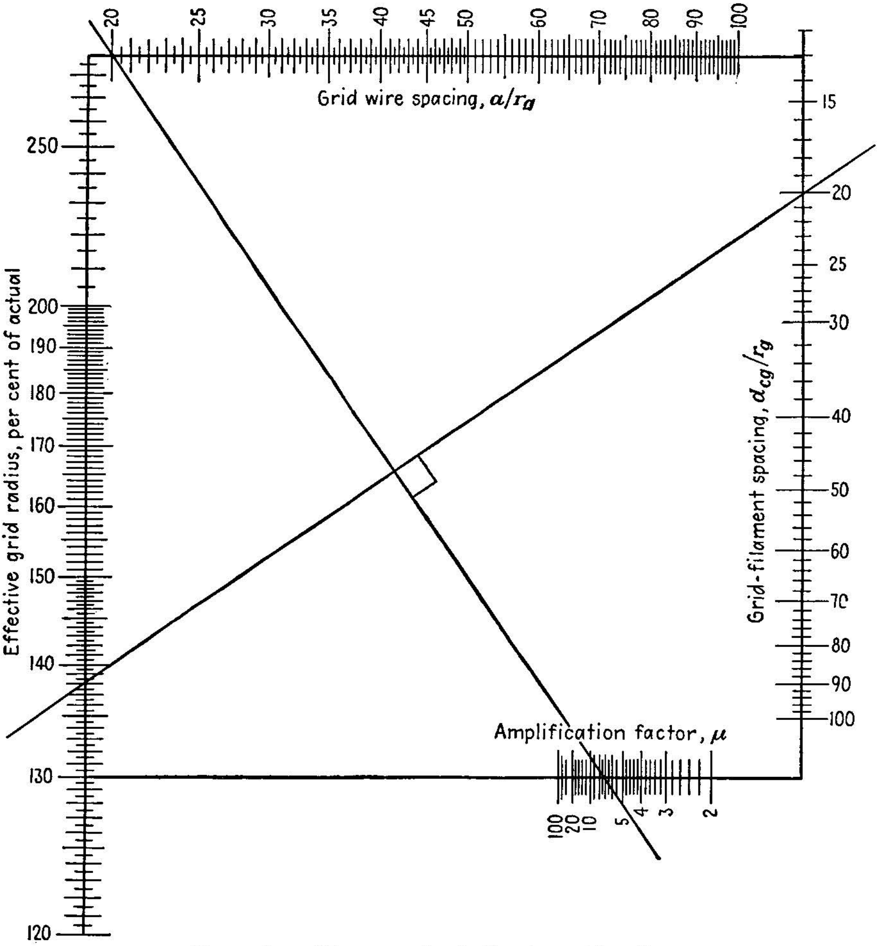


FIG. 9.21.—Nomograph of effective grid radius.

$$y_1 = \frac{a\mu[(d_{cg} + d_{gp})V_g - d_{cg}V_p]}{2\pi d_{gp}(V_p + V_g)} \frac{r_g}{2d_{cg}} \ln\left(\frac{4\epsilon d_{cg}}{r_g}\right) \qquad (9.26)$$

Equation (9.26) has been arrived at by solving for the sidewise displacement of the electrons grazing the grid as a function of electrode-voltage ratio. The corresponding current ratio is then readily determined. The electrode-current ratio is

$$\frac{I_p}{I_g} = \frac{a}{2(y_1 + r_g)} - 1 \tag{9.27}$$

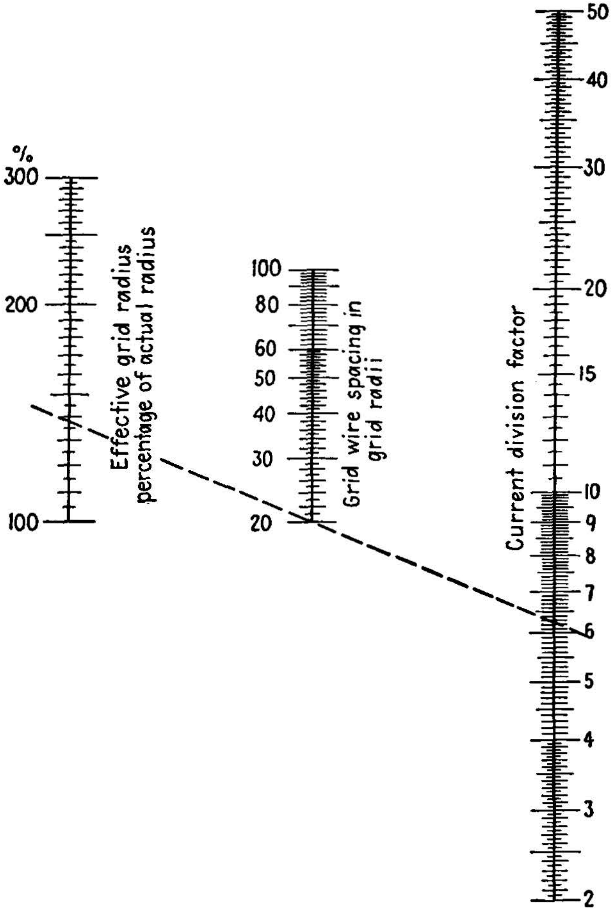


FIG. 9.22.—Nomograph of current-division factor.

Upon substituting the value of y_1 from Eq. (9.26), the current ratio is

$$\frac{I_p}{I_g} = \frac{a}{2\left\{\frac{a\mu\left[(d_{cg} + d_{gp}) - d_{cg}\frac{V_p}{V_g}\right]D}{2\pi d_{gp}\left(\frac{V_p}{V_g} + \mu\right)} + r_g\right\}} - 1^* \tag{9.28a}$$

---

* A somewhat more accurate formula has since been developed by J. H. L. Jonker and B. D. H. Tellegen, Current to a Positive Grid in Electron Tubes, *Philips Research Reprints*, vol. 1, pp. 13–32, October, 1945.

or

$$\frac{I_p}{I_g} = \frac{LV_g + MV_p}{PV_g + QV_p} \tag{9.28b}$$

where
$$\begin{aligned} L &= \pi a d_{gp}\mu - a\mu(d_{gp} + d_{cg})D - 2\pi d_{gp} r_g \mu \\ M &= \pi a d_{gp} - 2\pi d_{gp} r_g + a d_{cg} D\mu \\ P &= a(d_{gp} + d_{cg})\mu + 2\pi d_{gp} r_g \mu \\ Q &= a d_{cg} D\mu - 2\pi d_{gp} r_g \\ D &= \frac{r_g}{2d_{cg}} \ln \frac{4\epsilon d_{cg}}{r_g} \end{aligned}$$

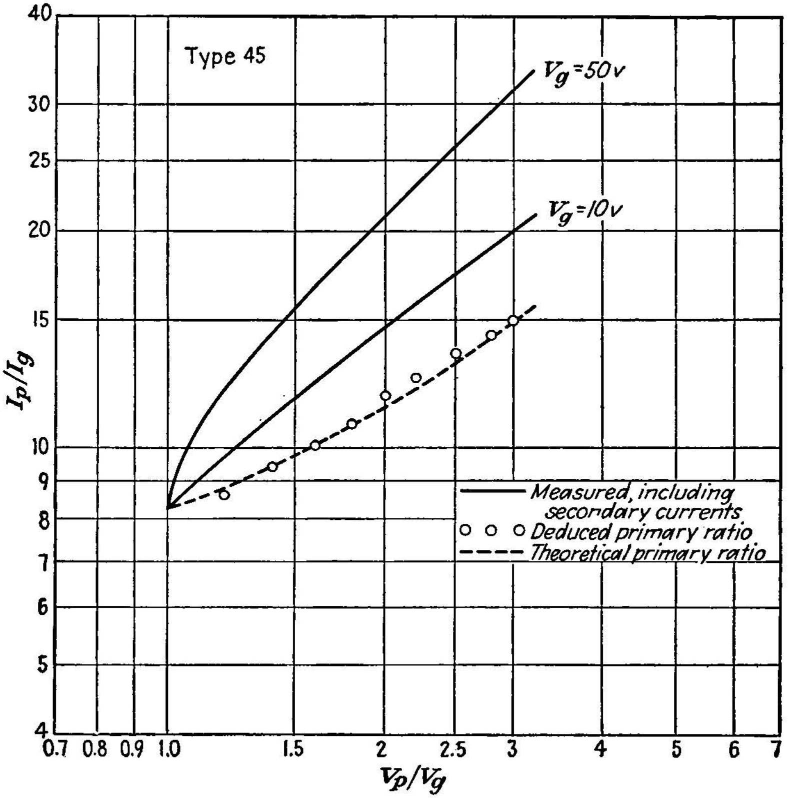


FIG. 9.23.—Primary-current division in a 45 triode.

A plot of Eq. (9.28*a*) in a typical case shows the curve to be concave upward as in the dotted curve of Fig. 9.23. In such a plot, the slope of the true current ratio is between ¼ and ¾ so that the assumption of a one-half-power law when this curve is slightly modified by space-charge and secondary-emission effects is a reasonable one.

To check the correctness of the above equation in actual tubes it is necessary to correct measured curves for the effect of secondary emission, which is always present to a degree. This is done by an extension of methods developed for screen-grid tubes.[1]

The curves from which the deduction of the true primary distribution are made are taken as follows: Filament emission is first reduced to the point where the current is temperature-limited rather than space-charge-limited. The grid voltage is then set at some value, and the ratio of plate to grid current is observed as a function of the ratio of plate to grid voltage by varying the plate voltage only. The grid voltage is then set at another positive value, and another similar run is made. The two solid curves of Fig. 9.23 were made by this method.

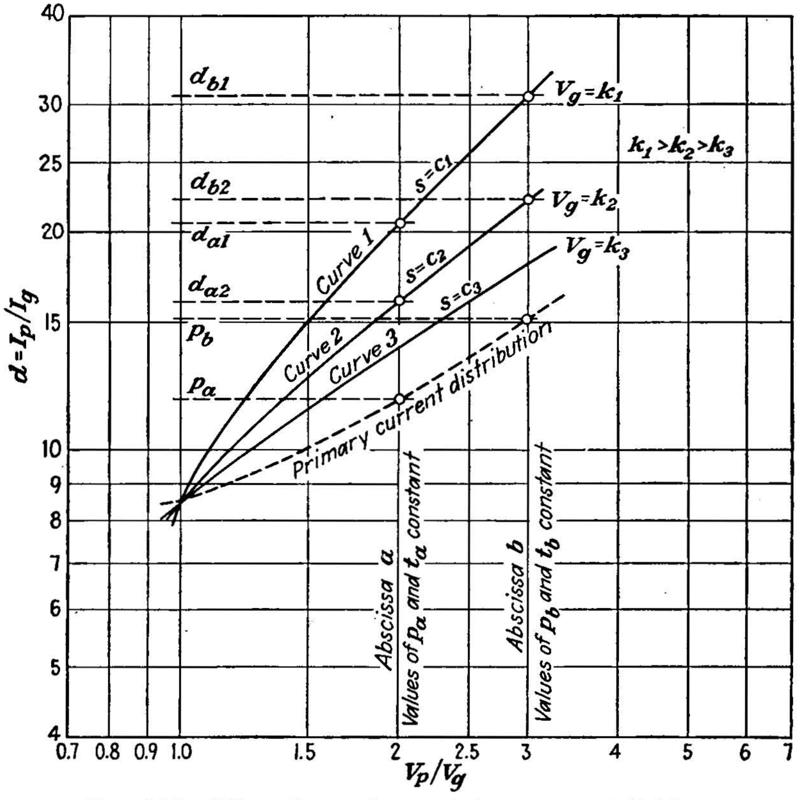


FIG. 9.24.—Effect of secondary emission on current division.

Because of the various factors that have been held constant and

---

[1] DE LA SABLONIERE, C. J. L., Die Sekundäremission in Schrimgitterrohren, *Hochfreq. u. Audio.*, vol. 41, pp. 195–202, June, 1933.

relations between the various current components a number of relations exist that must be borne in mind. Before summarizing these relations the notation to be used must be indicated in detail. Let I_p and I_g be total plate and grid current, respectively, including secondaries. Let I_{p1} and I_{g1} be those parts of the plate and grid currents which are due to primary electrons, *i.e.*, the primary plate and grid currents. Let I_{p2} and I_{g2} be the currents corresponding to all the secondary electrons that are knocked out of the plate and grid, respectively. This includes not only those secondary electrons which succeed in getting from one electrode to another but also those which are knocked out of one electrode and fall back into that same electrode. Let I_{gp} be that fraction of I_{g2} which does succeed in getting from grid to plate. Similarly, let I_{pg} be that fraction of I_{p2} which is able to get from plate to grid. Obviously, if the plate is much more positive than the grid, I_{gp} will be a large fraction of I_{g2}, while I_{pg} will not exist as a component of I_{p2} because all the secondary electrons knocked from the plate will be drawn back into the strongly positive plate.

Let $s = \frac{I_{g2}}{I_{g1}}$. The quantity s is a secondary-emission factor measuring the ratio of the number of secondary to primary electrons. Physical studies have shown that s depends only upon the velocity of the striking primary electrons for any given surface. Hence, along any curve such as those in Fig. 9.24, s will be constant since each curve is taken with a constant value of grid voltage.

Let $p = \frac{I_{p1}}{I_{g1}}$. This gives the division of primary current that from theoretical considerations is a function of the ratio of plate to grid voltage alone. Hence, for any particular value of $\frac{V_p}{V_g}$, p is a constant.

Let $d = \frac{I_p}{I_g}$. This is the ratio of plate to grid current, including the secondary-emission effects. The curves of Fig. 9.24 are curves of d against $\frac{V_p}{V_g}$.

Let $t = \frac{I_{gp}}{I_{g2}}$. This is a kind of transmission factor for secondary electrons. It measures the fraction of secondaries liberated that succeeds in getting to the plate. Some secondary electrons from the grid have such a low velocity that they are unable to climb the small potential hill between the grid and the plate. De la Sabloniere has assumed that for any value of the abscissa $\frac{V_p}{V_g}$ the value of t is constant. That is, for any value of $\frac{V_p}{V_g}$ the same fraction of the secondary electrons knocked from the grid succeeds in getting to the plate. This is perhaps the only assumption which is questionable. The matter is complicated by the velocity distribution of the secondary electrons, which changes as the striking voltage of the primary electrons changes. For the assumption to be strictly true the velocity-distribution curve of the secondary electrons must expand uniformly as the striking potential of the primary electrons increases. This is not strictly true but for small ranges of primary-electron velocity is approximately so. In the curves of Fig. 9.24 the primary-electron velocities are 10 and 50 volts. It was not found possible to get a good check for velocities of 10 and 200 volts, this being too great a range of primary velocities.

It will be noted further that the space current for each of the experimentally determined curves is approximately constant.

Consider the ratio

$$\frac{I_p}{I_g} = \frac{I_{p1} + I_{gp}}{I_{g1} - I_{gp}} \tag{9.29}$$

Dividing both numerator and denominator by I_{g1} there results

$$\frac{I_p}{I_g} = \frac{\frac{I_{p1}}{I_{g1}} + \frac{I_{gp}}{I_{g1}}}{1 - \frac{I_{gp}}{I_{g1}}} \tag{9.30}$$

But

$$\frac{I_{gp}}{I_{g1}} = \frac{I_{gp}}{I_{g2}} \frac{I_{g2}}{I_{g1}} = ts \tag{9.31}$$

so that the above ratio of net currents can be written as

$$d = \frac{p + ts}{1 - ts} \tag{9.32}$$

Solving this for ts,

$$ts = \frac{d - p}{d + 1} \tag{9.33}$$

Let the various curves of d against $\frac{V_p}{V_g}$ be numbered 1, 2, and so on, as shown in Fig. 9.24. Let the various values of $\frac{V_p}{V_g}$ have letters corresponding to them. Thus the abscissa of $\frac{V_p}{V_g} = 2$ might be lettered a, that of $\frac{V_p}{V_g} = 3$ might be lettered b, and so on. If we consider the four points formed by the intersection of the upper two curves of Fig. 9.24 and any two abscissas denoted by a and b, then it is possible to write four equations of the form of that last given. These will be

$$t_a s_1 = \frac{d_{a1} - p_a}{d_{a1} + 1} \tag{9.34}$$

and

$$t_a s_2 = \frac{d_{a2} - p_a}{d_{a2} + 1} \tag{9.35}$$

for the intersections of the curves 1 and 2 with the abscissa a and

$$t_b s_1 = \frac{d_{b1} - p_b}{d_{b1} + 1} \tag{9.36}$$

$$t_b s_2 = \frac{d_{b2} - p_b}{d_{b2} + 1} \tag{9.37}$$

for the other intersections. Since t and p are presumed constant for any particular value of $\frac{V_p}{V_g}$, they are given only a lettered subscript.

Dividing the two pairs of equations and equating them gives

$$\frac{d_{a1} - p_a}{d_{a2} - p_a} = \frac{d_{b1} - p_b}{d_{b2} - p_b} \tag{9.38}$$

which is the relation that has been sought. This may be solved for p_b to give

$$p_b = \frac{d_{b1} - d_{b2}\left(\dfrac{d_{a1} - p_a}{d_{a2} - p_a}\right)}{1 - \left(\dfrac{d_{a1} - p_a}{d_{a2} - p_a}\right)} \tag{9.39}$$

From this last equation it may be seen that if one point, p_a, on the true primary distribution curve is known, then points at any other abscissa b may be found from a pair of curves giving the net current division in the presence of secondary emission. The above treatment has been given for the case of V_p greater than V_g, but a similar treatment can be applied when this is not so. In this particular instance the primary-current distribution that was taken as known was that corresponding to the condition of the grid being at its "natural potential" relative to the plate. For this case, the electrons move in substantially parallel straight lines from filament to grid and plate, and the ratio of plate to grid current is determined by the ratio of intergrid to grid area. For the 45 tube this ratio of currents is 14.3 when the ratio of voltages is 2.81.

CHAPTER 10

TETRODES

10.1. Types of Tetrode. A tetrode, as its name implies, is a four-electrode tube. The four electrodes are invariably, in the order of their arrangement, the cathode, the control grid, the screen grid, and the plate. There are two types of tetrode. These are the so-called "screen-grid tube" and the "beam-power tube."

The screen-grid tube was the successor to the triode and the predecessor of the pentode, though, as indicated in the chapter on Basic Tube Types, it is now virtually obsolete and seldom used because of unfavorable current-voltage characteristics. The ordinary screen-grid tube has a fine control grid surrounding the emitter, which in turn is surrounded by a coarser screen grid a considerably greater distance out. The screen grid is in turn surrounded by a plate. The intended function of the screen grid was to shield the control grid electrostatically from the plate and so reduce the tendency toward oscillation that existed in r-f amplifiers. The screen grid performed this function, but it also introduced some other characteristics that were not desirable. Specifically, it introduced secondary emission, which distorts the current-voltage characteristics.

The beam-power tube is a special tetrode with aligned control and screen grids. It was the historical successor to the pentode. The pentode was developed to eliminate the secondary-emission action that appears in the screen-grid tube. The beam-power tube was later found capable of doing the same thing without an extra grid if proper attention were paid to grid alignment and to dimensioning.

10.2. Current-voltage Characteristics of the Screen-grid Tube. The screen-grid tube is usually operated with its screen at a fixed direct potential and by-passed with a large condenser to ground so that no alternating components of potential appear on it. The screen grid acts as a shield between the plate and control grid. Electrostatic lines from the plate terminate for the most part on the screen grid. This electrostatic behavior does not interfere with electronic action. An electron stream of varying intensity can still pass between the screen-grid wires.

The current-voltage characteristics of the screen-grid tube are determined by two principal effects that are at work. (1) The relative screen-grid and plate potentials determine how the space current will divide between these two electrodes. (2) The relative positiveness of plate and screen grid determines how secondary electrons will be interchanged between these two electrodes. In general, the behavior with regard to both factors is similar to that which exists in the positive-grid triode.

As far as space-current effects are concerned, the control grid and screen grid have the principal influence. With respect to first-order effects, it may be said that the screen grid plays the same role in the screen-grid tube as the plate does in the triode. The plate has only a very small influence in modifying space current in the screen-grid tube because of the shielding effect of the screen grid. The space current is given by

$$I_s = G\left(V_1 + \frac{V_2}{\mu_{sg}} + \frac{V_p}{\mu_p}\right)^a \tag{10.1}$$

where V_1 is control-grid potential
V_2 is screen-grid potential
G is perveance
V_p is plate potential
μ_{sg} is equivalent amplification factor of the screen grid, $-\left(\frac{dV_2}{dV_1}\right)_{I_{p\,\text{const}}}$
μ_p is plate amplification factor, $-\left(\frac{dV_p}{dV_1}\right)_{I_{p\,\text{const}}}$
a is a constant, nearly $3/2$

In this expression, μ_{sg} is considerably smaller than μ_p. The equivalent screen-grid amplification factor may be calculated quite accurately from the triode mu formulas by treating the screen grid as though it were the plate. The accuracy of this approximation decreases as the shielding effect of the screen grid decreases. The plate amplification factor may be calculated from some special formulas, which will be developed subsequently. It may be determined approximately by calculating a triode amplification factor, considering the control grid as the cathode, the screen grid as the control grid, and the plate as the plate, and then multiplying this amplification factor by the screen-grid amplification factor. This relation holds because the fictitious amplification factor cited first above measures the screening effect of the screen grid upon the control-grid plane just as the screen-grid amplification factor measures the screening effect of the control grid upon the cathode. The product of these two amplification factors, which are reciprocal screening factors, gives the over-all amplification factor. Thus, if the screen-grid amplification factor were 20 and the triode amplification factor obtained by considering the control grid as the cathode were 10, the plate amplification factor would be approximately 200. If the cold cathode were at zero potential and all the other electrodes were at the same positive potential, then one-twentieth of the electrostatic flux lines from the cathode would penetrate the control grid into the space beyond (actually, the ratio would be 1 in 21). Of the lines that passed through the control grid, one-tenth would pass on to the plate, and the rest would terminate on the screen grid. The over-all screening effect would be such that only 1 line would reach the plate for every 200 that reached the control grid. The resulting plate amplification factor is 200.

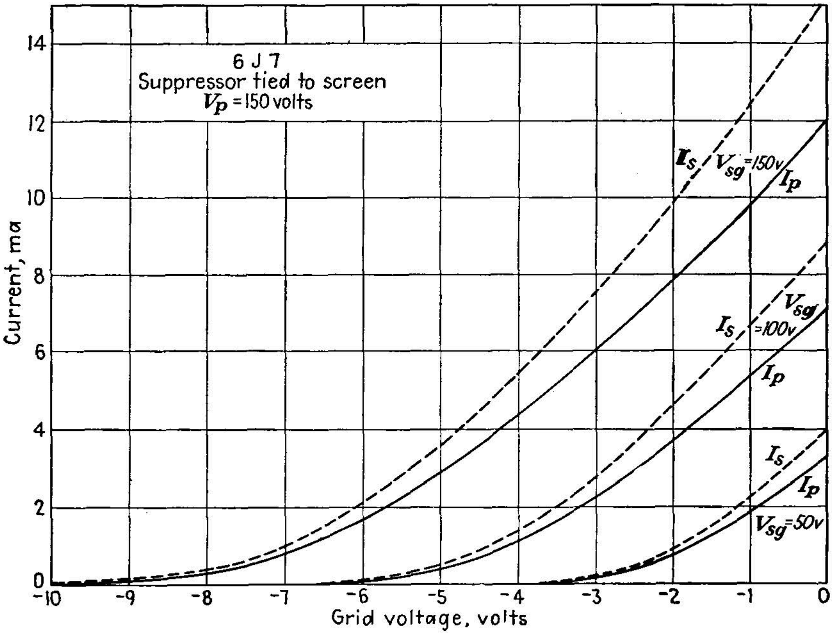


FIG. 10.1.—Plate and space current of a screen-grid tube as a function of control-grid voltage.

The plate-current–control-grid characteristics of the screen-grid tube are almost the same as the triode characteristics that result if the screen grid and plate are connected together. The only difference is that a small part of the space current is taken by the screen grid. Some typical plate-current and space-current characteristics as a function of control-grid voltage are shown in Fig. 10.1. Because of the usually high value of the plate amplification factor the plate potential has only a small effect upon the plate and space current compared with the screen-grid potential. This in turn has much less influence than the control-grid potential.

Plate-current–Plate-voltage Characteristics of the Screen-grid Tube. With a negative control-grid voltage and a positive screen-grid voltage, the plate-current–plate-voltage characteristics of a screen-grid tube have the form shown in Fig. 10.2. The shape of the plate-current curve departs considerably from the shape of the primary plate-current curve because of secondary emission. The probable shape of the primary plate-current curve has been sketched for $V_g = 0$. The primary plate current is not readily measured directly. It is seen to be an increasing fraction of the approximately constant space current I_s. The division of space current between screen grid and plate follows approximately the same law as does the division of the space current between grid and plate in a triode. When the plate voltage is zero, the plate gets none of the space current. As the plate is made positive, it rapidly acquires a major portion of the space current. When the plate is as positive as the screen grid, it gets a slightly smaller fraction of the total space current than the ratio of the area between the screen-grid wires to the total area of the screen-grid plane. As the plate potential is made still more positive, the plate acquires still more of the space current until at very large voltages the plate is getting nearly all the space current.

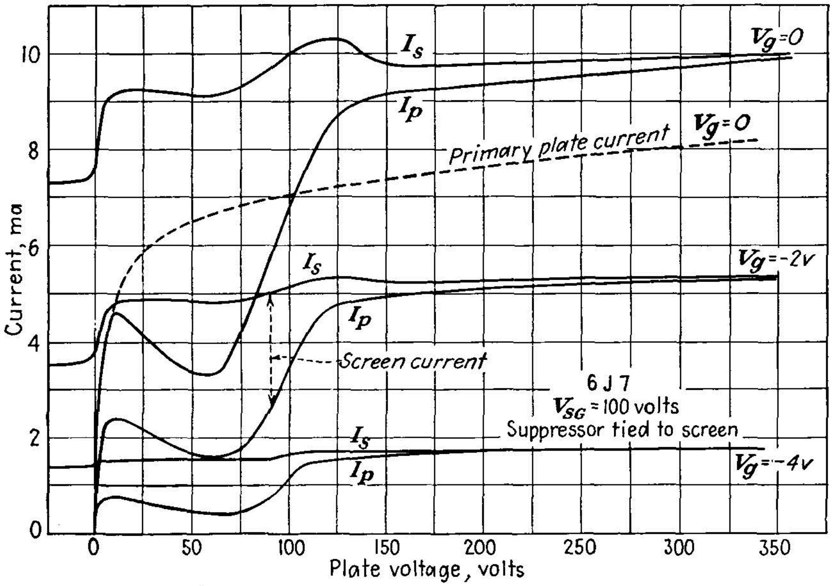


FIG. 10.2.—Plate-current–plate-voltage characteristics of a screen-grid tube.

The difference between the primary plate-current curves and the actual plate-current curves is obviously due to secondary-emission effects. The effects are the same as in the positive-grid triode. When the plate is less positive than the screen grid, secondary electrons liberated at the plate surface are attracted to the screen grid, thus reducing the plate current. This accounts for the pronounced dip in the plate-current characteristic. When the plate potential is equal to the screen-grid potential, the interchange of secondary electrons between plate and screen grid nearly balances and the actual plate current is nearly equal to the primary plate current. As the plate becomes more positive, it collects secondary electrons that are liberated from the screen grid, and, as a result, the actual plate current exceeds the primary plate current. The variation of plate current with control-grid voltage follows the high-mu-triode law.

Also shown in Fig. 10.2 are curves of space current as a function of plate voltage. If the plate amplification factor of the tube were extremely high, the space current would be completely independent of plate voltage. As it is, the space current tends to be fairly constant. Departures from constancy are observed, however, at zero plate potential and at the plate potential equal to the screen potential. The changes in the space current observed in these places are due to changes in the space-charge condition around the screen-grid wires. When the plate potential is negative, the electrons that initially miss the screen grid are reflected back from the plate and in general will oscillate around the wires a few times before being drawn in. The presence of these oscillating electrons constitutes an addition to the space charge and depresses the potential before the screen grid and even reaches back through the control grid to reduce the emitted current. When the plate potential becomes slightly positive, part of the electrons that initially miss the screen grid are received by the plate. This means that the current reflected back toward the screen grid is suddenly reduced, the space charge around the screen grid is correspondingly reduced, as is also its depressing effect upon the potential before the cathode, and as a result the emitted current suddenly increases. The nature of the change in the potential distribution within the screen-grid tube as the plate potential is changed from negative to positive is sketched in Fig. 10.3. The dotted lines in this figure show potential profiles for a negative plate potential, while the solid lines show potential profiles for a positive plate potential. The manner in which the plate potential controls the off-cathode gradient through the medium of the oscillating space charge about the screen grid may also be seen.

When the plate potential becomes more positive than the screen-grid potential, there is a change from a condition of partial reflection of electrons from the plate to one of no reflection, for all electrons reach the plate, no matter how strongly deflected by the screen grid. Here the space charge around the screen grid is again suddenly reduced, and the space current increases. Thus the space current is influenced most by the condition of current transmission to the plate and is hardly affected by secondary emission.

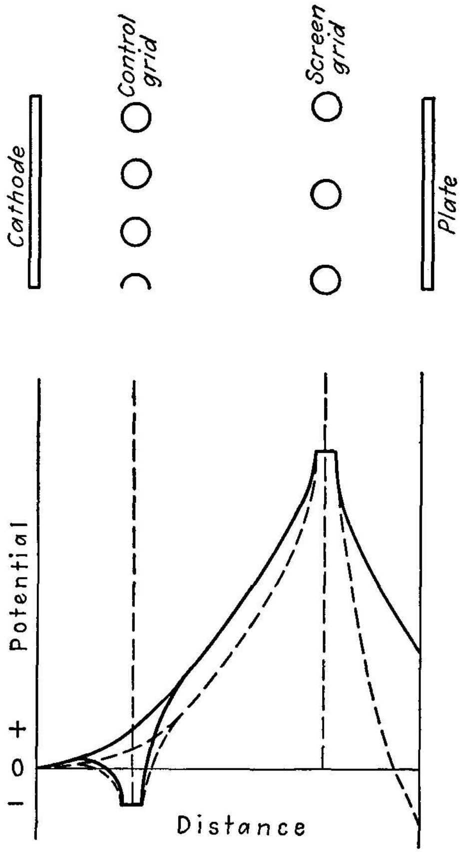


FIG. 10.3.—Potential distributions within a screen-grid tube for negative and positive plate potentials.

Screen-current–Plate-voltage Characteristics of the Screen-grid Tube. The screen-grid current is the difference between the space current and the plate current in Fig. 10.2. This difference is plotted as screen current in Fig. 10.4 as a function of plate voltage. The screen-current–plate-voltage curves are like the positive-grid-current–plate-voltage curves of a triode. Exactly the same factors enter into its composition. The primary distribution is such that the screen current decreases uniformly with increasing plate voltage exactly as is the case for the triode grid. When the plate voltage is less positive than the screen-grid voltage, the screen grid acquires current from the plate and hence rises above the primary-current value. If the secondary emission is sufficient, the net screen-grid current will rise with voltage until it falls as the plate potential becomes more positive than the screen potential. When this happens, the screen loses secondary electrons to the plate and as a result the net screen-grid current drops below the primary value and may even go negative in some cases.

General Characteristics of Screen-grid Tubes. Because of the distortions in the plate-current curves caused by secondary emission, the screen-grid tube has rather restricted ranges of potentials in which it operates satisfactorily. For very high plate potentials relative to the screen-grid potential, the current characteristics are very uniform. The range of uniform current characteristics is necessarily quite limited. The screen-grid potential must be relatively high to draw sufficient current. The plate potential must be at least this positive to avoid secondary-emission distortions and yet cannot be too much more positive because then the plate dissipation becomes excessive. In this operating region the plate resistance of the tube is very high. The amplification factor is also high, but the mutual conductance is of the same order as in a triode.

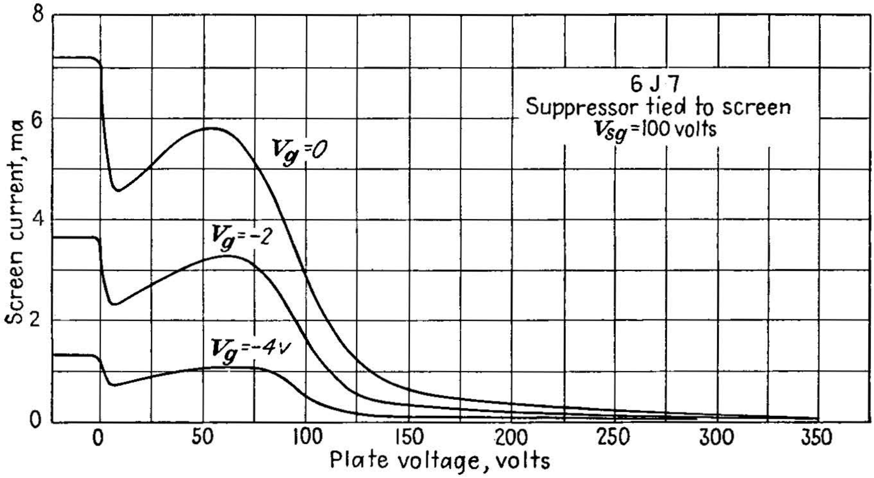


FIG. 10.4.—Screen-current–plate-voltage characteristics of a screen-grid tube.

Use is sometimes made of the negative plate-resistance characteristic that the screen-grid tube displays at low plate potentials. It will be recalled that the plate resistance of a tube is given by the reciprocal of the slope of the plate-current–plate-voltage characteristic. Hence the plate resistance is negative whenever the slope of this characteristic is negative. The negative resistance that can be realized from a screen-grid tube has a limited amplitude of current and voltage to which it can be subjected. It is further not very stable because secondary-emission characteristics are extremely variable. The negative resistance that can be realized will be different from tube to tube and will even change in the same tube with time.

10.3. Beam-power Tubes. The beam-power tube is a special tetrode designed to eliminate the interchange of secondary electrons between screen grid and plate. Historically, it was developed later than the pentode. Its development followed the discovery that when the screen-grid–plate distance in a tube was made rather long there was a maximum current which could be transmitted to the plate. This led to a study of the space-charge effects within the tube, which in turn led to the development of the final form of the beam-power tube.

The internal electrode arrangement of the beam-power tube is shown in Fig. 2.6. The distinctive features of the construction of this tube are the aligned control and screen grids of the same pitch. This is coupled with a flat cathode and side deflecting plates to keep the current sheets, which are formed by the aligned grids, from spreading. The screen-grid–plate spacing is made rather large, and the successive electrodes are curved so that they are at right angles to the electron flow.

The resulting plate-current–plate-voltage characteristics are shown in Fig. 2.7. It is seen that the dips in the current curves due to secondary emission have been eliminated at all but the very lowest control-grid voltages, and even here the dips are not very pronounced. The reason for this improvement in behavior is found in the space-charge effects that occur in the screen-grid–anode region. Before examining this subject in detail it is desirable to investigate briefly the electrostatic field of a beam-power tube.

10.4. The Electrostatic Field of a Beam-power Tube. The same general methods that have been described in the chapter on Triode Characteristics can be applied to multielectrode tubes in some cases. For tube structures in which the grid wires have a regular pattern the method of conformal transformations is easily applied. This is the case for the tetrode with aligned grids, the structure of the beam-power tube, which will be treated here by an extension of the method employed with the triode.

The line-charge configuration of Fig. 10.5*a* gives rise to the configuration of Fig. 10.5*b* upon application of the transformation $W = \frac{a}{2\pi} \ln Z$. The relation between the parts in the two planes is apparent from the previous study of triodes. The small circle about the origin in the Z plane goes into a cathode line in the W plane. The control-grid-wire circle at (1,0) in the Z plane goes into the series of equally spaced control-grid wires in the W plane. The screen-grid wire at $(\epsilon^{\frac{2\pi d_{gs}}{a}}, 0)$ in the Z plane goes into the line of screen-grid wires in the W plane. A large plate circle about the origin in the Z plane goes into the plate line in the W plane.

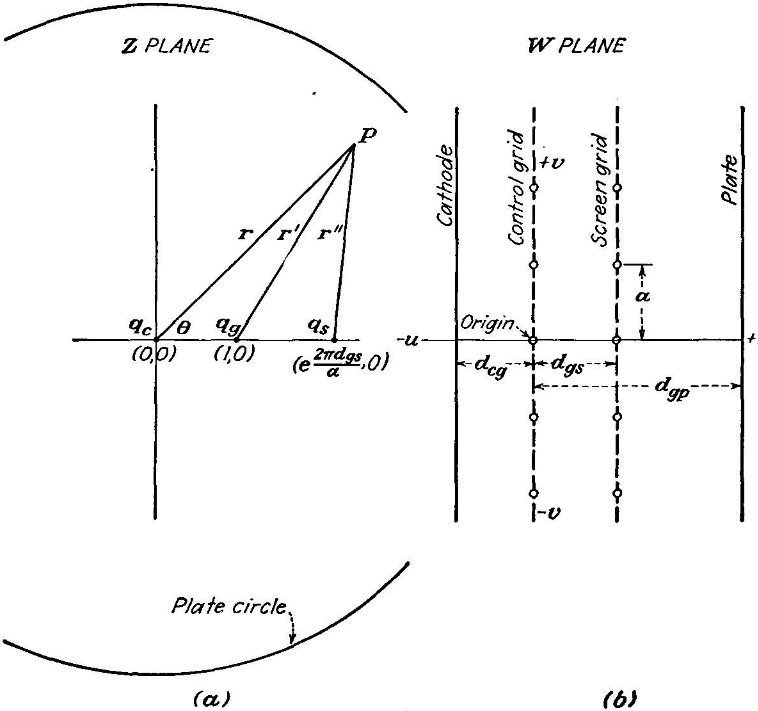


FIG. 10.5.—(*a*) Transformed beam-power tube, (*b*) beam-power-tube electrode arrangement.

It is necessary only to write an expression for the potential at any point P in the Z plane, transform it by the logarithmic transformation, and then evaluate the electrode potentials in terms of the charges and potentials. This is the procedure that was used for the triode, though the form of the resulting expressions may be expected to be more complicated because of the introduction of another electrode. In the treatment that follows the small-grid-wire approximations will be made.

The potential at any point P in the Z plane is given by

$$V = -\frac{q_c}{2\pi\varepsilon_0}\ln r - \frac{q_g}{2\pi\varepsilon_0}\ln r' - \frac{q_s}{2\pi\varepsilon_0}\ln r'' + C \qquad (10.2)$$

but

$$(r')^2 = (1 + r^2 - 2r\cos\theta) \tag{10.3}$$

and

$$(r'')^2 = (\epsilon^{\frac{4\pi d_{gs}}{a}} + r^2 - 2\epsilon^{\frac{2\pi d_{gs}}{a}} r\cos\theta) \tag{10.4}$$

Substitution of Eqs. (10.3) and (10.4) into Eq. (10.2) and application of the logarithmic-transformation coordinate equations gives

$$\begin{aligned} V = {}& -\frac{u}{a\varepsilon_0} q_c - \frac{q_g}{4\pi\varepsilon_0} \ln\left(1 - 2\epsilon^{\frac{2\pi u}{a}} \cos\frac{2\pi v}{a} + \epsilon^{\frac{4\pi u}{a}}\right) \\ & - \frac{q_s}{4\pi\varepsilon_0} \ln\left(\epsilon^{\frac{4\pi d_{gs}}{a}} - 2\epsilon^{\frac{2\pi(u+d_{gs})}{a}} \cos\frac{2\pi v}{a} + \epsilon^{\frac{4\pi u}{a}}\right) + C \end{aligned} \tag{10.5}$$

To determine the electrode potentials in terms of the charges and dimensions, let $u = -d_{cg}$, $v = 0$, where $d_{cg} \gg a$, and set the cathode potential equal to zero. This gives an expression for the constant C that can be put into subsequent expressions.

$$C = \frac{1}{a\varepsilon_0}(-q_c d_{cg} + q_s d_{gs}) \tag{10.6}$$

To find the control-grid potential let $u = 0$, $v = r_g$, where $r_g < \frac{a}{20}$. Then

$$V_g = -\frac{d_{cg} q_c}{a\varepsilon_0} - \frac{q_g}{2\pi\varepsilon_0} \ln\frac{2\pi r_g}{a} \tag{10.7}$$

To find the screen potential let $u = d_{gs}$, $v = r_s$ where $d_{gs} > a$. Then

$$V_s = -\frac{d_{sg} + d_{cg}}{a\varepsilon_0} q_c - \frac{d_{gs}}{a\varepsilon_0} q_g - \frac{q_s}{2\pi\varepsilon_0} \ln\frac{2\pi r_s}{a} \tag{10.8}$$

To find the plate potential let $u = d_{gp}$, $v = 0$, where $d_{gp} \gg 2a$. Then

$$V_p = -\frac{d_{gp} + d_{cg}}{a\varepsilon_0} q_c - \frac{d_{gp}}{a\varepsilon_0} q_g - \frac{d_{gp} - d_{gs}}{a\varepsilon_0} q_s \tag{10.9}$$

The last three equations give three electrode potentials in terms of three charges. The system can, of course, be solved for the charges in terms of the potentials. Solving for the cathode charge,

$$\begin{aligned} q_c = \frac{1}{\Delta a^2 \varepsilon_0{}^2} \Bigg\{ & \left[d_{gs} d_{sp} - \frac{a}{2\pi} d_{gp} \ln\left(\frac{2\pi r_s}{a}\right)\right] V_g - V_s \frac{a}{2\pi} \ln\left(\frac{2\pi r_g}{a}\right) d_{sp} \\ & + V_p \left(\frac{a}{2\pi}\right)^2 \ln\frac{2\pi r_g}{a} \ln\frac{2\pi r_s}{a} \Bigg\} \end{aligned} \tag{10.10}$$

where Δ is the determinant of the coefficients of the q's in Eqs. (10.7), (10.8), and (10.9). From Eq. 10.10 the grid-plate amplification factor is given by the ratio of the coefficients of V_g and V_p as

$$\mu_p = \frac{d_{gs}d_{sp} - \dfrac{ad_{gp}}{2\pi} \ln \left(\dfrac{2\pi r_s}{a}\right)}{\left(\dfrac{a}{2\pi}\right)^2 \ln \left(\dfrac{2\pi r_g}{a}\right) \ln \left(\dfrac{2\pi r_s}{a}\right)} \tag{10.11}$$

This expression is accurate to within a few per cent provided that the spacings between the various electrode planes are all greater than the grid-wire spacing and provided that the screening fraction of the grids (ratio of grid-wire diameter to grid-wire spacing) is less than 0.1.

It should be pointed out that the amplification factor derived above gives the relative effectiveness of the control grid and plate in controlling the *total space current* and not the plate current so that the above constant will not correspond exactly with that given in the tube manuals. That given in the tube manuals gives the relative effectiveness of the control grid and plate in controlling the *plate current,* and this depends upon the factor of Eq. (10.11) and also upon the way the space current divides between screen and plate. However, since the current-division function of a beam-power tube does not vary greatly with electrode potentials, the above expression for amplification factor is accurate enough for most purposes.

10.5. Space-charge Effects in the Screen-grid–Anode Region of Beam-power Tubes. In Sec. 8.9 of the chapter on Space-charge Effects it was shown that the effect of initial velocities in a diode was to create a virtual cathode between the actual cathode and the plate. Similarly, in tetrodes of proper design it is possible to get a virtual cathode or potential minimum between the screen grid and plate. If a satisfactory potential minimum can be achieved, it will suppress secondary electrons from the plate and do away with the need for a suppressor grid. Such a tetrode is the beam-power tube. It is in many respects superior to the conventional pentode.

In order that space-charge effects be appreciable, it is necessary that there be a very nearly parallel flow of electrons. This is not the case in the ordinary tetrode, for the use of control and screen grids with different pitches breaks up the electron flow. It is, however, possible to get what is nearly a parallel flow in a tetrode by making the control and screen grid have equal pitch and aligning the grid wires so that electrons which pass through the spaces of the control grid will also pass through the spaces in the screen grid. In Fig. 10.6 are shown some typical electron paths in a beam-power tube.[1] The paths give a sufficiently close approximation to a parallel flow in the screen-grid–plate region so that observed tube characteristics correlate well with theoretical properties deduced from this assumption.

Assuming a parallel flow of electrons starting at a high positive potential at the screen, a number of different potential distributions are possible depending upon the plate potential and the magnitude of the current injected into the screen-grid–anode region. Associated operating conditions are correspondingly different. The types of distributions encountered are shown in Fig. 10.7. The characteristics associated with these various distributions are best listed in tabular form. They are essentially determined by the sign of the constant which appears after the first integration of Poisson's equation as in Eq. (8.5) which may be written

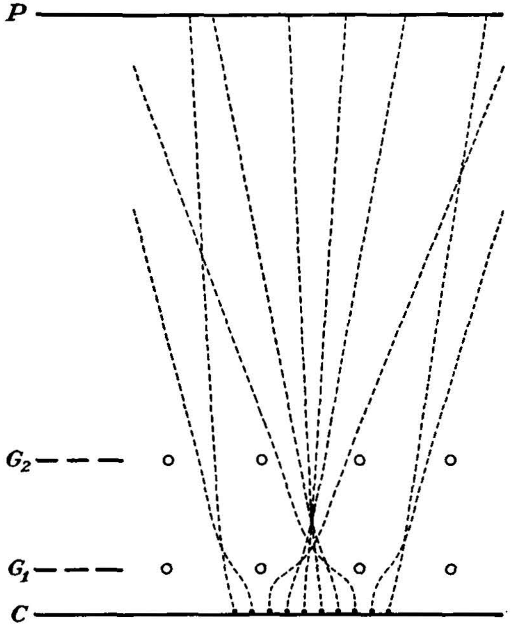


FIG. 10.6.—Electron paths in a beam-power tube. (*Jonker.*)

$$\left(\frac{dV}{dx}\right)^2 = \frac{16J}{9a^2}\,(V^{1/2} + C_1) \qquad (10.12)$$

---

[1] From JONKER, J. H. L., Pentode and Tetrode Output Valves, *Wireless Engr.*, vol. 26, [No. 189], pp. 274–286.

where $J = \frac{a^2 V^{3/2}}{d^2}$ and a^2 has the value 2.335×10^{-6} amperes per volt$^{3/2}$. The properties of the distributions as determined by the sign of the constant C_1 are as follows:

Type	V_p	C_1	$\frac{dV}{dx}$	Current transmission to plate	V
A	$-$	$+$	$-$	None	$+, 0, (-)$
			(Virtual cathode)		
B	0 or $+$	0	$-, 0, +$	Partial	$+, 0, +$
			(Potential minimum)		
C	$+$	$-$	$-, 0, +$	Complete	$+$
D	$+$	$-$	$+$	Complete	$+$

In types A and B there is a virtual cathode at the point of zero potential. In type C there is a potential minimum but no virtual cathode at the point of zero gradient of potential. It is seen that the current transmission is complete only when no virtual cathode exists. The various types of distribution will be analyzed in some detail in the following paragraphs. A number of extensive analyses of the space-charge effects in the grid-anode region of tubes have been published.[1–5] The treatment given here makes use of dimensionless parameters giving rise to universal characteristics as proposed by Fay, Samuel, and Shockley.

Type A Distribution. This type of distribution corresponds to that of a temperature-limited diode and is encountered when the plate is negative. The electrons injected into the screen-grid–plate space encounter a retarding field and are thus slowed down until they finally reach a zero velocity at some point before the plate, reverse, and return

---

[1] HARRIES, J. H. O., The Anode to Accelerating Electrode Space in Thermionic Valves, *Wireless Engr.*, vol. 13, pp. 190–199, April, 1936.

[2] PLATO, G., W. KLEEN, and H. ROTHE, The Space Charge Equations for Electrons with Initial Velocity, Part I, *Zeit. für Phys.*, vol. 101 [No. 5], pp. 509–520, 1936.

[3] KLEEN, W., and H. ROTHE, The Space Charge Equations for Electrons with Initial Velocity, Part II, *Zeit. für Phys.*, vol. 104 [Nos. 11, 12], pp. 711–723, 1937.

[4] SALZBERG, B., and A. V. HAEFF, Effects of Space Charge in the Grid-anode Region of Vacuum Tubes, *RCA Rev.*, vol. 2, pp. 336–374, January, 1938. Excellent discussion of dynamic characteristics.

[5] FAY, C. E., A. L. SAMUEL, and W. SHOCKLEY, On the Theory of Space Charge between Parallel Plane Electrodes, *Bell Sys. Tech. Jour.*, vol. 17, pp. 49–79, January, 1938.

to the screen. There is thus a virtual cathode at the point of zero potential. The potential distribution from the virtual cathode to the negative plate is linear.

The equations for the relations between potential, distance, and current are obtained by letting the constant in Eq. (10.12) assume the positive value of $m^{1/2}V_1{}^{1/2}$, where m is related to the slope of the potential-distribution curve as will be shown and V_1 is the screen potential. The differential equation then has the form

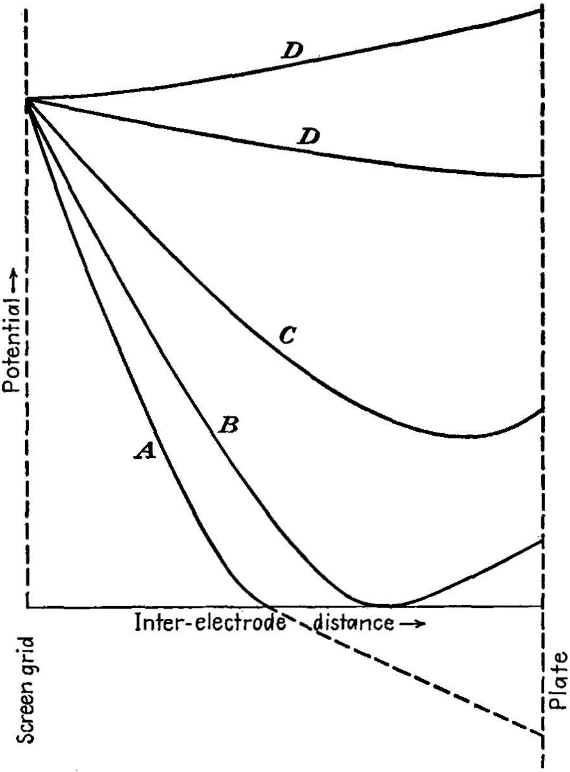


FIG. 10.7.—Types of potential distributions in the screen-grid–plate region of a beam-power tube.

$$\left(\frac{dV}{dx}\right)^2 = \frac{16(2J)}{9a^2}\,(V^{1/2} + m^{1/2}V_1{}^{1/2}) \tag{10.13}$$

or

$$\frac{dV}{dx} = \frac{-4\cdot 2^{1/2}J^{1/2}}{3a}\,(V^{1/2} + m^{1/2}V_1{}^{1/2})^{1/2} \tag{10.14}$$

Let the following changes in variables be introduced:

$$\phi = \frac{V}{V_1} \tag{10.15}$$

$$\sigma = \frac{x}{x_0} \tag{10.16}$$

where $x_0 = \dfrac{aV_1^{3/4}}{J^{1/2}}$ is the distance over which a potential V_1 will produce a current density J in a space-charge-saturated diode. The Eq. (10.14) becomes

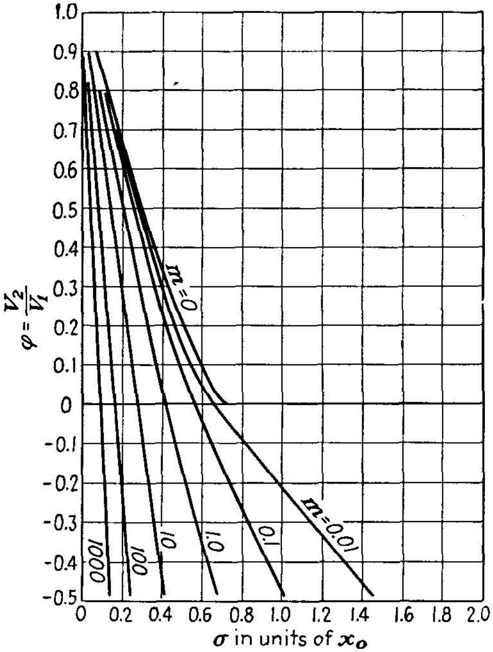


FIG. 10.8.—Potential-distribution curves of the type A.

$$\frac{d\phi}{d\sigma} = \frac{-4 \cdot 2^{1/2}}{3} (\phi^{1/2} + m^{1/2})^{1/2} \tag{10.17}$$

This may be written as

$$\frac{d\phi}{(\phi^{1/2} + m^{1/2})^{1/2}} = \frac{-4 \cdot 2^{1/2}}{3}\, d\sigma \tag{10.18}$$

for convenience of integration. Let $d\phi = 2\phi^{1/2}\, d(\phi^{1/2})$, and then, upon integration,

$$-(\phi^{1/2} - 2m^{1/2})(\phi^{1/2} + m^{1/2})^{1/2} = 2^{1/2}\sigma + C_2 \tag{10.19}$$

Since $\phi = 1$ when $\sigma = 0$,

$$2^{1/2}\sigma = (1 - 2m^{1/2})(1 + m^{1/2})^{1/2} - (\phi^{1/2} - 2m^{1/2})(\phi^{1/2} + m^{1/2})^{1/2} \qquad (10.20)$$

The factor $2^{1/2}$ has been introduced because the reversal of current at the virtual cathode has required that J be replaced by $2J$ in the above derivation. Curves of the type A obtained from Eq. (10.20) are shown in Fig. 10.8. If the slope of the curves be evaluated it is seen that

$$\frac{d\phi}{d\sigma} = \frac{-4\cdot 2^{1/2}}{3}\, m^{1/4} \qquad \text{for } \phi \leqq 0 \qquad (10.21)$$

and

$$\frac{d\phi}{d\sigma} = \frac{-4\cdot 2^{1/2}}{3}\, (1 + m^{1/2})^{1/2} \qquad \text{for } \sigma = 0 \qquad (10.22)$$

It will be recognized that the potential distributions resulting in this case are the same as those encountered in the temperature-limited diode, the only difference being that the current is flowing in equal amounts in both directions and is in this case injected at a positive rather than at a zero potential.

Type B Distribution. This occurs when the integration constant C_1 is zero and as can be seen from simple physical considerations *gives rise to a Child's law distribution on each side of a virtual cathode that exists at the point of zero potential and zero gradient.* Let it be assumed that, of the injected current, a fraction P is transmitted beyond the virtual cathode. Then the net current on the screen-grid side of the virtual cathode, as far as its space-charge effects are concerned, is $(2 - P)J$. Child's law then assumes the form

$$(2 - P)J = \frac{a^2 V^{3/2}}{x_1^2} \qquad (10.23)$$

but since $J = \dfrac{a^2 V_1^{3/2}}{x_0^2}$, then in terms of the factors ϕ and σ

$$\sigma_1^2 = \frac{\phi^{3/2}}{(2 - P)} \qquad (10.24)$$

Since the actual potential factor is $\phi = 1$ when σ is zero and the potential decreases with increasing σ, the relation must be put into the form

$$\sigma_L = \frac{1 - \phi^{3/4}}{(2 - P)^{1/2}} \qquad (10.25)$$

in which the subscript L indicates that the relation holds to the left of the virtual cathode for values of ϕ between 0 and 1 and σ is measured from the point of current injection.

To the right of the virtual cathode the current density is PJ so that

$$PJ = \frac{a^2 V^{3/2}}{x^2} \tag{10.26}$$

where x is now measured from the virtual cathode. Combining this with the Child's law relation,

$$\sigma^2 = \frac{\phi^{3/2}}{P} \tag{10.27}$$

Actually, the potential factor is zero when σ has a value of $\dfrac{1}{(2 - P)^{1/2}}$ as may be seen from Eq. (10.25), so that the desired relation is

$$\sigma_R = \frac{1}{(2 - P)^{1/2}} + \frac{\phi^{3/4}}{P^{1/2}} \tag{10.28}$$

where the subscript R indicates that the expression holds only to the right of the virtual cathode and σ is again measured from the point of current injection. Curves of the type B as determined from Eqs. (10.25) and (10.26) are shown in Fig. 10.9. It will be recognized that these are all three-halves-power-law curves drawn with different scales from both sides of the virtual cathode.

A curve of considerable importance in the family (Fig. 10.9) is the limiting curve that gives the maximum value of ϕ for a fixed value of σ to the right of the virtual cathode. This is an envelope to the family of type B curves. If the expression of Eq. (10.28) be solved for ϕ and maximized with respect to P, there results

$$\sigma = \frac{2}{(2 - P)^{3/2}} \tag{10.29}$$

When this is substituted in Eq. (10.28), there is obtained

$$P = \frac{2\phi^{1/2}}{1 + \phi^{1/2}} \tag{10.30}$$

The factor P can be eliminated between Eqs. (10.29) and (10.30) to give the relation between ϕ and σ.

$$\sigma = (1 + \phi^{1/2})^{3/2} 2^{-1/2} \tag{10.31}$$

This curve is shown dotted in Fig. 10.9. Equation (10.30) tells what the maximum transmitted current for any plate potential is. If the attempt is made to increase the transmitted current beyond this value, the distribution will jump from a type B to a type C or D distribution.

Type C Distribution. This type of distribution is characterized by the existence of a potential minimum that is not at zero potential. The distributions are obtained by letting $C_1 = -(\alpha V_1)^{1/2}$ in Eq. (10.12).

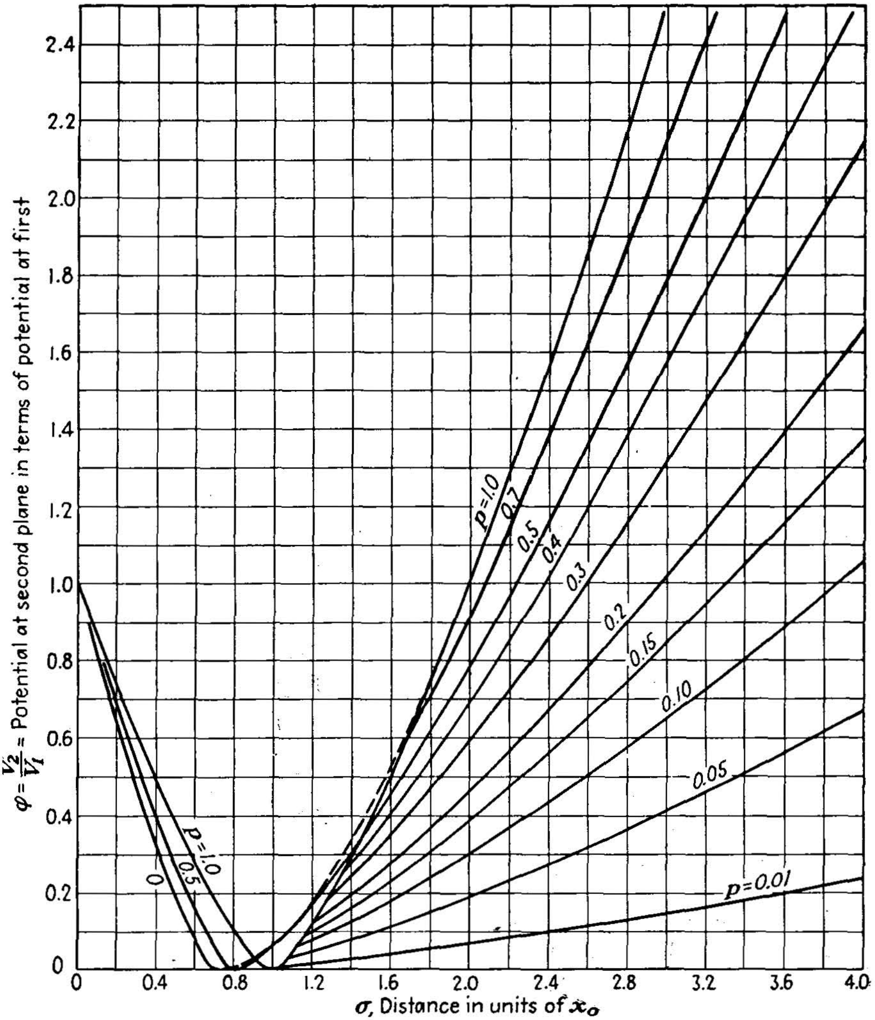


FIG. 10.9.—Potential-distribution curves of the type *B*.

This gives a positive value of V, equal to αV_1, when $\frac{dV}{dx}$ equals zero. Integration of Eq. (10.12) with the above value of the constant gives

$$\sigma = \pm(\phi^{1/2} + 2\alpha^{1/2})(\phi^{1/2} - \alpha^{1/2})^{1/2} + C_2 \tag{10.32}$$

in which the negative sign goes with a negative slope at the screen and the positive sign gives a distribution of the type *D*. Since ϕ is unity when σ is zero,

$$\sigma_L = -(\phi^{1/2} + 2\alpha^{1/2})(\phi^{1/2} - \alpha^{1/2})^{1/2} + (1 + 2\alpha^{1/2})(1 - \alpha^{1/2})^{1/2} \tag{10.33}$$

which holds to the left of the potential minimum with σ measured from the point of current injection. The distance at which the potential minimum exists is found by setting $\phi = \alpha$.

$$\sigma_{\min} = (1 + 2\alpha^{1/2})(1 - \alpha^{1/2})^{1/2} \tag{10.34}$$

To the right of the potential minimum the potential distribution is given by

$$\sigma_R = (\phi^{1/2} + 2\alpha^{1/2})(\phi^{1/2} - \alpha^{1/2})^{1/2} + (1 + 2\alpha^{1/2})(1 - \alpha^{1/2})^{1/2} \tag{10.35}$$

σ being measured from the screen grid. The slope of the potential-distribution curves at the screen is given by

$$\frac{d\phi}{d\sigma} = -\frac{4}{3}(1 - \alpha^{1/2})^{1/2} \tag{10.36}$$

Curves of the *C* type are shown in Figs. 10.10*a* and *b*. Various limiting curves are of interest. By letting ϕ equal α there is obtained the curve which passes through all the minima and of which the equation is

$$\sigma = (1 + 2\phi^{1/2})(1 - \phi^{1/2})^{1/2} \tag{10.37}$$

This curve is shown dotted in Fig. 10.10*a*.

By setting α equal to zero, another limiting curve is obtained,

$$\sigma_R = 1 + \phi^{3/4} \tag{10.38}$$

which is the boundary between the *B* and *C* type of curve. This curve runs through the field of the type *C* curves because of the way in which the curves overlap. The significance of the overlap curves of Fig. 10.10*b* is that two potential distributions are possible for one set of electrode potentials.

By setting α equal to unity,

$$\sigma = (\phi^{1/2} + 2)(\phi^{1/2} - 1)^{1/2} \tag{10.39}$$

which sets an upper limit to the type *C* curves.

Another limiting curve is obtained by making σ_R a maximum with respect to α and holding ϕ constant. This gives

$$\alpha = \phi(1 + \phi^{1/2})^{-2} \tag{10.40}$$

and

$$\sigma = (1 + \phi^{1/2})^{3/2} \tag{10.41a}$$

In these expressions σ and ϕ are coordinates of a lower-limit envelope that is tangent to the type C curves. The parameter α determines which curve is tangent to the envelope at the point in question. Type C distributions cannot exist beyond this condition. If current or voltage is changed beyond this boundary, the distribution jumps to a type B curve.

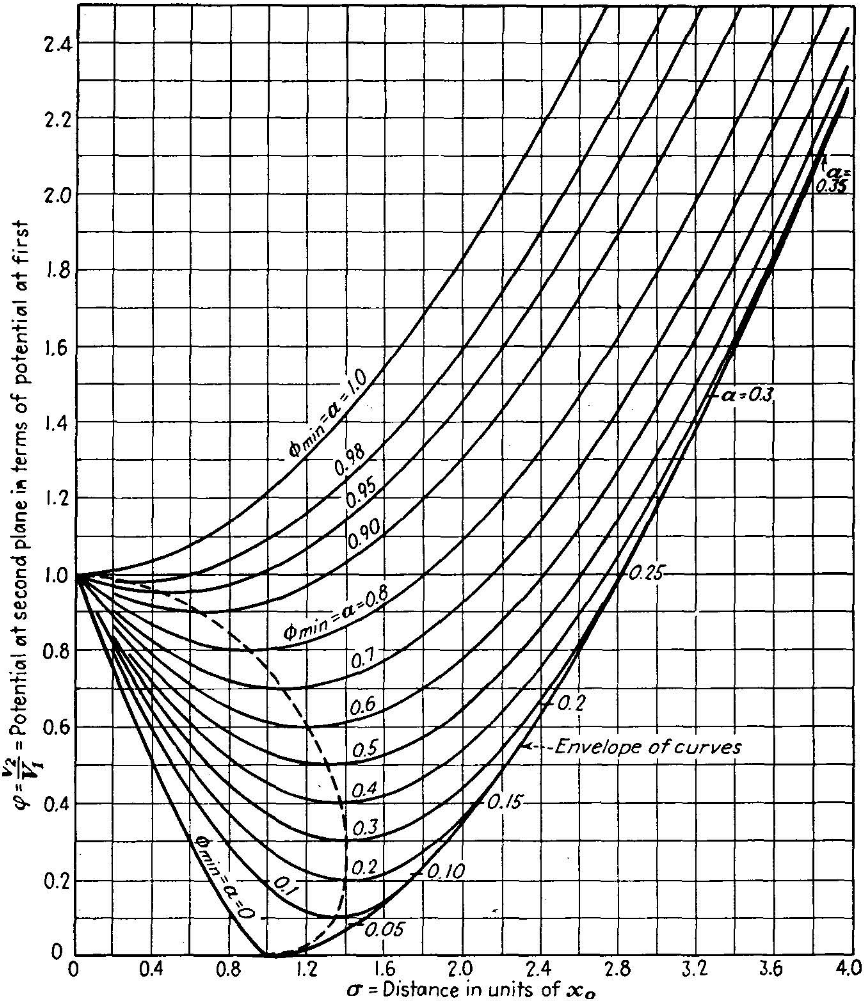


FIG. 10.10*a*.—Potential-distribution curves of the type C.

From Eq. (10.41*a*) is obtained the expression that gives the maximum current that can be transmitted between electrodes at potentials V_1

and V_2 separated by a distance x. The limiting current density is

$$J_{\max} = \frac{2.335 \times 10^{-6}(V_1^{1/2} + V_2^{1/2})^3}{x^2} \quad \text{amperes per unit area} \qquad (10.41b)$$

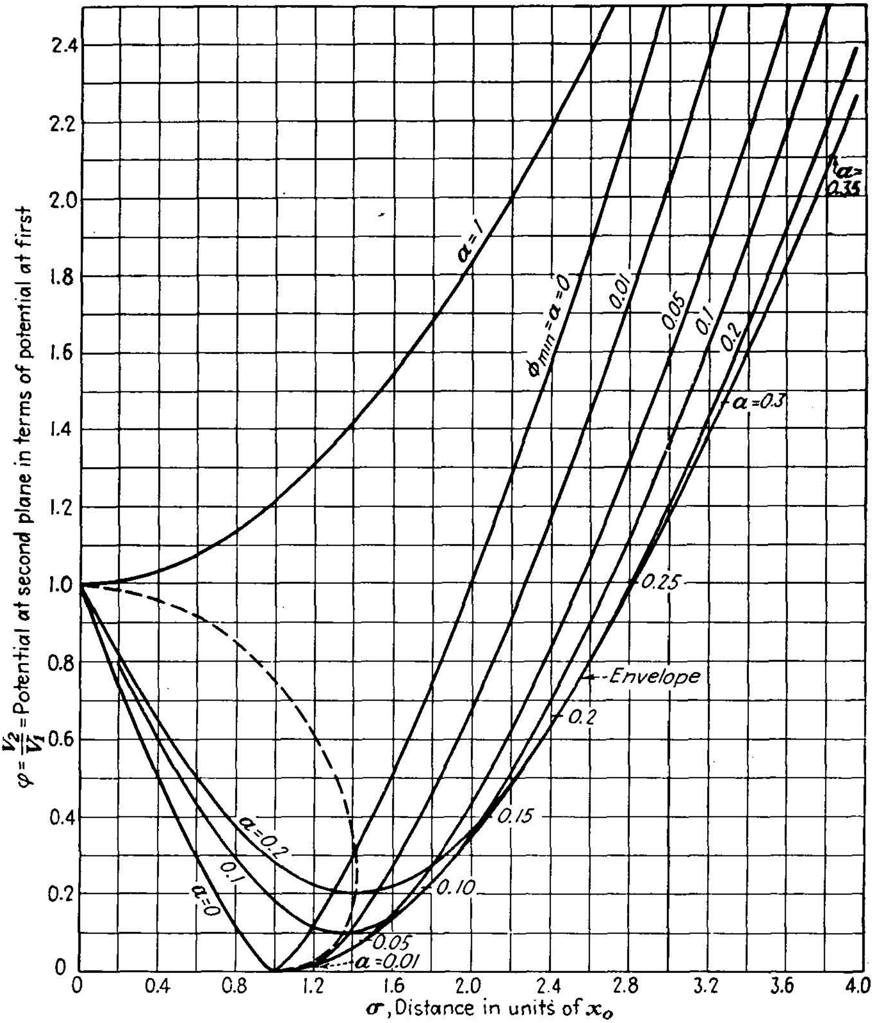


FIG. 10.10*b*.—Potential-distribution curves of the type *C* overlap.

If the injected current exceeds this amount, the potential distribution will jump from a type *C* to a type *B* distribution, with an attendant reduction in transmitted current.

Examination of the curves of Fig. 10.10 and their equations shows that the curves are all of the same form but differ only in scale and position. Hence the upper limiting curve for which α equals 1 is a universal curve showing how the potential varies on either side of the potential minimum. The universal form of this curve is obtained by setting α equal to 1 in Eq. (10.35), giving the universal form

$$\sigma_u = (\phi^{1/2} + 2)(\phi^{1/2} - 1)^{1/2} \tag{10.42}$$

which holds on either side of the minimum at which the potential value is now V_1 and σ_u is measured from the minimum. Current is introduced into this expression by the relation $J = \dfrac{a^2 V_1^{3/2}}{x_u^2}$, and $\sigma_u = \dfrac{x}{x_u}$.

Type D Distributions. These include curves of the C type restricted to the region of the curve before the potential minimum is reached. They also include curves that start with a positive gradient and increase. This latter type is given by using the positive sign in Eq. (10.32). It is not of much practical importance. Since the curves of the D type are included in the other types previously discussed, they will not be discussed in detail.

10.6. Dynamic Characteristics of Beam-power Tubes. In the above discussion of the different types of potential distribution possible it has been indicated that there are limiting conditions under which the separate types could exist. It is also true that several potential distributions are possible for a given set of externally imposed conditions. In actual tube operation this means that there may be discontinuities in the current-voltage characteristics; for as potential conditions are changed, the internal distributions may jump from one form to another and these changes are sometimes accompanied by changes in the fraction of the current transmitted to the plate. Furthermore, it sometimes happens that there may be hysteresis effects as a cycle of voltages is impressed upon a tube in that the current cycle produced does not retrace itself exactly.

The beam-power tube makes use of a potential minimum produced by a type C distribution to reduce the secondary emission from the plate. As long as the potential minimum is 20 or more volts more negative than the plate, very effective suppression of secondary electrons is achieved. This expedient dispenses with the need for a suppressor grid but may cause the dynamic characteristics to be different from those of other tubes.

Two of the most important dynamic characteristics will be discussed in a qualitative fashion. Quantitative analyses have been given,[1] but these are somewhat limited in value in that the ideal conditions of parallel electron flow cannot be realized exactly in any actual tube.

Injected Current Varied, Potentials Constant. One case that is of

---

[1] SALZBERG and HAEFF, *loc. cit.*

considerable interest is that in which the screen and plate of a beam-power tube are maintained at the *same* potential and the current injected into the screen-grid–anode region is increased from zero to a large value and then decreased. The changes in the potential distribution encountered in the tube are shown in Fig. 10.11. Initially, for no injected current, the potential distribution from screen grid to plate is a straight horizontal line as shown at *a*. As the injected current is increased, the potential-distribution curve is depressed, assuming the form of the type *C* distributions as shown at *b*, in this case symmetrical with respect to a potential minimum at the center. As the injected current is further increased, the potential-distribution curve is depressed still further, maintaining its symmetry. The physical reasons for the action are quite apparent. As the current is increased, the space charge is increased, which reduces the potential, which decreases the electron velocity, which increases the space charge still further and thus depresses the potential still more. Thus an increase in injected current starts a cycle of action that is very sensitive to changes in current, so much so that an equilibrium may not always be reached. This occurs *in this case* when the potential curve has been moved about three-quarters of the way down to a zero potential, as at *c*, at which point any further increase in current causes the potential curve to drop as far as it will go because of the instability in the sequence of actions described above. The potential

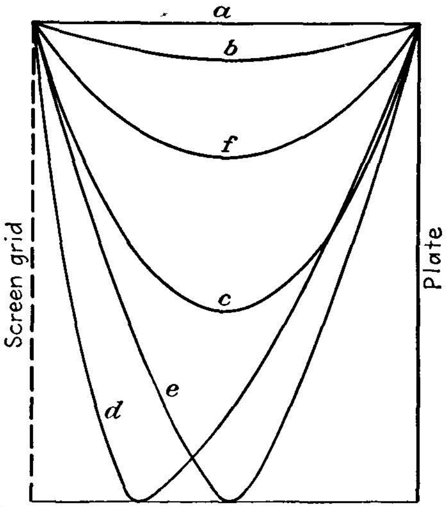


FIG. 10.11.—Potential distributions in a beam-power tube as the injected current is varied.

curve can drop only to zero, at which point a virtual cathode is formed halfway between grid and plate. When this occurs, part of the current, which has previously all been going to the plate, is turned back toward the grid, thus increasing the space charge or effective current to the left of the virtual cathode. Child's law demands that when the current is increased without a change in potential the distance must be decreased so that the virtual cathode moves toward the screen, finally coming to rest at some position, as shown at *d*. Thus the distribution changes immediately from that at *c* to that at *d*, with an abrupt reduction in current to the plate. Any further increase in injected current increases the space charge and current on the screen side of the virtual cathode and causes it to move closer to the screen, with a further reduction in transmitted current.

If now the current is decreased, the sequence of operations will not be exactly the same, for the initial conditions are different. For a given set of voltages and current, two potential distributions may be possible but only one can exist at a time, of course, and the physical choice between the possible distributions is determined by the order in which limiting conditions are established. If the current is decreased, the virtual cathode moves toward the plate, an increasing fraction of the current going to the plate. Finally the virtual cathode reaches the mid-point, all the current going to the plate. The virtual cathode is now "saturated." The potential field and electron paths for such a condition are shown in Fig. 10.12. The type *B* distribution cannot exist with any smaller injected current, and thus a further decrease causes the potential distribution to jump to the type *C* distribution, jumping from the distribution at *e* to that at *f*. Any further decrease in current now maintains the same type of symmetrical distribution, the potential minimum rising until finally it is flat, with no current. Because of the fact that there is a maximum value of plate current for any set of screen and plate voltages in a beam-power tube, difficulties may sometimes be encountered with pulsed operation.

Although the above discussion has been given for *equal screen and plate voltages*, the same sort of behavior results if the electrode potentials are not the same. In general, specific limiting conditions will be different for different cases.

The associated current behavior is shown in Fig. 10.13, which shows the relation between the plate current and the injected current. As the injected current is increased, at first all the current is transmitted to the plate, giving the straight-line characteristic shown. When the potential distribution jumps from *c* to *d*, the plate current suddenly drops in value and then decreases uniformly, as shown, as the injected

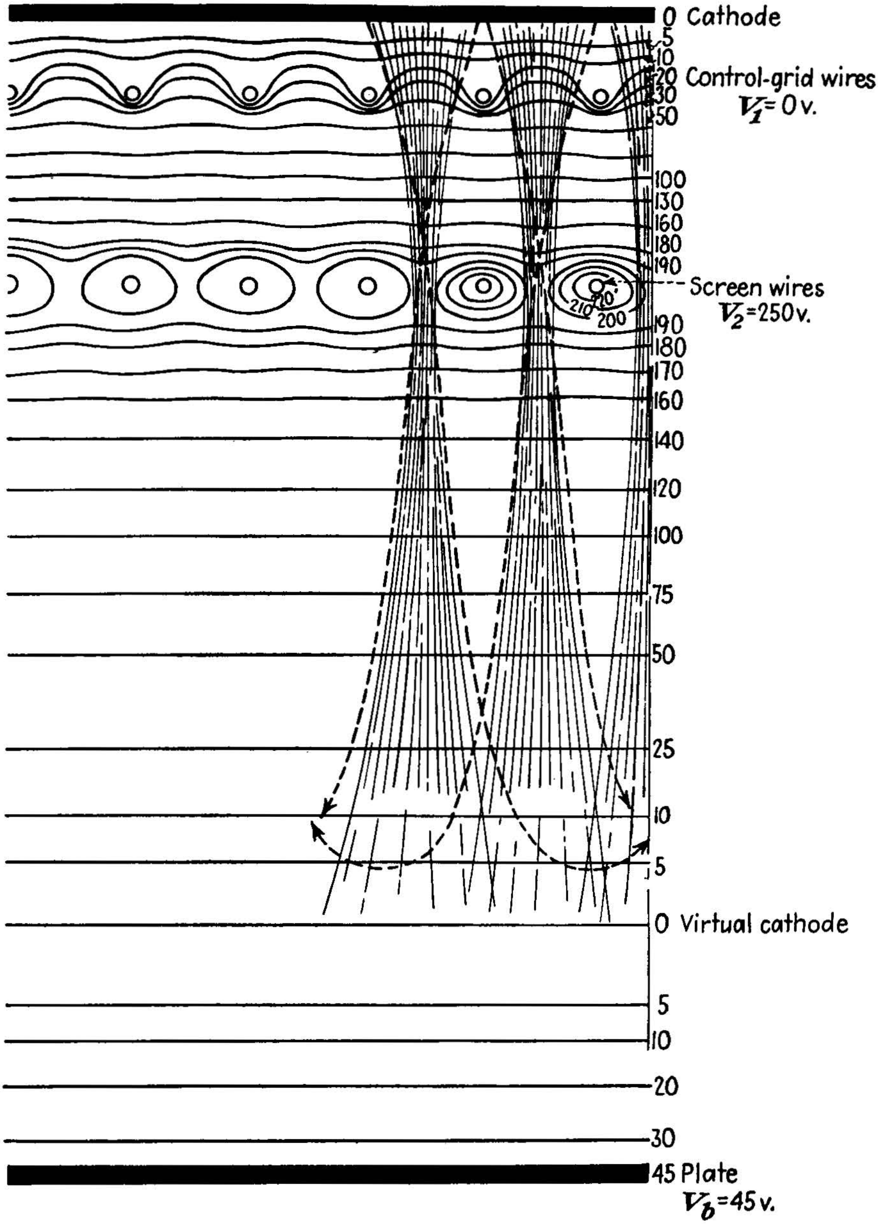


FIG. 10.12.—Potential field of a beam-power tube with a saturated virtual cathode (*Schade.*)

current is increased. When the injected current is then decreased, the transmitted current is increased uniformly until the distribution shown at *e* in Fig. 10.11 is reached. In this case the jump in distribution from *e* to *f* produces no change in current, though it will be observed that the highest current reached on the retrace of the cycle is less than that obtained as the injected current was increased. All the injected current now goes to the plate again; and as the injected current is decreased further, the plate current decreases correspondingly, moving down the straight-line portion of the curve of Fig. 10.13. The portion at the extreme right of the current characteristic is seen to exhibit a negative transconductance since the injected current is a continuous function of the control-grid voltage. Oscillators have been built utilizing this property. The characteristic shown in Fig. 10.13 can actually be observed on an oscilloscope if a beam-power tube is connected so that the vertical deflection is proportional to plate current and the horizontal deflection proportional to space current as the control-grid voltage is varied sinusoidally.

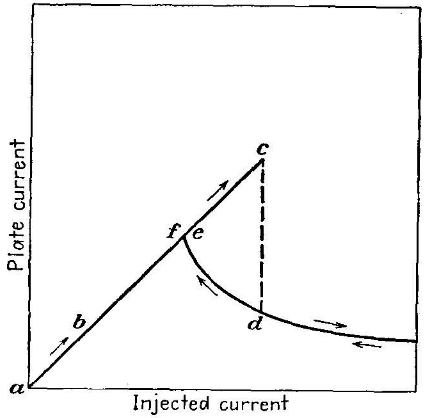


FIG. 10.13.—Transmitted current in a beam-power tube as a function of injected current.

Plate Potential Varied, Screen Potential and Injected Current Constant. Another case of operation which is of particular importance is that which occurs when the plate potential alone is varied. Consider the case in which the injected current is quite high, corresponding to a positive control-grid voltage on a beam-power tube. Starting with a negative plate potential, the potential distribution is of the type *A* (temperature-limited), as shown at *a* in Fig. 10.14. When the plate potential reaches a value of zero, a distribution of the type *B* (space-charge-limited) exists as shown at *b*. Then, as the potential is further increased, a three-halves-power-law distribution holds on each side of the virtual cathode, which moves toward the plate. Finally, a limiting type *B* curve is reached, and even though all the current is not being transmitted to the plate the distribution jumps from that at *c* to that at *c*′, giving a type *C* (potential minimum) distribution with a complete instead of a partial transmission of current. As the potential is further increased, the potential minimum moves toward the screen as the curve moves up. Then, as the cycle is reversed and the potential is decreased, the curve moves down through the stages indicated by *e* and *f*. At *g* there is reached a limiting curve of the type *C*, and the distribution jumps to that at *g*′, giving a partial transmission of current. From this the virtual cathode moves toward the screen as the potential and plate current decrease to zero. The curve *a* is obtained again as the plate voltage is made negative.

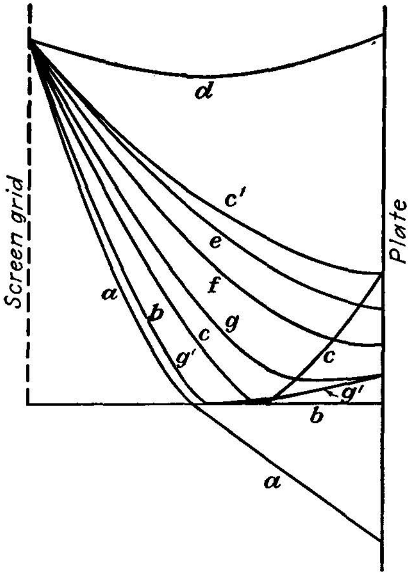


FIG. 10.14.—Potential distributions in a beam-power tube as plate voltage is varied.

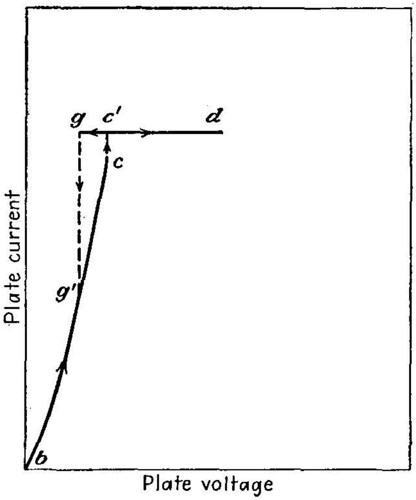


FIG. 10.15.—Plate current in an ideal beam-power tube as a function of plate voltage.

The corresponding current behavior is shown in Fig. 10.15, in which is shown the variation of plate current with plate potential. Plate current begins to flow at *b* and continues to increase until *c* because of the partial transmission of current. If the position of the virtual cathode remained fixed, the current would increase as the three-halves power of the plate potential in this region. Since it moves toward the plate, the current increases faster than the three-halves power, giving a very steep characteristic and accounting for the sharp shoulder of the plate-current–plate-voltage curves of the beam-power tube. At *c* the current jumps to that at *c′*, at which the transmission is complete and the current remains constant for further increases in potential. When the cycle is reversed, the current remains constant as the potential is decreased until *g* is reached, at which it drops to the value on the curve between *b* and *c*. It will be noted that the jump in current on the retrace occurs at a lower value of potential than when the potential is increasing.

The type of behavior described above occurs for other values of current, though the effects are most pronounced when the current is high. At lower values of current, potential minima may not be formed. The area of the hysteresis loop is in all cases quite small and becomes smaller as the current is decreased. In an actual tube such as the 6L6, having the characteristics shown in Fig. 2.7, the distribution of initial velocities and more particularly the spreading of the beam sheets cause the current characteristic to depart from the idealized behavior indicated here.[1] Sometimes at high currents, instead of the jumps indicated, the curve will run through a small s giving rise to a small region with a negative plate resistance. The kinks are observed only for high currents. At lower currents there is a sharp shoulder. At very low currents, potential minima are not formed; and as a result secondary electrons from the plate are not suppressed, and the curves have the characteristic depressions associated with this effect. It is also not true that the current is completely independent of the plate potential when the current transmission is complete, for the cathode region is not completely shielded from the plate.

---

[1] SCHADE, O. H., Beam Power Tubes, *Proc. I.R.E.*, vol. 26, pp. 137–181, February, 1938.

CHAPTER 11

PENTODES

11.1. Electrode Arrangement in a Pentode. The pentode, as its name implies, is a five-electrode tube. The five electrodes, in order, are cathode, control grid, screen grid, suppressor grid, and plate. The effects of all these electrodes except the suppressor grid have been studied.

The suppressor grid was added to the screen-grid tube to eliminate the exchange of secondary electrons between screen grid and plate. It is invariably a coarse-mesh grid placed between the screen grid and plate and operated at cathode potential. At this potential it is able to suppress secondary electrons by causing a deep dip in the potential between screen grid and plate while at the same time its coarse mesh allows electrons to pass on through it to the plate. The potential profiles of a pentode are shown in Fig. 11.1. From these it is seen that both the plate and the grid present negative gradients of potential to secondary electrons created at their surface. This eliminates the exchange of secondary electrons between screen grid and plate and results in current-voltage characteristics which are almost exactly those which would occur in a perfect screen-grid tube having no secondary emission.

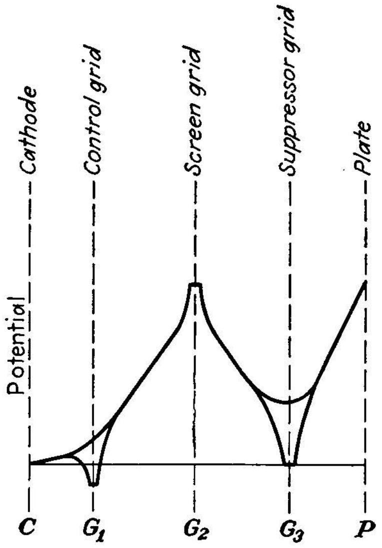


FIG. 11.1.—Potential profiles in a pentode.

Of all the various types of vacuum tube, the pentode is probably the one in most extensive use. For voltage amplification whether at audio or radio frequencies it is the invariable choice. It is used at audio frequencies because a higher gain per stage can be realized than with a triode. It is used at radio frequencies because the extremely low control-grid-to-plate capacity virtually eliminates the possibility of regeneration. Even as a power tube, it finds considerable use because its control-grid current and hence the power necessary to drive it are lower than for the corresponding triode.

11.2. Current-voltage Characteristics of the Pentode. As stated above, the suppression of secondary electrons in the pentode gives it the characteristics of a screen-grid tube that is free of secondary emission. The addition of the extra grid increases the extent to which the control grid is shielded from the plate and results in a somewhat higher amplification factor and a somewhat higher plate resistance.

Plate-current–Control-grid Voltage Characteristics. The plate-current–control-grid characteristics of a pentode are similar to those of a tetrode and not greatly different from those of a triode. The plate current is most easily influenced by the control grid, less so by the screen grid, and hardly at all by the plate. Some typical plate-current–control-grid voltage characteristics of a pentode are shown in Fig. 11.2. Here there is shown a group of curves for different screen-grid voltages. These curves are almost identical with the corresponding curves in a triode. If a group of curves for different plate voltages were shown, they would be very closely grouped and, for the same screen-grid potential, would have the same cutoff potential.

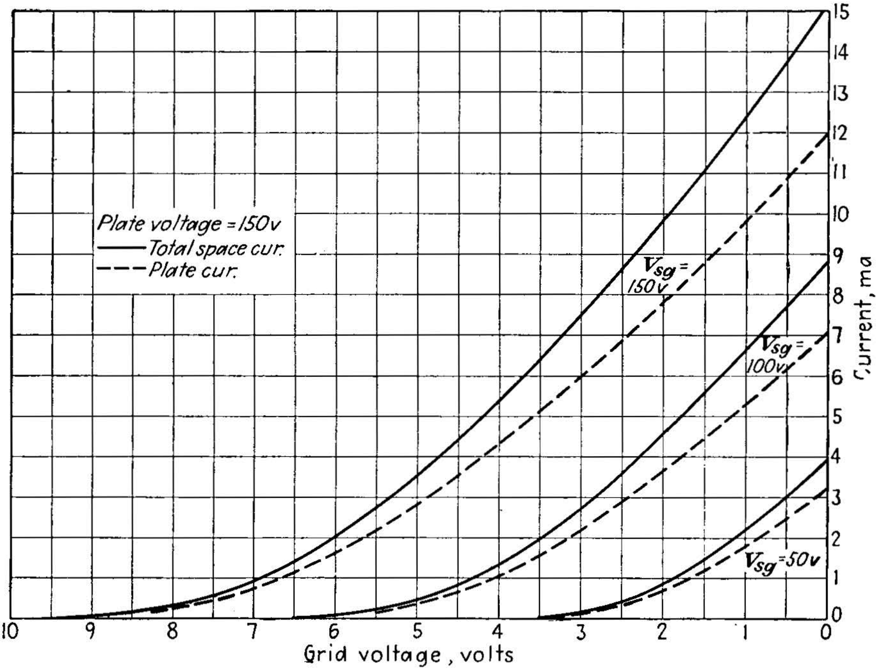


FIG. 11.2.—Plate-current–control-grid voltage characteristics of a type 6J7 pentode.

Plate-current–Plate-voltage Characteristics of a Pentode. The plate-current–plate-voltage characteristics of a pentode are shown in Fig.

11.3*a*. In a well-designed pentode the plate current varies only slightly with plate potential for all plate voltages greater than 50 per cent of the screen-grid voltage. Below this value of voltage the current falls rapidly to zero. The magitude of the space current in a pentode is determined almost entirely by the control-grid and screen-grid potentials. The plate potential determines only what fraction of the space current is transmitted to the plate. It does, of course, have a second-order influence upon the plate current, with the result that the plate current rises slowly as the plate voltage is increased, but this rise is even slower than in the screen-grid tube, where there is only one shielding grid between the plate and the control grid. As a result of the action of the plate voltage in determining the fraction of the space current that is transmitted to the plate, all the plate-current–plate-voltage curves are similar in shape and differ only in scale.

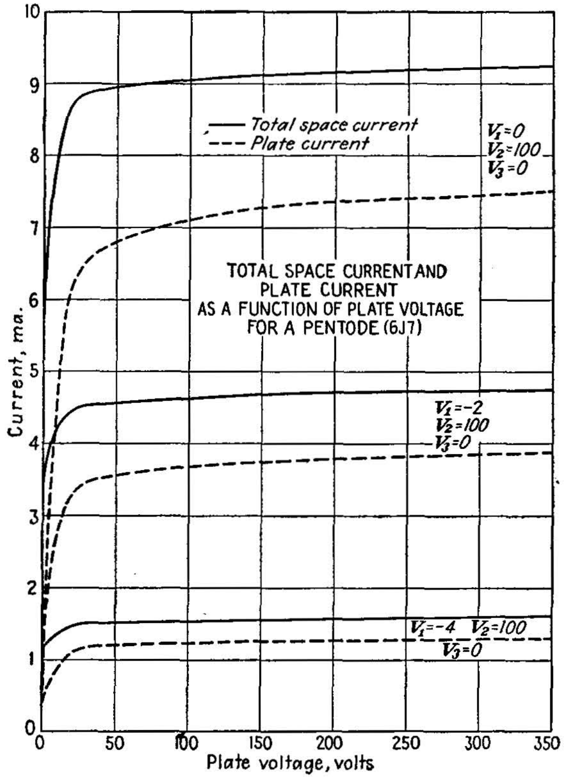


FIG. 11.3*a*.—Plate-current–plate-voltage characteristics of a type 6J7 pentode.

Space-current–Plate-voltage Characteristics of the Pentode. Also shown in Fig. 11.3*a* are the space-current–plate-voltage characteristics of a pentode. The space current is even more constant with plate voltage than is the plate current. The only departure from near constancy occurs near zero plate potential. Here the space current increases by about 20 to 40 per cent as the plate potential is increased to about half of the screen-grid potential. This increase in space current occurs because there is a change in the space-charge condition around the screen grid as the condition of reflection of electrons from the plate changes to one of transmission. The action is exactly the same as that which was encountered in screen-grid tubes (see Fig. 10.2).

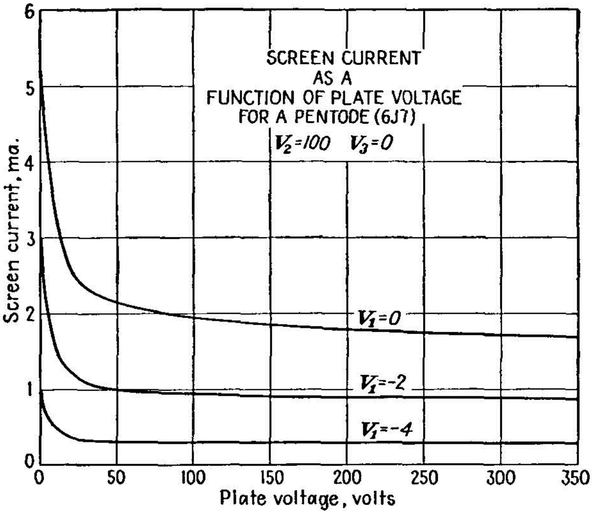


FIG. 11.3*b*.—Screen-current–plate-voltage characteristics of a type 6J7 pentode.

Screen-grid-current–Plate-voltage Characteristics of a Pentode. The screen-grid current in a pentode is the difference between the space current and the plate current, provided that the other grids in the tube are drawing no current. The nature of the screen-grid-current variation with plate voltage is shown in Fig. 11.3*b*. The screen-grid current is seen to have a uniformly decreasing characteristic with plate voltage. The screen current will generally lie between one-fifth and one-third of the plate current at large plate voltages. The effectiveness of the suppressor grid in suppressing secondary electrons is so complete that the screen current rarely shows even a trace of distortion due to this cause.

Suppressor-grid Effects. With small pentodes such as are used for voltage amplification the suppressor grid is operated at cathode potential and is not used to influence the plate current. In power-output pentodes, however, the suppressor grid may be used as an active electrode. It can be used to modulate the plate current in amplifiers or oscillators.[1] Use is here made of the fact that the suppressor grid is able to control the fraction of the current transmitted past the plane of the screen grid that goes on to the plate. When the suppressor grid is at a low potential relative to plate and screen grid, as it usually is, it can sort out the electrons having a large component of energy directed toward the plate from those which, because of deflection on passing close to a screen-grid wire, have a lower component of plate-directed energy. Some typical curves showing the current transmitted to the plate as a function of suppressor-grid voltage are shown in Fig. 11.4. It is seen that the plate current is moderately sensitive to suppressor-grid voltage and that the suppressor grid is readily capable of completely cutting off the plate current. The action of the suppressor grid in controlling the plate current is a combination of its action as a velocity sorter and a direct control on the gradient of potential before the suppressor grid. At voltages near zero the first action predominates. As the suppressor grid is made more negative, the second action becomes predominant. As the suppressor grid approaches cutoff, there is a strong tendency for a virtual

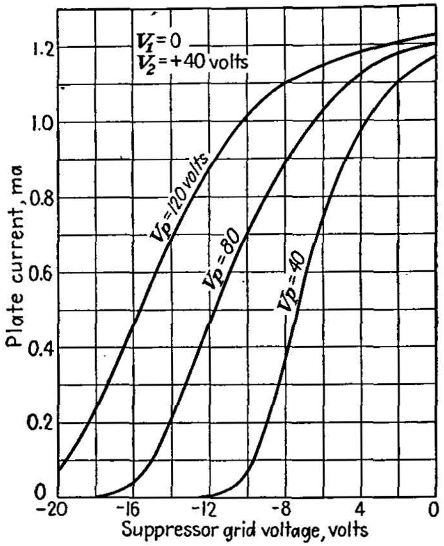


FIG. 11.4.—Plate-current–suppressor-grid characteristics of a type 6J7 pentode.

---

[1] GREEN, C. B., Suppressor Grid Modulation, *Bell Lab. Rec.*, vol. 17, pp. 41–44, October, 1938.

cathode to form before the suppressor grid owing to the space charge of the approaching electrons, which have been reduced to a very low velocity. The suppressor grid then has an action very similar to that of the control grid in front of an actual cathode, as in a triode.

An alternative representation of the effect of the suppressor grid is shown in Fig. 11.5. Here the control-grid action of the suppressor grid is evident at large negative values of suppressor-grid potential. Power

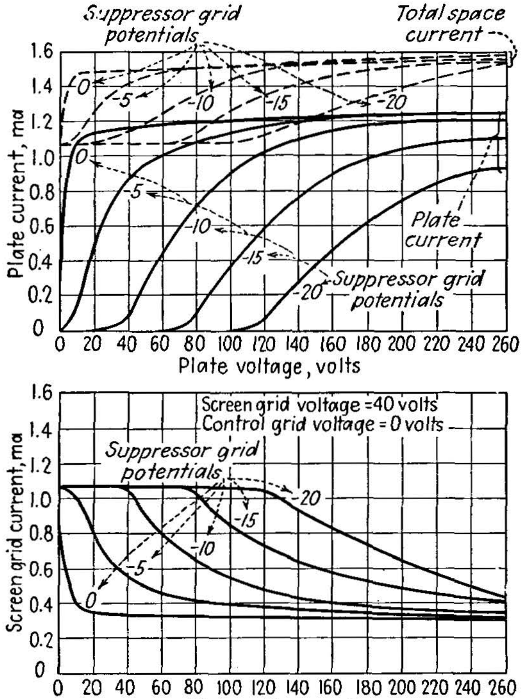


FIG. 11.5.—Effect of suppressor-grid voltage upon the plate and screen-grid-current characteristics of a pentode.

pentodes are frequently operated with slightly positive suppressor grids in order to get a sharper shoulder on the plate-current–plate-voltage characteristics. Some typical power-pentode characteristics are shown in Fig. 11.6. The use of a positive suppressor-grid potential is seen to give a considerable sharpness to the shoulder of the characteristics. The reason why this is necessary in the power tubes is twofold. (1) The current densities involved are greater, increasing the tendency for a virtual cathode to form in front of the suppressor grid and hence requiring a more positive value of suppressor-grid voltage to pass the major portion of the space current on to the plate. (2) The screen-grid pitch is made relatively large to reduce the total current and power taken by the screen. As this is done, the deflection imparted to the electrons passing through the screen grid is increased and hence a smaller fraction of them have enough plate-directed velocity to pass through the positive potential spaces between the suppressor-grid wires.

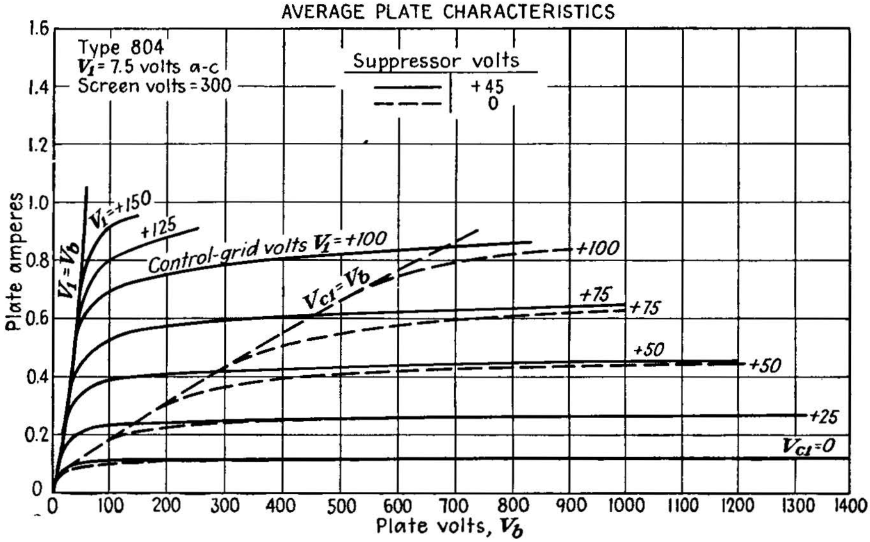


FIG. 11.6.—Plate-current–plate-voltage characteristics of a power pentode for zero and positive suppressor-grid potentials.

11.3. Current Division in Pentodes. The pentode is often operated with a suppressor grid at cathode potential and with screen grid and plate at the same positive potential. With this arrangement of electrode potentials the screen grid will always intercept an appreciable fraction of the space current. Some of this current is intercepted directly on the first passage of the electrons. Some of the screen-grid current consists of electrons which were so strongly deflected by the screen-grid wires that they did not have enough plate-directed velocity to pass between the suppressor-grid wires. *All* electrons that fail to penetrate the suppressor grid upon their first attempt may be expected to return to the screen grid.

The fraction of the total current transmitted to the plate is expected to be a function of the ratio of plate to screen-grid voltages. So also is the ratio of plate to screen-grid current. In Fig. 11.7 are shown curves of these current ratios as a function of the electrode-voltage ratio. Both current ratios are seen to vary rather slowly with the electrode-voltage ratio. The ratio of plate to space current in the vicinity of equal plate and screen-grid voltages varies something like the one-tenth power of the ratio of plate to screen-grid voltage. The ratio of plate to screen current varies approximately as the one-fifth power of the ratio of plate to screen-grid voltage. No simple theoretical analysis is available to give the current-division law directly in either case.

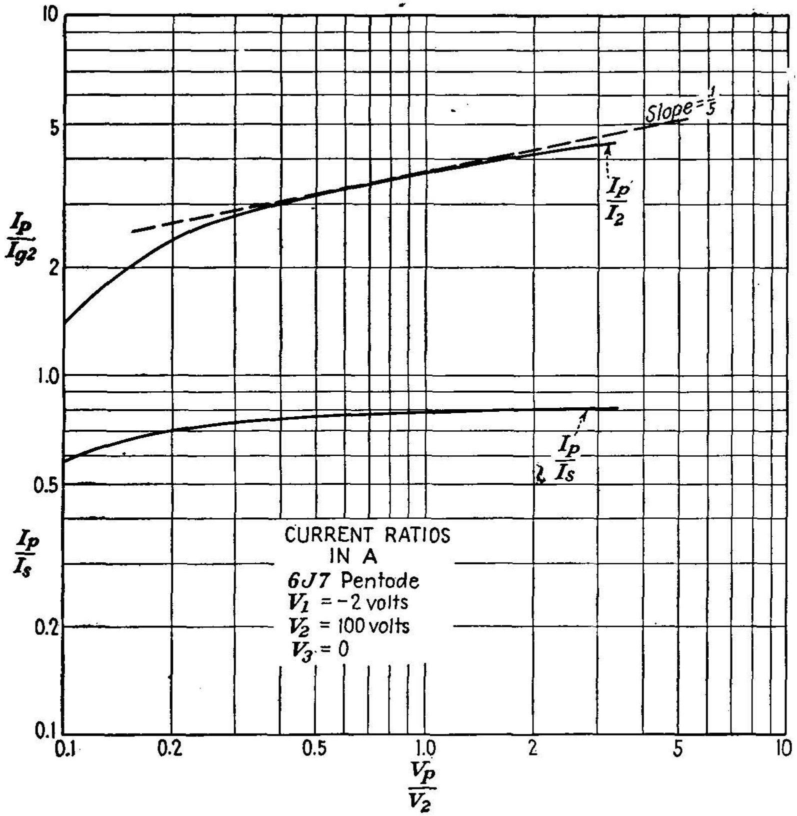


FIG. 11.7.—Pentode current ratios as a function of electrode-voltage ratios.

Just as it was possible to define a current-division factor for positive-grid triodes, so is it possible to define a current-divison factor for pentodes. Let

$$\frac{I_p}{I_2} = \mathfrak{D}\left(\frac{V_p}{V_2}\right)^m \tag{11.1}$$

Here $\mathfrak{D}$ is a current-division factor that measures the ratio of plate to screen-grid current when the plate and screen grid have the same voltage, other grids being presumed to draw no current. This pentode current-division factor may be determined quite closely from the considerations that applied to the positive-grid triode.

Figure 11.11 shows the general nature of the potential field and electron paths within a pentode. The control grid has a slight focusing action upon the electrons that pass between its wires. The screen grid intercepts a fraction of the current that passes through its plane and has a dispersing action on the rest. The suppressor grid will for the most part pass the electrons that approach it, with the exception of some electrons that fall into two groups. (1) The electrons that are aimed directly at a suppressor wire. These are naturally reflected and collected eventually by the screen grid. In general, all the electrons that are aimed at a suppressor-grid wire within half a radius of the center of the wire will be deflected back into the screen grid. This group comprises the great majority of the electrons initially passed by the screen grid that are returned to it. (2) The electrons that just barely miss a screen-grid wire and are so strongly deflected that they do not have enough plate-directed velocity to reach the suppressor-grid plane. This group of electrons is distinctly in the minority and may not even exist in some tubes if the screen-grid wires are large enough.

The same factors that determined the current division in a positive-grid triode also determine the fraction of the space current transmitted and intercepted by the screen grid of a pentode. The following identification of the elements of a triode gives approximately the conditions existing in a pentode:

Triode	Pentode
Cathode	Control grid
Control grid	Screen grid
Plate	Suppressor grid

When this correspondence of the electrodes is used, then Eq. (9.28) for current division in triodes may be applied directly provided that the mean suppressor-plane potential is used as the triode plate voltage and the equivalent triode cathode-plate distance is recognized as being somewhat larger than the pentode control-grid–screen-grid distance. When these considerations are applied, a transmission factor for the screen grid is determined. It is then necessary only to correct this for the additional electrons reflected from elastic collisions with the suppressor-grid wires.

For the condition of equal plate and screen-grid voltages prescribed for the pentode current-division factor, the mean suppressor-plane potential is a very small fraction of the plate and screen-grid potential. In terms of the equivalent-triode current division this means that interest is centered on the region of very small voltage ratios, far to the left on the true primary-current-division curve of Fig. 9.18. The value of the current ratio will not be a great deal less than the triode current-division factor because the current-ratio curve is concave upward. The true primary-division curve rather than the measured triode division law may be expected to apply in the case of the pentode because there is no modification of the electrode currents by secondary emission and because reflection of electrons from the plate of a triode at low potentials may be considerable, whereas in a pentode suppressor grid of the same mean potential the electrons will, for the most part, be able to penetrate in the spaces between the wires.

In order to apply the above ideas it is necessary to evaluate the mean suppressor-plane potential. Consider the configuration of electrodes shown in Fig. 11.8. In this somewhat idealized configuration the screen grid is replaced by a plane. Since the suppressor grid is usually of a coarse mesh, the field approximation of Eq. (7.72) is sufficiently accurate. If a linear potential term is included to account for the effect of the screen grid and plate, the expression for potential becomes

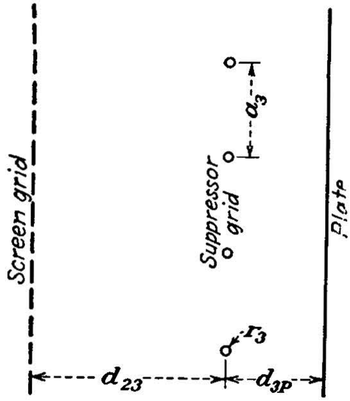


FIG. 11.8.—Electrode dimensions in the suppressor-grid region.

$$V(x,y) = -\frac{q}{4\pi\varepsilon_0}\left[\ln 2\left(\cosh\frac{2\pi x}{a} - \cos\frac{2\pi y}{a}\right)\right] + Bx + C \qquad (11.2)$$

where q is the charge per unit length of a single grid wire, x is measured in a direction perpendicular to the grid plane and has a zero value in the grid plane, and y is measured in the grid plane relative to a grid-wire center in a direction perpendicular to the wires themselves. Upon setting the potential equal to V_2 at the screen grid at which $x = -d_{23}$ and $y = 0$, then approximately

$$V_2 = -\frac{q}{2\varepsilon_0}\frac{d_{23}}{a_3} - Bd_{23} + C \qquad (11.3)$$

Upon setting the potential equal to zero at the suppressor at which $x = 0$ and $y = r_3$, then approximately

$$0 = -\frac{q}{2\pi\varepsilon_0}\ln\left[2\sin\left(\frac{\pi r_3}{a_3}\right)\right] + C \qquad (11.4)$$

Upon setting the plate potential also equal to V_2 at $x = d_{3p}$, $y = 0$, there results, approximately,

$$V_2 = -\frac{q}{2\varepsilon_0}\frac{d_{3p}}{a_3} + Bd_{3p} + C \tag{11.5}$$

The same approximations apply above as do in the case of the low-mu triode. The expressions are valid only for screening fractions less than one-tenth and interelectrode spacings greater than the distance between suppressor-grid wires.

The three equations above may be solved for the three unknowns q, B, and C and these values substituted in Eq. (11.2). When this is done and the general expression for potential resulting is restricted to the point $x = 0$, $y = \frac{a_3}{2}$, there is obtained an expression for the maximum potential between the suppressor-grid wires,

$$V_{3\,\text{max}} = \frac{V_2 \ln\left[\sin\left(\frac{\pi r_3}{a_3}\right)\right]}{\ln 2\left[\sin\left(\frac{\pi r_3}{a_3}\right)\right] - \frac{2\pi}{a_3}\frac{d_{23}d_{3p}}{d_{2p}}} \tag{11.6}$$

The *mean* value of the suppressor-grid-plane potential is just half this.

$$V_{3\,\text{av}} = \frac{V_2 \ln\left[\sin\left(\frac{\pi r_3}{a_3}\right)\right]}{2\left\{\ln\left[2\sin\left(\frac{\pi r_3}{a_3}\right)\right] - \frac{2\pi}{a_3}\frac{d_{23}d_{3p}}{d_{2p}}\right\}} \tag{11.7}$$

When this mean value of the suppressor-plane potential is known, Eq. (9.28*a*), which is repeated here, can be applied with the electrode correspondence previously mentioned to obtain the ratio of the current transmitted by the screen grid to the current intercepted on the initial passage.

$$\frac{I_p}{I_g} = \frac{a}{2\left\{\frac{a\mu\left[(d_{cg} + d_{gp}) - d_{cg}\frac{V_p}{V_g}\right]D}{2\pi d_{gp}\left(\frac{V_p}{V_g} + \mu\right)} + r_g\right\}} - 1 \tag{9.28a}$$

Upon substituting equivalent factors to suit the pentode problem this expression becomes

$$T_2 = \frac{a_2}{2\left\{\dfrac{a_2\mu_{23}\left[(d_{12}+d_{23}) - d_{12}\dfrac{V_{3\,\mathrm{av}}}{V_2}\right]D}{2\pi d_{23}\left(\dfrac{V_{3\,\mathrm{av}}}{V_2}+\mu_{23}\right)} + r_2\right\}} - 1 \qquad (11.8)$$

where $D = \dfrac{r_2}{2d_{12}} \ln \dfrac{4\epsilon d_{12}}{r_2}$

T_2 = ratio of current transmitted by screen grid upon initial passage of electrons to current intercepted

μ_{23} = geometrical amplification factor of control grid relative to screen grid

It is necessary only to modify the factor T_2 by the transmission factor of the suppressor grid to obtain the screen current-division factor $\mathfrak{D}$ The transmission through the suppressor plane will generally be so large that it will affect the over-all result by only a small fraction. The transmission of the suppressor plane is given approximately by

$$T_3 = \frac{a_3}{r_3} - 1 \qquad (11.9)$$

since membrane-model studies show that only those electrons aimed at the center *half* of a suppressor-grid wire will strike it. The over-all current ratio is then given approximately by

$$\frac{1}{\mathfrak{D}} = \frac{1}{T_2} + \frac{1}{T_3} \qquad (11.10)$$

Example: Consider the 6J7 pentode, which has the following dimensions:

$d_{c1} = 8.65$ mils $\quad a_1 = 21.1$ mils $\quad r_1 = 1.1$ mils
$d_{12} = 56.8$ mils $\quad a_2 = 15.9$ mils $\quad r_2 = 1.25$ mils
$d_{23} = 109.5$ mils $\quad a_3 = 50$ mils $\quad r_3 \doteq 2.25$ mils
$d_{3p} = 110$ mils

$$\frac{V_{3\,\mathrm{av}}}{V_2} = \frac{\ln\left[\sin\left(\dfrac{\pi \times 2.25}{50}\right)\right]}{2\left\{\ln\left[2\sin\left(\dfrac{\pi \times 2.25}{50}\right)\right] - \dfrac{2\pi}{50}\dfrac{109.5 \times 110}{219.5}\right\}}$$

$$= 0.12 \qquad \text{from Eq. (11.7)}$$

$$\mu_{23} = 56 \qquad \text{from Fig. 7.11}$$

for

$$\frac{d_{23}}{a_2} = \frac{109.5}{15.9} = 6.89$$

and

$$S = \frac{2r_2}{a_2} = \frac{2.5}{15.9} = 0.1572$$

$$T_2 = \frac{15.9}{2\left\{\dfrac{15.9 \times 56[(56.8 + 109.5) - 56.8 \times 0.12]0.0683}{2\pi \times 109.5(0.12 + 56)} + 1.25\right\}} - 1$$

$$= 4.3 \qquad \text{from Eq. (11.8)}$$

Note that this value of current transmission is only slightly less than the δ factor of 4.63 calculated from Eq. (9.25), considering the screen grid of the pentode as the control grid of a triode.

$$T_3 = \frac{50}{2.25} - 1 = 21.2 \qquad \text{from Eq. (11.9)}$$

$$\mathfrak{D} = \frac{1}{\dfrac{1}{4.3} + \dfrac{1}{21.2}} = 3.57 \qquad \text{from Eq. (11.10)}$$

This calculated value of $\mathfrak{D}$ agrees well with a measured value of 3.65. The agreement is, in fact, better than there is any reason to expect in view of the fact that the 6J7 does not have a plane-electrode structure at all but has a circular cathode and suppressor grid, elliptical control and screen grids, and a plane plate.

11.4. Amplification Factor of a Pentode. The design of a pentode presents a rather complex problem. Relatively little has been published on this important subject. Most of the design equations exist in the private notebooks of a few workers in the field and are largely empirical modifications of simple theoretical relations. In this and subsequent sections there is given a sketch of the factors involved in the determination of pentode tube constants. The results that are given can serve only as a rough guide to the fundamental relations and should not be taken as anything more than approximate design equations.

The amplification-factor formulas of a pentode may be expected to be considerably more complicated than those of a triode for two reasons. (1) There are three grids instead of one. (2) The division of current between the various electrodes is a function of the relative electrode potentials. If the amplification factor is calculated from electrostatic considerations as was done for the triode, there results an expression that gives the relative influence of the plate and control grid in keeping the *space current* constant. This is not the true amplification factor but what will be referred to as the "electrostatic amplification factor" since it gives the relative influence of the plate and control grid in controlling the off-cathode gradient of potential in a cold tube (or space current in a hot tube). The true amplification factor is a modification of this value that gives the relative influence of plate and control grid in controlling the *plate current*. The subject will be treated by first studying the field in a pentode, deducing the electrostatic amplification factor from it, and then modifying this to obtain the true amplification factor.

Electrostatic Field of a Pentode. The method of conformal transformations is not readily applied to tubes having several grids of different pitches. It is, however, possible to construct the field from the expression for the potential due to a single row of grid wires given in Eq. (7.72),

$$V(x,y) = -\frac{q}{4\pi\varepsilon_0}\ln\left[2\left(\cosh\frac{2\pi x}{a} - \cos\frac{2\pi y}{a}\right)\right] + C \qquad (7.72)$$

where the wires are spaced a distance a apart upon the y axis and q is the charge per unit length of wire. If three terms like the function in Eq. (7.72) are combined properly with a linearly varying component of potential, the resultant expression is a satisfactory representation of the field of a pentode.

First consider some of the properties of Eq. (7.72). In the first place the constant has the value

$$C = V_g + \frac{q}{2\pi\varepsilon_0}\ln\left(\frac{2\pi r_g}{a}\right) \qquad (11.11)$$

where V_g is the potential of the isolated grid and r_g is the radius of the grid wires. Near the grid wires the equipotential contours are circles concentric with the grid wire. In this vicinity the potential is given approximately by

$$V(x,y) = V_g - \frac{q}{2\pi\varepsilon_0}\ln\left(\frac{r}{r_g}\right) \qquad (11.12)$$

At a considerable distance from the grid the equipotential contours are straight lines parallel to the grid-wire plane whose approximate potential is given by

$$V(x) = V_g + \frac{q}{2\pi\varepsilon_0}\ln\left(\frac{2\pi r_g}{a}\right) \mp \left(\frac{qx}{2a\varepsilon_0}\right) \qquad (11.13)$$

where the upper sign is associated with potentials to the right of the grid plane and the lower with potentials to the left. It is seen that the potential varies linearly with distance from the grid-wire plane, just as it would from a plane with a surface-charge density of $\frac{q}{2a}$. The second term above gives the difference between the actual grid-wire potential and the *equivalent* potential of the grid plane, found by extending the straight-line portions of the potential profiles back to an intersection, as shown in Fig. 11.9. The depth of the fillet about the grid wires is given by letting

$x = 0$ and $y = \frac{a}{2}$ in Eq. (7.72). This substitution gives for the maximum deviation from grid-wire potential in the grid plane

$$V_g - V\left(0, \frac{a}{2}\right) = -\frac{q}{2\pi\varepsilon_0} \ln\left(\frac{\pi r_g}{a}\right) \tag{11.14}$$

It is therefore always true that the potential difference between the grid wire and the equivalent potential of the grid plane is 0.693 (= ln 2) of the maximum difference of potential encountered in the grid plane. Also shown in Fig. 11.9 is the gradient of potential about a row of grid wires.

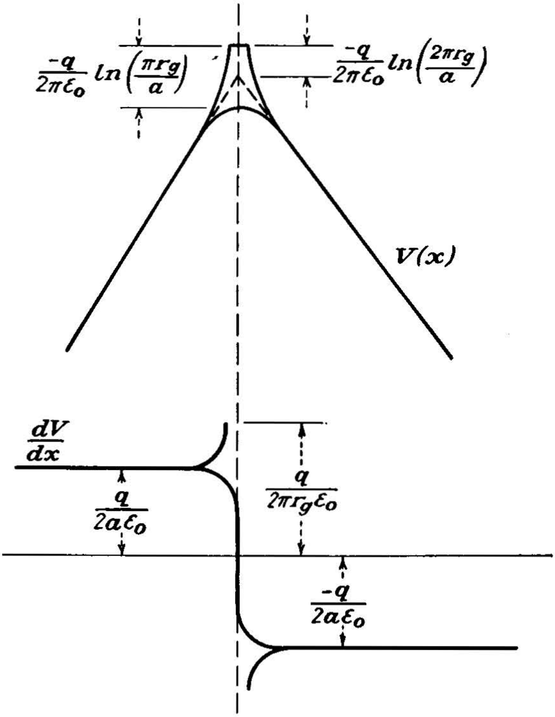


FIG. 11.9.—Potential profiles of a single row of grid wires.

Provided that the distance between electrodes is greater than the distance between grid wires, which is a good approximation with the exception that the suppressor-grid–plate distance is often less than the suppressor-grid pitch, the field of a pentode is given by the sum of three expressions like the right-hand side of Eq. (7.72) plus a linear component of field plus a constant.

$$V(x,y) = \sum_{n=1}^{3} - \frac{q_n}{2\pi\varepsilon_0} \ln \left\{ 2 \left[\cosh 2\pi \left(\frac{x - d_{cn}}{a_n} \right) - \cos 2\pi \left(\frac{y - b_n}{a_n} \right) \right] \right\} - \frac{q_c x}{\varepsilon_0} + C \quad (11.15)$$

in which the subscript n assumes the values 1,2,3 to correspond to each of the three grids, d_{cn} is the distance from the cathode to the nth grid, b_n gives the relative location of the grid wires along a reference line normal to the cathode, a_n is the pitch of the nth grid, q_c is the cathode charge per unit area, and q_n is the charge per unit length of grid wire on the nth grid.

Upon making the usual approximations for large values of x and small values of y, it is possible to write the equations relating the potential at each electrode to the electrode charges. These equations are

$$V_c = 0 = -\frac{q_1 d_{c1}}{a_1 \varepsilon_0} - \frac{q_2 d_{c2}}{a_2 \varepsilon_0} - \frac{q_3 d_{c3}}{a_3 \varepsilon_0} + 0 + C \quad (11.16)$$

$$V_1 = -\frac{q_1}{2\pi\varepsilon_0} \ln \left(2 \sin \frac{\pi r_1}{a_1} \right) - \frac{q_2 d_{12}}{a_2 \varepsilon_0} - \frac{q_3 d_{13}}{a_3 \varepsilon_0} - \frac{q_c d_{c1}}{\varepsilon_0} + C \quad (11.17)$$

$$V_2 = -\frac{q_1 d_{12}}{a_1 \varepsilon_0} - \frac{q_2}{2\pi\varepsilon_0} \ln \left(2 \sin \frac{\pi r_2}{a_2} \right) - \frac{q_3 d_{23}}{a_3 \varepsilon_0} - \frac{q_c d_{c2}}{\varepsilon_0} + C \quad (11.18)$$

$$V_3 = -\frac{q_1 d_{13}}{a_1 \varepsilon_0} - \frac{q_2 d_{23}}{a_2 \varepsilon_0} - \frac{q_3}{2\pi\varepsilon_0} \ln \left(2 \sin \frac{\pi r_3}{a_3} \right) - \frac{q_c d_{c3}}{\varepsilon_0} + C \quad (11.19)$$

$$V_p = -\frac{q_1 d_{1p}}{a_1 \varepsilon_0} - \frac{q_2 d_{2p}}{a_2 \varepsilon_0} - \frac{q_3 d_{3p}}{a_3 \varepsilon_0} - \frac{q_c d_{cp}}{\varepsilon_0} + C \quad (11.20)$$

The above expressions may be solved for the charges in terms of the electrode potentials and then applied to Eq. (11.15) for the potential field. This process is somewhat involved, however; for ordinary purposes a simpler procedure that yields results accurate enough for most purposes is recommended. This simplified procedure consists in sketching the potential profiles and then correcting the originally assumed values. Ordinarily only one correction is necessary.

The simplified procedure for determining potential fields in pentodes is applied as follows: Ordinarily a complete plot of the potential field is not required, and potential profiles are sufficient. The procedure first calls for a sketch of the potential profiles within the tube. For convenience the profiles will be drawn through a wire of each grid and midway between the grid wires and the segments of such profiles joined. Actually, there may be no actual straight line in the tube that goes through a wire of each grid, but this makes no difference. Such a sketch is shown in Fig. 11.10. The specified electrode potentials are marked, fillets of reasonable size are attached, and the fillets are then joined by straight lines. The next step in the procedure is to draw a curve of the gradient of potential between electrodes as taken from the slope of the straight-line portions of the profiles joining the electrodes. Such a curve of gradients is also shown in Fig. 11.10. If the gradients are taken from the profiles in units of volts per meter, then the gradients will have the values indicated on the figure in terms of the charges. From the four values of the gradient between the electrodes, the four electrode charges in units of coulombs per square meter can be calculated. When the electrode charges are known, then the factor $-\frac{q_n}{2\pi\varepsilon_0}\ln\left(\frac{2\pi r_n}{a_n}\right)$ applied to each grid, as shown in Fig. 11.9, to see how good the original guess on the size of the fillet of potential around the wires was. Generally, the original guesses are not exact, and some values of electrode potentials different from those desired are found to fit the straight-line portions of the potential profiles. Correcting these values is a simple matter, and usually the first correction will be close enough to the exact one for ordinary purposes.

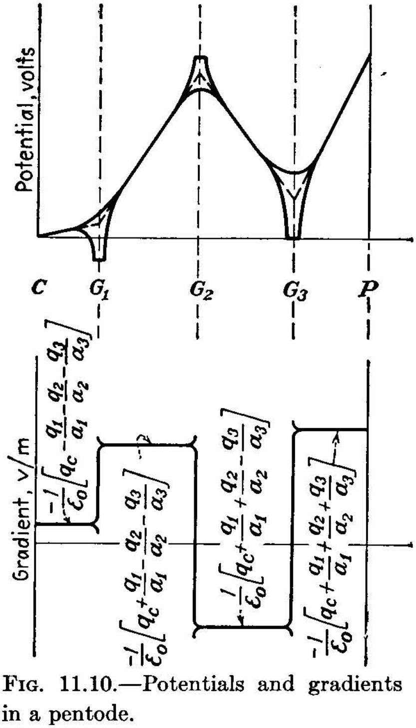


FIG. 11.10.—Potentials and gradients in a pentode.

A plot of the complete field within a pentode is shown in Fig. 11.11. The figure shows some typical electrode dimensions and electrode potentials and gradients in a pentode.

Electrostatic Amplification Factor of a Pentode.[1] From work on the triode it is possible to find an expression giving the potential of a grid in terms of its charge and the charge to the left of it. By combining such expressions for all the grids of a pentode there is obtained an expression for the cathode charge in terms of the electrode potentials. From this the relative effectiveness of the various electrodes in controlling the off-

---

[1] See also Dow, W. G., Equivalent Electrostatic Circuits for Vacuum Tubes, *Proc. I.R.E.*, vol. 28, pp. 548–556, December, 1940.

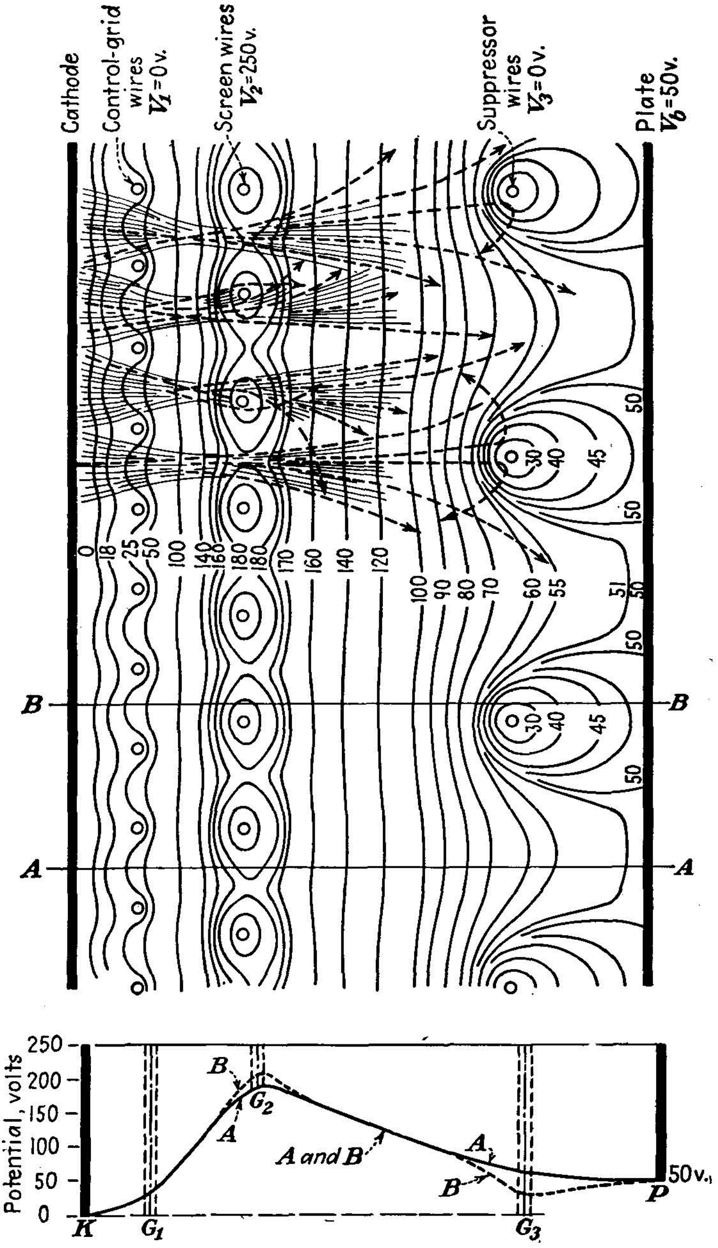


FIG. 11.11.—Electrostatic field and electron paths in a pentode. (*Schade.*)

cathode gradient can be determined. This gives the electrostatic amplification factor.

If the component logarithmic-transformation equations are applied to Eq. (7.31), there results

$$V_g = \frac{q_c}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_g}{a} - \frac{q_c d_c}{a\varepsilon_0} + \frac{q_p}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_g}{a} - \frac{q_p}{2\pi\varepsilon_0} \ln \cosh \frac{2\pi r_g}{a} \tag{11.21}$$

The last term in the above equation rarely exceeds 1 per cent of the second last term and so will be dropped. The remaining terms can be arranged to give

$$V_1 = -\frac{q_L d_L}{a_1\varepsilon_0} - \frac{q_1}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_1}{a_1} \tag{11.22}$$

in which q_L is the total charge per grid-wire section to the left of the grid, d_L is the distance from the grid to the next electrode to the left of the grid in question (the cathode in the case of the triode), and the subscript 1 means that the particular symbol applies only to grid 1.

The first term of Eq. (11.22) establishes the average level of the grid-plane potential since it is the gradient of the straight-line portion of the potential profile to the left of the grid. The second term gives the rest of the potential necessary to make up the actual grid potential. There are no restrictions on Eq. (11.22) that confine it to a triode; it can just as well be applied to any grid of a tube as long as the symbols are given the proper interpretation. Since in summing potential expressions like Eq. (11.22) it is necessary to take account of the fact that the various grids may have different pitches, the charge per unit area instead of the charge per unit length of grid wire will be used. The charge per unit area is given by

$$Q_n = \frac{q_n}{a_n} \qquad Q_c = q_c \tag{11.23}$$

where a_n is the grid-wire spacing.

To apply the summation procedure indicated above to a pentode, the three grids will be referred to by numbered subscripts and the distance between electrodes will be given by the symbol d with a double subscript, corresponding to the electrodes involved. The following equations are thus obtained:

$$V_1 = -\frac{Q_c d_{c1}}{\varepsilon_0} - \frac{Q_1 a_1}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_1}{a_1} \tag{11.24}$$

$$V_2 - V_1 = -\frac{(Q_c + Q_1) d_{12}}{\varepsilon_0} - \frac{Q_2 a_2}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_2}{a_2} \tag{11.25}$$

$$V_3 - V_2 = -\frac{(Q_c + Q_1 + Q_2)d_{23}}{\varepsilon_0} - \frac{Q_3 a_3}{2\pi\varepsilon_0} \ln \sinh \frac{2\pi r_3}{a_3} \tag{11.26}$$

$$V_p - V_3 = -\frac{(Q_c + Q_1 + Q_2 + Q_3)d_{3p}}{\varepsilon_0} \tag{11.27}$$

It will be recognized that the first term of each of the above equations establishes the average potential of an electrode in terms of the charge and distance to its left. The second term takes account of the deformation introduced by the presence of the grid. In the case of Eq. (11.27) there is no second term because the potential of the plate is constant. The above set of equations gives the electrode potentials in terms of the charges. The system can, of course, be solved for the charges in terms of potentials by Cramer's rule.

The determinant of the coefficients is

$$\Delta = \begin{vmatrix} d_{c1} & \left(\frac{a_1}{2\pi} \ln \sinh \frac{2\pi r_1}{a_1}\right) & 0 & 0 \\ d_{12} & d_{12} & \left(\frac{a_2}{2\pi} \ln \sinh \frac{2\pi r_2}{a_2}\right) & 0 \\ d_{23} & d_{23} & d_{23} & \left(\frac{a_3}{2\pi} \ln \sinh \frac{2\pi r_3}{a_3}\right) \\ d_{3p} & d_{3p} & d_{3p} & d_{3p} \end{vmatrix} \tag{11.28}$$

This can be simplified by the introduction of the symbols

$$K = \frac{d_{c1}d_{12}d_{23}d_{3p}}{\varepsilon_0{}^4} \tag{11.29}$$

and

$$B_n = \frac{a_n}{2\pi d_{(n-1)n}} \ln \sinh \frac{2\pi r_n}{a_n} \tag{11.30}$$

when $n = 1$, $n - 1 = c$.

After substituting the above, the determinant assumes the form

$$\Delta = K \begin{vmatrix} 1 & B_1 & 0 & 0 \\ 1 & 1 & B_2 & 0 \\ 1 & 1 & 1 & B_3 \\ 1 & 1 & 1 & 1 \end{vmatrix} \tag{11.31}$$

The value of this determinant has the simple form

$$\Delta = K(1 - B_1)(1 - B_2)(1 - B_3) \tag{11.32}$$

The cathode charge is given by

$$Q_c = V_1\left[\frac{(1 - B_2)(1 - B_3)}{d_{c1}} + \frac{B_1(1 - B_3)}{d_{12}}\right] + V_2\left[-\frac{B_1(1 - B_3)}{d_{12}} - \frac{B_1B_2}{d_{23}}\right] + V_3\left[\frac{B_1B_2}{d_{23}} + \frac{B_1B_2B_3}{d_{3p}}\right] + V_p\left[-\frac{B_1B_2B_3}{d_{3p}}\right] \tag{11.33}$$

Reference to Eq. (7.35) shows that the electrostatic amplification factor is given by the ratio of the coefficients of the first and last terms.

$$\mu_{1p} = \frac{\dfrac{(1 - B_2)(1 - B_3)}{d_{c1}} + \dfrac{B_1(1 - B_3)}{d_{12}}}{-\dfrac{B_1B_2B_3}{d_{3p}}} \tag{11.34}$$

Substitution of the values of the various B's shows that the above expression for amplification factor is independent of d_{c1}. In general, the electrostatic amplification factor of the pentode will have a value approximately equal to that given by the product of three triode mu's calculated by considering the plate, the suppressor grid, and the screen grid as plates of a triode and the next two electrodes in order toward the cathode as grid and cathode.[1]

Also available from Eq. (11.33) is the electrostatic amplification factor giving the relative effectiveness of the control grid and screen grid in controlling space current. It is

$$\mu_{12} = \frac{\dfrac{(1 - B_2)(1 - B_3)}{d_{c1}} + \dfrac{B_1(1 - B_3)}{d_{12}}}{\dfrac{-B_1(1 - B_3)}{d_{12}} - \dfrac{B_1B_2}{d_{23}}} \tag{11.35}$$

The above expressions for amplification factor have some small inherent inaccuracies due to the fact that the filleting of potentials about some of the grids has been neglected. The inaccuracy is probably not more than a few per cent. More accurate expressions for electrostatic amplification factor may be derived from Eqs. (11.16) to (11.20), but these are so cumbersome as to be almost totally useless.

True Amplification Factor of a Pentode. The true amplification factor of a pentode must take into account not only the electrostatic action

---

[1] THOMPSON, B. J., Space Current Flow in Vacuum Tube Structures, *Proc. I.R.E.*, vol. 31, pp. 485–491, September, 1943.

within the tube but also the division of current between the electrodes. Let the true amplification factor of the pentode be defined by

$$\mu_p = -\left(\frac{dV_p}{dV_1}\right)_{I_p\ \text{const}} = -\left(\frac{\dfrac{\partial I_p}{\partial V_1}}{\dfrac{\partial I_p}{\partial V_p}}\right) \tag{11.36}$$

In deriving the true amplification factor, use will be made of the following relation,

$$I_s = G\left(V_1 + \frac{V_2}{\mu_{12}} + \frac{V_3}{\mu_{13}} + \frac{V_p}{\mu_{1p}}\right)^n \tag{11.37}$$

where n is a constant of approximate value 1.5, the μ's are electrostatic amplification factors measuring the relative effectiveness of the electrode in question and the control grid in controlling the space current, G is perveance of value $\dfrac{2.335 \times 10^{-6}}{d^2}$ amperes per unit area per volt^n, and d is the equivalent control-grid–cathode spacing as calculated from Eq. (8.45) but with the screen grid considered as the triode plate.

Let the ratio of plate to screen-grid current be given by

$$\frac{I_p}{I_2} = g\left(\frac{V_p}{V_2}\right) \tag{11.38}$$

and let the functional relation be indicated subsequently by the symbol g.

The ratio of plate to space current is given by

$$\frac{I_p}{I_s} = \frac{I_p}{I_2 + I_p} = \frac{\dfrac{I_p}{I_2}}{1 + \dfrac{I_p}{I_2}} = \frac{g}{1+g} \tag{11.39}$$

and hence

$$I_p = I_s\frac{I_p}{I_s} = G(V_{\text{eq}})^n\frac{g}{1+g} \tag{11.40}$$

where V_{eq} is the equivalent voltage $V_1 + \dfrac{V_2}{\mu_{12}} + \dfrac{V_3}{\mu_{13}} + \dfrac{V_p}{\mu_{1p}}$.

The partial derivatives that enter into the determination of the true amplification factor can now be evaluated.

$$\frac{\partial I_p}{\partial V_1} = nG(V_{\text{eq}})^{n-1}\frac{g}{1+g} \tag{11.41}$$

$$\frac{\partial I_p}{\partial V_p} = \frac{nG}{\mu_{1p}}(V_{\text{eq}})^{n-1}\frac{g}{1+g} + G(V_{\text{eq}})^n\frac{\dfrac{dg}{dV_p}}{(1+g)^2} \tag{11.42}$$

Upon combining these last two expressions, the resulting expression for the true amplification factor is

$$\mu_p = \frac{\mu_{1p}}{1 + \dfrac{\mu_{1p} V_{eq} \dfrac{dg}{dV_p}}{ng(1 + g)}} \tag{11.43}$$

If the ratio of plate to screen-grid current is assumed to have the form previously given

$$g = \frac{I_p}{I_2} = \mathfrak{D} \left(\frac{V_p}{V_2} \right)^m \tag{11.1}$$

then

$$\mu_p = \frac{\mu_{1p}}{1 + \dfrac{\mu_{1p} m V_{eq}}{n V_p \left[1 + \mathfrak{D} \left(\dfrac{V_p}{V_2} \right)^m \right]}} \tag{11.44}$$

From this it is seen that the true amplification factor is less than the electrostatic amplification factor by a considerable factor.

11.5. Transconductance of a Pentode. The transconductance of a pentode is readily obtained. Let

$$G_m = G_{1p} = \frac{\partial I_p}{\partial V_1} \tag{11.45}$$

but

$$I_p = I_s \frac{I_p}{I_s} \tag{11.46}$$

and so

$$G_{1p} = \frac{\partial I_s}{\partial V_1} \frac{I_p}{I_s} \tag{11.47}$$

since the ratio of plate to space current is independent of control-grid voltage. The quantity $\dfrac{\partial I_s}{\partial V_1}$ may be designated as G_{1s} and may be obtained from the triode mutual-conductance formula [Eq. (8.47)] by considering the screen grid of the pentode as the plate of a triode. Equation (11.47) states that the control-grid–plate transconductance of a pentode is equal to the triode mutual conductance of the first three electrodes, reduced by the ratio of the plate current to the screen current.

11.6. Plate Resistance of a Pentode. The plate resistance of a pentode is the reciprocal of the plate conductance defined by

$$\frac{1}{R_p} = G_p = \frac{\partial I_p}{\partial V_p} \tag{11.48}$$

Utilizing Eq. (11.40),

$$G_p = \frac{\partial I_s}{\partial V_p}\,\frac{g}{1+g} + \frac{I_s\,\dfrac{dg}{dV_p}}{(1+g)^2} \tag{11.49}$$

The first term in this last expression results from the variation of the emitted current with plate voltage. The second term results from the change in plate current occasioned by the variation in current division with plate voltage. When the current-division function of the pentode has the form of Eq. (11.1), the above expression reduces to

$$G_p = \frac{G_{1s}}{\mu_{1p}}\,\frac{I_p}{I_s} + \frac{I_s\mathfrak{D}m\left(\dfrac{V_p}{V_2}\right)^{m-1}}{V_2\left[1+\left(\dfrac{V_p}{V_2}\right)^m\right]^2} \tag{11.50}$$

11.7. Design Considerations. In a pentode the ordinary constants are readily made to assume satisfactory values. Prime interest is centered in the transconductance. Here the same considerations apply as in the triode, and no greater difficulty is encountered. The amplification factor and plate resistance are naturally high and require no particular attention. Thus interest is focused upon some of the other characteristics of the tube that affect its operation. These other characteristics are

1. Suppression of secondary electrons.
2. Sharpness of the shoulder of the plate-current–plate-voltage characteristic.
3. Plate-current to screen-grid-current ratio.

The above factors are controlled by some factors that have not appeared before in this study of vacuum-tube design. These are

1. Shape of the potential field before the plate.
2. Electron deflection by the grids.

In previous considerations of tube characteristics the principal concern has been with the potential field and with the space-charge flow. *In the pentode, in addition, the electron paths are critical.*

The suppression of secondary electrons from the plate is not a difficult problem, though some attention must be paid to the electrode dimensions. The critical factors are the pitch of the suppressor grid and its distance from the plate. In Fig. 11.12 is shown the effect of different suppressor-grid pitches upon the retarding potential offered to secondary electrons created at the plate as a function of plate voltage as calculated from Eq.

(11.6). The minimum potential referred to in the figure is the minimum of potential on a line normal to the plate passing midway between grid wires. As the suppressor-grid pitch is decreased, this minimum potential is decreased and the retarding potential offered to the secondaries thus increased. The effects are linear with the various potentials involved, provided that space-charge effects do not distort the potentials appreciably. In Fig. 11.13 is shown the effect of putting the suppressor grid in different positions between the screen grid and plate, as calculated from Eq. (11.6). This changes both the magnitude and the rate of change of the retarding potential. For suppressing plate secondaries it is desirable to have a fixed retarding potential. This cannot be realized, and therefore the arrangement of electrodes that gives the most retarding potential at *low* plate potentials is desired. This means that a small suppressor-grid–plate distance is indicated. So also is a small suppressor-grid pitch. Some other considerations, as we shall see, limit the degree to which the suppressor pitch can be reduced, but both the above factors should be considered to ensure secondary suppression.[1]

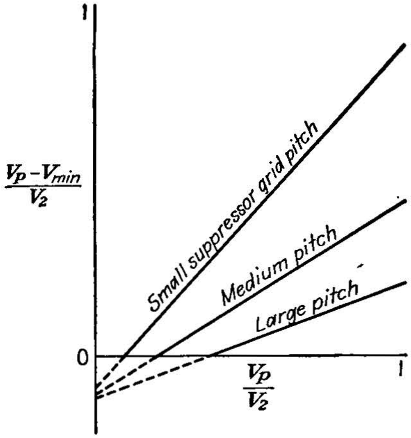


FIG. 11.12.—Retarding increment of potential before the plate of a pentode as a function of plate voltage for various suppressor-grid pitches.

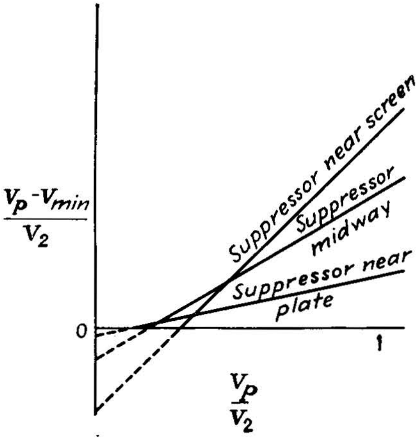


FIG. 11.13.—Retarding increment of potential before the plate of a pentode as a function of plate voltage for various suppressor-grid locations.

The other new factor in pentode design, *viz.*, the deflection of electrons by the grids, is probably more important than the secondary suppression, since the latter is usually achieved without too much difficulty. The electron deflection by the grids will affect strongly both the sharpness of the plate-current–plate-voltage characteristic and the ratio of plate to

---

[1] JONKER, J. H. L., Pentode and Tetrode Output Valves, Parts I, II, *Wireless Engr.*, vol. 16, pp. 274–286, 344–349, July, 1939.

screen-grid current. The more the electrons are deflected by the grids, the fewer will reach the plate at ordinary plate voltages and the less rapid will be the increase of plate current with plate voltage.

An exact study of electron deflection by the grid wires is rather difficult, but an excellent approximate analysis of the deflections can be made by using the fact that the potential field between the grid wires acts like a cylindrical lens and may have the effect of either focusing or dispersing the electrons which pass between them. Thus in Fig. 9.10*a* it is seen that the control grid gives the focusing action of a convergent lens. This lens, however, has some very pronounced aberrations. The focal length for parts of the lens near the grid wires is less than for the center of the space. In the language of optics, the lens has a positive spherical aberration. In Fig. 10.6 the focusing action of the control grid may again be seen. It is also seen that the screen grid exhibits the characteristics of a divergent lens. In Figs. 11.14 and 11.15 it is seen that the suppressor grid has a convergent-lens action. The nature of the lens caused by the potential field between grid wires depends upon the gradient of potential on the two sides of the grid and is, to first order, independent of such things as the grid radius and pitch.

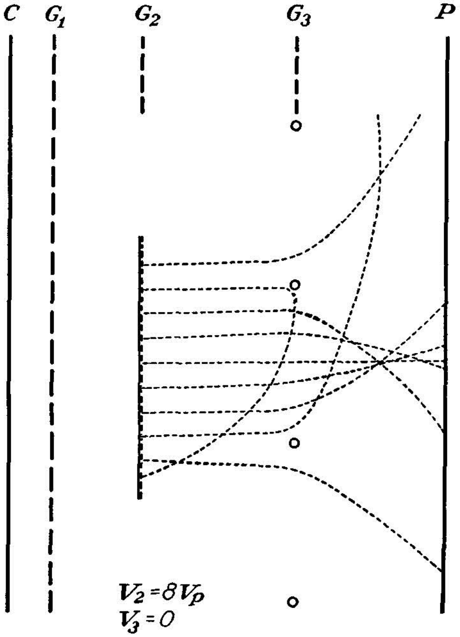


FIG. 11.14.—Scattering action of a small-pitch suppressor grid far from the plate of a pentode.

The focal length of the lenses formed by the grid wires will be shown in the chapter on Electrostatic Electron Optics to be given by

$$f = \frac{2V_n}{\left(\frac{dV}{dx}\right)_R - \left(\frac{dV}{dx}\right)_L} \qquad (11.51)$$

where V_n is the potential midway between the wires of grid n, $\left(\frac{dV}{dx}\right)_R$ is the gradient of potential to the right of the grid, and $\left(\frac{dV}{dx}\right)_L$ is the gradient of potential profile to the left of the grid plane. It is seen that when the gradient of potential increases upon passing through the grid, f is positive and the lens is convergent. When the opposite is true, the lens is divergent. The focal distance as given by the above formula will be modified somewhat in actual cases when the gradient of potential is not zero on the side where the focus occurs, for the electron trajectories

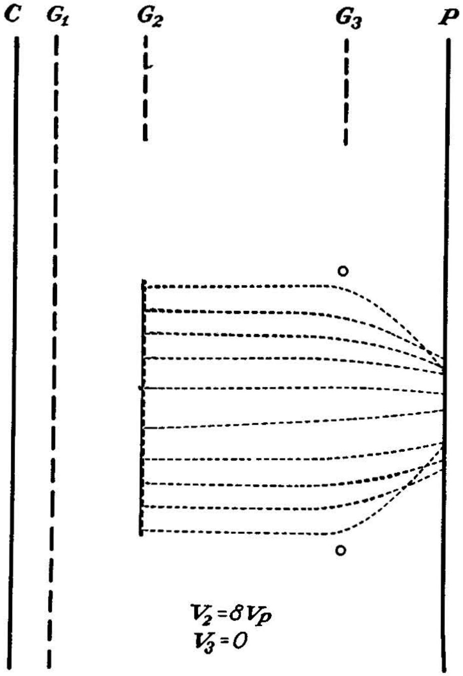


FIG. 11.15.—Focusing action of a large-pitch suppressor grid close to the plate of a pentode.

will tend to be sections of parabolic curves instead of straight lines. The formula does, however, give a fairly exact indication of the principal effects and the correct value of sidewise components of velocity.

The nature of the focusing of a set of grid wires is shown schematically in Fig. 11.16. An electron passing midway between grid wires suffers no sidewise deflection whatever. As the distance from the midplane increases, the electrons receive more and more deflection, by a linear law, so that they all cross over at the same point. This holds true almost exactly for the center half of the space between the grid wires. The initial offset from the midplane, the focal length, and the tangent of the angle of deflection are related by

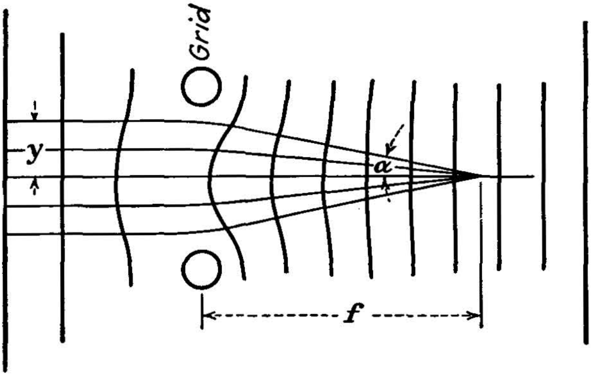


FIG. 11.16.—Focal action of a grid.

$$\tan \alpha = \frac{y_0}{f} \tag{11.52}$$

where y_0 is the offset from the midplane along the line of the grid wires of the electron's initial position, f is the focal length, and α is the angle of deflection. In terms of velocity components,

$$\frac{y_0}{f} = \frac{v_y}{v_x} \cong \frac{v_y}{v} \tag{11.53}$$

where v_x and v_y are the forward and sidewise components of velocity possessed by the electron shortly after passing through the grid plane and v is the total velocity.

The experimental agreement between this formula and the actual behavior is quite good, as shown in Fig. 11.17. Here is shown the actual deflection as measured on an elastic membrane (curve a) and the deflection calculated on the assumption of a constant grid-plane potential equal to the mean potential of the grid plane (curve b). The measured deflection is seen to become greater than the linear value as the electrons first approach the grid wires. This is due to lens aberration. The deflection then decreases. This is because, as may be seen in Fig. 9.10, those electrons which come very close to the grid wires are so strongly deflected that they come within the influence of the next grid wire and suffer a deflection in the opposite direction.

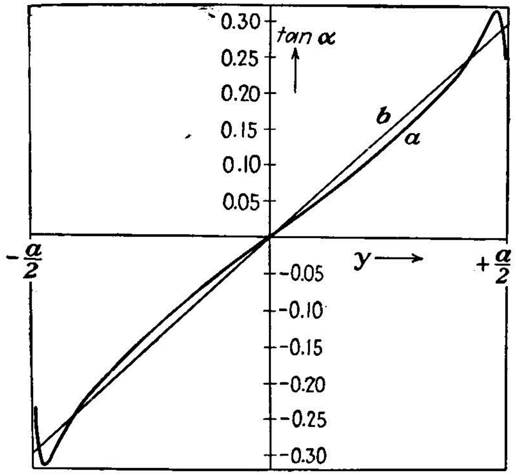


FIG. 11.17.—Deflection angle of an electron as a function of the offset from the mid-point between grid wires.

Introducing the value of f from Eq. (11.51) and the expression for velocity in terms of potential,

$$v_{yn} = \frac{2.96 \times 10^5}{\sqrt{V_n}} y_0 \left[\left(\frac{dV}{dx} \right)_R - \left(\frac{dV}{dx} \right)_L \right] \tag{11.54}$$

in which v_{yn} is the sidewise component of velocity acquired, in meters per second, from the nth grid, V_n is the mean grid potential of the nth grid, in volts, and $\left(\frac{dV}{dx}\right)_R$ and $\left(\frac{dV}{dx}\right)_L$ are the gradients of potential to the right and left of the nth grid plane, respectively, in volts per meter. The corresponding expressions for the three grids of a pentode become

$$v_{y1} = \frac{2.96 \times 10^5}{\sqrt{V_1}} y_0 \left(\frac{V_2 - V_1}{d_{12}} - \frac{V_1}{d_{c1}} \right) \tag{11.55}$$

$$v_{y2} = \frac{2.96 \times 10^5}{\sqrt{V_2}} y_0 \left(\frac{V_3 - V_2}{d_{23}} - \frac{V_2 - V_1}{d_{12}} \right) \tag{11.56}$$

$$v_{y3} = \frac{2.96 \times 10^5}{\sqrt{V_3}} y_0 \left(\frac{V_p - V_3}{d_{3p}} - \frac{V_3 - V_2}{d_{23}} \right) \tag{11.57}$$

in which the sidewise components of velocity are in meters per second, grid potentials are understood to be mean grid-plane potentials in volts, and distances are in meters.

The sidewise components of velocity are additive in the form shown in Fig. 11.18. Curve *a* of this figure shows the distribution of the sidewise components of velocity after passing through one grid. The sidewise components of velocity are uniformly distributed between plus and minus v_1 where this is the maximum component of this tangential velocity. After passing through two grids the distribution of velocities has the form shown in *b* of Fig. 11.18. This is a trapezoidal figure with velocities reaching from $v_1 + v_2$ to the negative of the same quantity. In this particular figure it has been assumed that v_2 is less than v_1. The distribution given is arrived at by sliding a rectangular aperture of width $2v_2$ and of the same height as the rectangle of part *a* of the figure over the rectangle of part *a* and plotting the exposed area as a function of the displacement of the aperture. Upon repeating the process with an aperture of width $2v_3$, the distributions of c_1 and c_2 result. The distribution of c_1 results from the assumption that v_3 is less than $v_1 - v_2$, whereas the distribution of c_2 results when v_3 is greater than $v_1 - v_2$.[1] Both

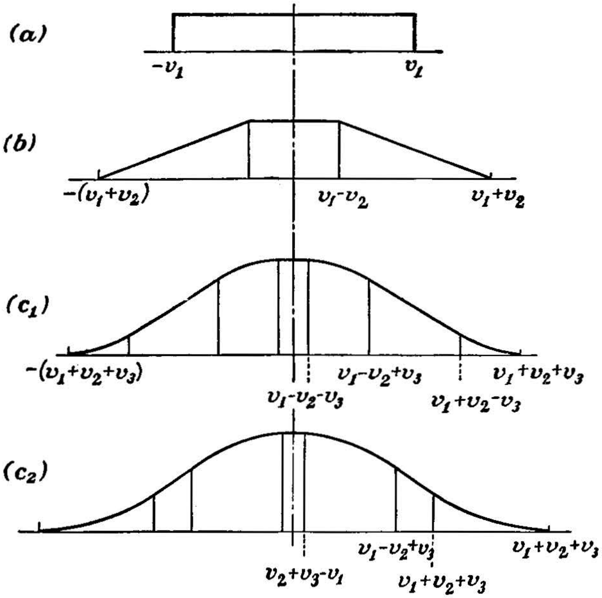


FIG. 11.18.—Distribution of sidewise component of velocity in a beam of electrons scattered by one, two, or three grids.

---

[1] JONKER, J. L. H., Electron Trajectories in Multigrid Valves, *Philips Tech. Rev.*, vol. 5, pp. 131–139, May, 1940.

these distributions consist of parabolic and straight-line sections with the same over-all span. The difference is that the distribution of c_1 has a straight-line center section, while that of c_2 has a parabolic center section of large curvature. Both these distributions already show approximately the form of a Gauss error curve, which they would obtain from the random deflection of a large number of grids.

If the distribution of tangential velocities as given above is known, the plate-current–plate-voltage characteristic can be calculated. It is necessary only to remember that at every plate voltage V_p only those electrons will reach the plate whose tangential velocity after being deflected by all grids is less than a maximum value given by

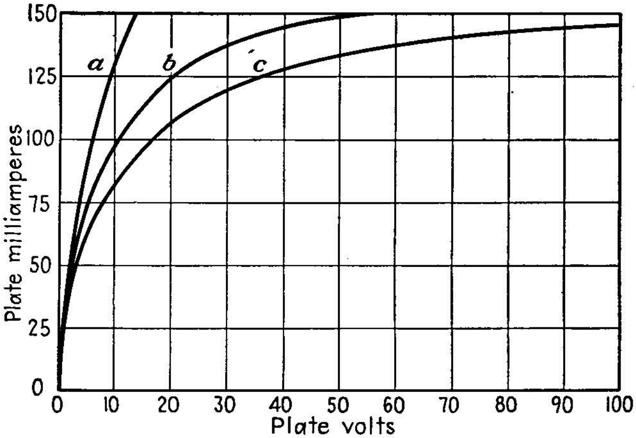


FIG. 11.19.—Plate-current–plate-voltage characteristic of a tube with electrons scattered by one, two, or three grids.

$$v_{y\,\max} = 5.93 \times 10^5 \sqrt{V_p} \tag{11.58}$$

This follows from the fact that an electron reaching the plate will have the value of velocity given by Eq. (11.58), and if all this velocity exists in a sidewise component then the electron will graze the plate and fall back through more positive spaces in the suppressor grid. If the plate-current–plate-voltage characteristic be calculated on this basis, the curves of Fig. 11.19 result. The curves shown are for the corresponding velocity diagrams of Fig. 11.18. Curve *a* is similar to that which occurs in a beam-power tube. This is a tetrode with aligned grids in which the net effect of deflection by two grids is not very different from that of one grid as seen in Fig. 10.6. This effect undoubtedly contributes to the sharpness of the shoulder of the beam-power tube. The effect of two grids is shown at *b*. This would correspond to the curve of an ordinary tetrode free from secondary emission. The curve c is a typical pentode characteristic resulting from the random action of three grids. It is evident that it is necessary to keep the total sidewise velocity component of the electrons low in order to achieve a sharp shoulder to the characteristic. Since all three grids contribute to this in approximately the same amount, it is necessary to study the effect of each of the grids to see what can be done to reduce the resulting sidewise component of velocity. Examination of Eq. (11.54) shows that, in general, the sidewise component of velocity introduced by grid deflection may be reduced by either reducing the grid pitch or decreasing the change in the gradient of potential on passing through the grid.

Deflections Due to the Control Grid. The same factors that give rise to a large mutual conductance also give rise to small deflection. These factors are small grid pitch and a small value of cathode-grid spacing. It might be thought that a small cathode-grid spacing would give rise to a large change in the slope of the potential curves on the two sides of the control grid, but this is not the case, for the mean grid-plane potential increases as the cathode-grid distance decreases. Nothing much can therefore ordinarily be done with the control grid to decrease the electron deflection.

Deflections Due to the Screen Grid. The electron deflections due to the screen grid may be reduced by increasing the distances d_{12} and d_{23} in cases in which transit time is not a consideration. They can also be reduced by decreasing the grid pitch, though there is a limit to this method, for the current intercepted by the screen grid increases as this is done.

Deflections Due to the Suppressor Grid. The suppressor pitch cannot be made too small, for then the mean suppressor-plane potential becomes too small and offsets the effect of the small grid-wire spacing as far as electron deflections are concerned. It is, however, possible to make the suppressor-grid–plate distance quite small, with resulting improvement in the deflection characteristics. This has the added advantage, as is apparent in Fig. 11.15, that the current is concentrated in front of the plate, giving rise to considerable space charge, which aids in increasing the retarding potential presented to the secondary electrons originating at the plate. The bad deflection situation that results from the use of a large suppressor-grid–plate distance is shown in Fig. 11.14.

By making use of the possibilities indicated in the above outline it is possible to make pentode tubes with shoulders of the plate-current–plate-voltage characteristic nearly as sharp as those of the beam-power tube.

CHAPTER 12

NOISE IN VACUUM TUBES

12.1. Noise as a Limiting Factor in the Ultimate Sensitivity of Electronic Devices. Vacuum-tube amplifiers making use of triodes and pentodes are capable of giving extremely large amplification of power and voltage. In fact, it may be said that an amplification of any desired magnitude may be achieved by the use of vacuum tubes. At first glance this would seem to imply that arbitrarily small signals could be detected. Ultimately, however, it is found that there is a limit determined by the noise generated by the random motion of electrons at the input of the circuit. Any signal whose level is appreciably less than that of the electron noise will be masked by it. The order of the equivalent voltage of the electron noise is extremely small, of the order of millimicrovolts, but many electronic devices have enough amplification to bring this up to a detectable level.

Electron noise shows up in both resistors and in vacuum tubes. Even in a passive resistor, the molecular and the electronic agitation is evident with sufficient amplification. Here the noise is referred to as "thermal-agitation noise." In vacuum tubes the random emission and fluctuation of space-charge-limited currents contribute a similar noise. In temperature-limited tubes this noise is called "shot noise" and is due to random emission. In space-charge-limited emission tubes the noise is much less and is called "reduced shot noise." Both types of noise are characterized by a uniform distribution of energy over the frequency spectrum. Depending upon the application, the noise from either the tubes or the resistors at the input of an electronic device may predominate.

Needless to say, resistor and tube noise is an exclusive concern of electronic devices. No other type of device can have sufficient sensitivity to be limited by random electron motion. Resistor and tube noise set the ultimate sensitivity of high-gain amplifiers, receivers, phototube input circuits, and television pickup tubes. Although resistor and tube noise can never be avoided, much can be done by circuit design and selection of tubes to approach the minimum attainable noise.

Although the formulas for various types of electron noise and their application are quite simple, their derivation is dependent upon some intimate aspects of thermodynamics and the kinetic theory of gases that have not been developed in this book. For this reason, only the basis of the development will be given, and emphasis will be upon the interpretation and application of the formulas.

12.2. Noise in Resistors. Noise in resistors is due to the random motion of electrons within them. The noise energy is proportional to the resistance, the absolute temperature, and the frequency band width over which the noise is observed and is independent of the material of which the resistor is made. The noise energy increases with absolute temperature because the molecular agitation is proportional to this. The noise is probably made up of extremely short and sharp pulses resulting from the impact of the electrons with the molecules. These pulses are probably so short and sharp that they are made up of a continuous distribution of frequency components of equal amplitude up to the highest frequencies in use today. As a result, the noise energy is uniformly distributed over the useful r-f spectrum.

The mean-square thermal-agitation noise voltage $\bar{e}^2$ across the terminals of a resistor R at an absolute temperature T, associated with a frequency band width B, is

$$\bar{e}^2 = 4kT_rRB \tag{12.1}$$

where e is rms value of the noise voltage, volts
k is Boltzmann's constant, 1.3805×10^{-23} watt-second per °K
R is resistance, ohms
T_r is room temperature, °K (°C + 273)
B is frequency band width, cycles per sec[1,2]

If room temperature is taken as 290°K (63°F), the expression for the rms noise voltage becomes

$$e_{\text{rms}} = 126.0\sqrt{RB} \qquad \text{micromicrovolts} \tag{12.2}$$

A nomograph of this equation is shown in Fig. 12.1. For the sample construction line shown, the rms value of the noise voltage across a 1,000-ohm resistor in a frequency band of 10 mc is 12.6 microvolts.

The effect of thermal-agitation noise can be expressed either as an emf in series with the resistor considered noiseless or as a constant-current generator in parallel with the resistor considered noiseless. This follows from application of Thévénin's and Norton's theorems. The two equivalent circuits of a noisy resistor are shown in Fig. 12.2. The circle in the figure indicates a zero-impedance constant-voltage generator

---

[1] JOHNSON, J. B., Thermal Agitation of Electricity in Conductors, *Phys. Rev.*, vol. 32, pp. 97–109, July, 1928.

[2] NYQUIST, H., Thermal Agitation of Electric Charge in Conductors, *Phys. Rev.*, vol. 32, pp. 110–113, July, 1928.

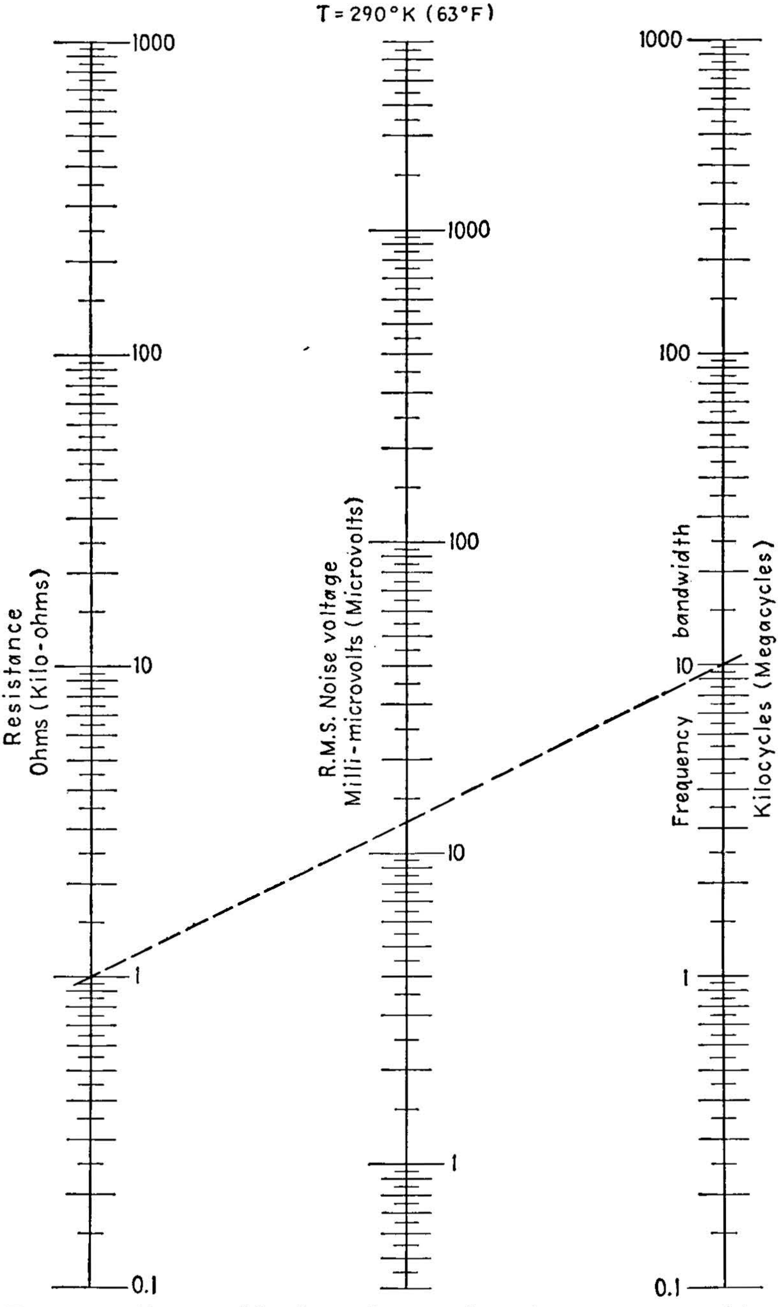


FIG. 12.1.—Nomographic chart of rms noise voltage across a resistor.

of the value given by Eq. (12.2). The square in the figure indicates an infinite-impedance constant-current generator whose output is

$$i_{\text{rms}} = 126.0\sqrt{\frac{B}{R}} \qquad \text{micromicroamperes} \tag{12.3}$$

at room temperature. This is obtained by dividing the expression for the rms noise voltage by the resistance. The general form of the above equation is

$$i_{\text{rms}} = \sqrt{\frac{4kTB}{R}} \qquad \text{amperes} \tag{12.4}$$

where all the symbols have their previous significance. A nomograph of the rms noise-current equivalent of a noisy resistor to fit the right circuit of Fig. 12.2 is given in Fig. 12.3. For the sample construction line shown, the equivalent rms noise current of a 1,000-ohm resistor is 4 millimicroamperes for a frequency band width of 1 mc.

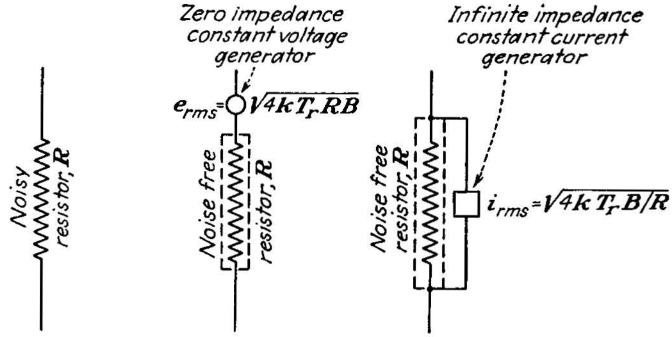


FIG. 12.2.—Equivalent circuits of a noisy resistor.

The noise power associated with thermal-agitation noise in a resistor is i^2R, or

$$N = 4kTB \qquad \text{watts} \tag{12.5}$$

Note that this is independent of the value of resistance. The *available* noise power that can be obtained from a resistor by perfect matching is just one-fourth of this value. This follows immediately from maximizing the power obtainable from the middle circuit of Fig. 12.2 by varying the resistance load on such a generator. Maximum output power is obtained when the load resistance equals the generator resistance and has a value of $\frac{e^2}{4R}$. Thus the *available* noise power is

$$N_a = kTB \qquad \text{watts} \tag{12.6}$$

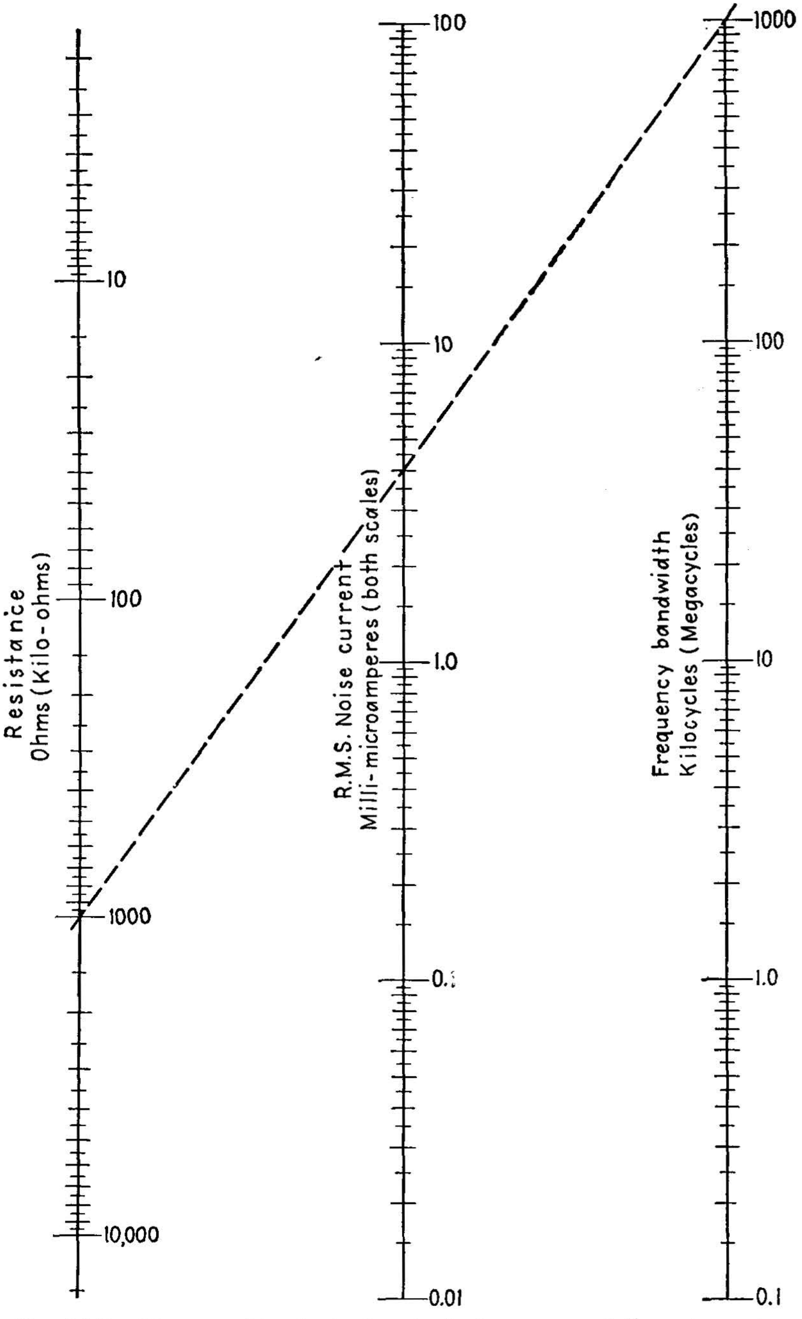


FIG. 12.3.—Nomographic chart of equivalent rms current through a noisy resistor.

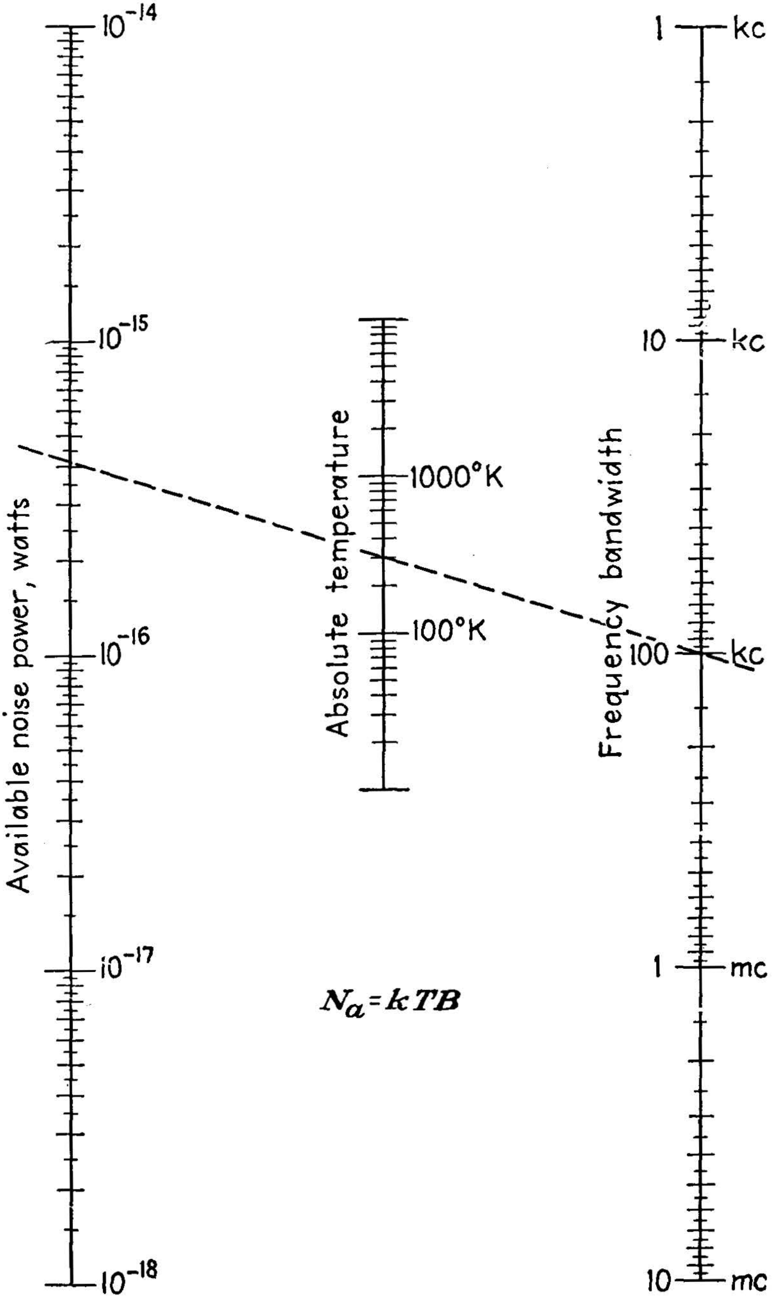


FIG. 12.4.—Nomographic chart of noise power available from a resistor.

At room temperature of 290°K this has a value of

$$N_a = 4.0 \times 10^{-21} B \qquad \text{watts} \tag{12.7}$$

A nomograph of this equation is shown in Fig. 12.4. A temperature of 300°K and a frequency band width of 10 kc is seen to give an available noise power of 4.1×10^{-16} watt. *It is convenient to remember that at room temperature the available noise power for a* 1-*mc band width is* 144 *db below* 1 *watt.*

When two resistances at different temperatures are connected in parallel, the mean-square noise voltage becomes

$$\bar{e}^2 = 4k \frac{R_1 R_2 (R_2 T_1 + R_1 T_2)}{(R_1 + R_2)^2} B \qquad \text{volts}^2 \tag{12.8}$$

where the resistances are in ohms and the temperatures in degrees Kelvin, and Boltzmann's constant is 1.3805×10^{-23} watt-sec. When several resistances are connected in parallel, the results are better expressed in terms of conductances. Let conductances $G_1, G_2, G_3, \ldots, G_n$ at temperatures $T_1, T_2, \ldots, T_n$, respectively, be connected in parallel. The resulting mean-square noise voltage is

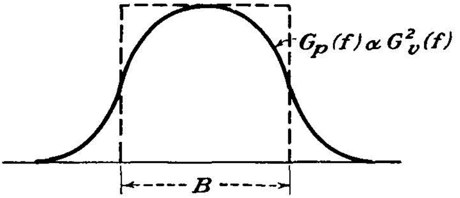


FIG. 12.5.—Definition of equivalent band width.

$$\bar{e}^2 = 4k \frac{\sum_1^n G_n T_n}{\left(\sum_1^n G_n\right)^2} B \qquad \text{volts}^2 \tag{12.9}$$

Some care must be used in determining the band width to fit the above expressions. Since the concepts involved are basically those concerned with power, the equivalent band width of any device must be defined in terms of the power-frequency curve. In the case of an amplifier the band width is defined as that frequency interval for which a power gain equal to the gain at mid-band would transmit the same noise energy as does the actual power-gain–frequency curve. Analytically, this becomes

$$B = \frac{\int_0^\infty G_p(f)\,df}{G_p(f_0)} \tag{12.10a}$$

where G_p is the power gain. All this amounts to is finding the width of a rectangular power-gain–frequency curve of height equal to the mid-frequency power gain of the actual curve as shown in Fig. 12.5. In terms of voltage amplification, the equivalent band width is

$$B = \frac{\int_0^\infty G_v^2(f)\,df}{G_v^2(f_0)} \tag{12.10b}$$

where G_v is the voltage gain.

12.3. Sources of Noise in Tubes. Noise can occur in vacuum tubes from a rather imposing list of sources. The principal sources of noise in tubes are

1. Shot effect (temperature-limited emission).
2. Reduced shot effect.
3. Flicker effect.
4. Collision ionization.
5. Random division of current between electrodes.
6. Induced noise at ultra-high frequencies.
7. Faulty tube construction.
    - *a.* Hum.
    - *b.* Poor insulation.
    - *c.* Vibration.
    - *d.* Varying wall charges.

Shot effect is the noise associated with random emission in a tube whose emission is temperature-limited. This is probably the loudest of the electronic tube noises but not the most serious, for tubes are seldom operated so that their emission is temperature-limited.

The so-called "reduced shot effect" is observed in tubes whose emission is space-charge-limited. The magnitude of the noise is much less than in tubes whose emission is temperature-limited. In this case the space charge exerts a smoothing action upon the true shot effect, and the noise is principally due to variations in the space-charge currents.

Flicker effect is observed with oxide cathodes. This effect is associated with variations in the activity of the emitting surface. The effect is much more noisy than the true shot effect for temperature-limited emission. When the emission is space-charge-limited, the magnitude of the noise is greatly reduced.

Noise in tubes is raised by the presence of an appreciable amount of gas. This is due to the fact that gas molecules are ionized by collision with emitted electrons and the positive ions formed subsequently liberate little bursts of electrons as they penetrate the virtual cathode in front of the emitting surface. Gas noise is inappreciable unless the positive-ion gas current is more than a few hundredths of a microampere. Such gas noise appears mostly below 10 mc.

Random division of current between electrodes contributes to the noise of multielectrode tubes and makes pentodes three to five times as noisy as the same tube connected as a triode. It may be said that multielectrode tubes will always be noisier than triodes because of this additional factor contributing to the tube noise.

The ultra-high-frequency components of the random fluctuations of space charge in a tube will induce voltages in the grid circuit, which in turn will react back upon the space-charge flow. This effect is important only for frequency components above 30 mc.

Noise due to faulty tube construction is always present to a degree. If the filament is not sufficiently noninductive, hum will result. If insulation is poor at any point in the tube, there will be leakage currents, which will generally create noise because of nonconstant leakage resistance. Vibration may be a factor in an electromechanical feedback circuit. Dirt on the glass inside of a tube may give rise to varying wall charges, which will influence the tube current in a noisy manner. All these items can, however, with sufficient care in construction be held to a very low level.

Items 1, 2, 4, 5, and 6 listed above can never be removed entirely. They are, however, subject to an analysis that shows how their effects may be minimized. These items will be the subject of the subsequent sections. It has been found in most cases that it is convenient to express tube-noise effects in terms of equivalent noisy resistors. These resistors in turn are considered to have internal-noise emfs.

12.4. Shot Noise in Diodes with Temperature-limited Emission. Noise in diodes was probably the first form of tube noise ever detected. It is generally referred to as shot noise but also as "*Schrot* noise" and "Schottky noise," after the scientist who first analyzed the effect.[1,2] The noise is due to the random emission and arrival of electrons

---

[1] SCHOTTKY, W., Spontaneous Current Fluctuations in Electron Streams, *Ann. Physik*, vol. 57, pp. 541–567, Dec. 20, 1918.

[2] See also FRY, T. C., The Theory of the Schroteffekt, *Jour. Franklin Inst.*, vol. 199, February, 1925.

at the plate. It cannot be explained in terms of individual electron emission or arrival alone, for if the electrons were emitted at a uniform rate the lowest frequency component of noise would be above the highest frequency that vacuum tubes can handle. Thus a current of 1 ma corresponds to a flow of approximately 10^{16} electrons per sec. If these did flow at a uniform rate, there would be no noise components below 10^{16} cycles per sec. The electron stream evidently exhibits rather pronounced variations in density caused by the electrons arriving in groups. The mean square of the fluctuation components of current is

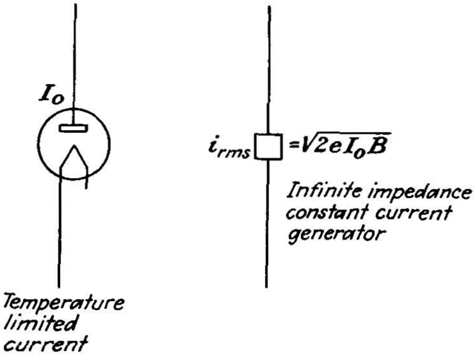


FIG. 12.6.—Constant-current-generator equivalent of a diode with temperature-limited current.

found to depend only upon the magnitude of the emitted current and the frequency band width
where e is electronic charge, 1.6020×10^{-19} coulomb

$$\overline{i^2} = 2eI_0B \qquad \text{amperes}^2 \tag{12.11}$$

I_0 is emission current, amperes
B is band width, cycles per sec

This expression may more conveniently be written

$$\overline{i^2} = 3.2041 \times 10^{-19} I_0B \qquad \text{amperes}^2 \tag{12.12}$$

If the current from a diode with temperature-limited emission is put through a resistor R, the diode effectively puts a noise power of value $\overline{i^2}R$ into the resistor.

The above relations have been verified experimentally and they are reproducible to a high degree of accuracy. This property makes the diode with temperature-limited emission valuable as a standard noise source for such purposes as receiver and amplifier sensitivity measurement.

The diode with temperature-limited emission acts like a constant-current generator as far as noise energy is concerned and will put a noise current given by the above equations through a resistor of any size. The equivalent circuit of the diode with temperature-limited emission is shown in Fig. 12.6.

12.5. Reduced Shot Effect in Diodes with Space-charge-limited Emission. In diodes in which the emission is space-charge-limited, the shot noise is much less than in the same diode passing the same current when its emission is temperature-limited.[1–4] The noise power is of the order of 10 per cent of that encountered for the same current when the emission is temperature-limited. The space charge thus has a very definite "smoothing" action upon the shot effect, giving rise to what may be called the "reduced shot effect." The mechanism of the smoothing action of the space charge is something like this: The virtual cathode in front of the emitting surface has a potential lower than that of the emitter by a value determined by the mean velocity of emission. Electrons with all velocities are storming this potential hill, and those with velocities greater than the mean velocity will *on the average* get past the virtual cathode and go on to the plate. Occasionally there will come a group of electrons with a velocity slightly in excess of that needed to get past the virtual cathode. When this occurs, the potential minimum at the virtual cathode is momentarily depressed by the additional space charge and as a result a few electrons that normally would have got past the potential minimum fail to do so and are returned to the emitter. This means that for every burst of electrons which might give rise to noise there is a compensating current set up in the opposite direction which tends to cancel the noise produced by the burst. The net result is an over-all reduction in noise that is considerable. The resulting noise levels are, however, high enough still to be of concern in the design of electronic equipment.

By considering the action of each increment of the velocities encountered in the process of emission some fairly good theoretical expressions for the reduced shot noise may be obtained. If the ratio of the noise

---

[1] RACK, A. J., Effect of Space Charge and Transit Time on the Shot Noise in Diodes, *Bell Sys. Tech. Jour.*, vol. 17, pp. 1–28, October, 1938.

[2] NORTH, D. O., Fluctuations in Space-charge-limited Currents at Moderately High Frequencies, *RCA Rev.*, vol. 4, Part II, pp. 441–473, April, 1940; vol. 5, pp. 244–260, July, 1940.

[3] WILLIAMS, F. C., Fluctuations of Space Charge Limited Currents in Diodes, *Jour. I.E.E.*, vol. 89, Part III, pp. 219–229, December, 1941.

[4] BELL, D. A., Fluctuations in Space Charge Limited Currents, *Jour. I.E.E.*, vol. 89, Part III, pp. 207–212, December, 1942.

See also SCHOTTKY, W., *Wiss. Veröffent. Siemens-Werken*, vol. 14, p. 15, 1937.

power of a diode passing a given current with and without space-charge limitation of current be designated by Γ^2, then, for a plane-electrode diode with a large ratio of plate voltage V_p to average emission voltage (velocity equivalent) V_e,

$$\Gamma^2 = \frac{9V_e}{V_p} \tag{12.13}$$

seems to be functionally correct.[1,2] The mean-square noise current can be expressed by an equation similar to that for the diode with temperature-limited emission as

$$\bar{i}^2 = \Gamma^2 2eI_0B \qquad \text{amperes}^2 \tag{12.14}$$

by analogy to Eq. (12.11). The mean-square noise current can also be written in the form

$$\bar{i}^2 = \theta\,\frac{4kT_cB}{R_1} \qquad \text{amperes}^2 \tag{12.15}$$

by analogy to Eq. (12.4). In this form, θ is a dimensionless parameter that has an asymptotic value of 0.644 for large ratios of plate to minimum potential, T_c is the absolute cathode temperature, and R_1 is the a-c diode resistance. Theoretically, the parameter θ is within a fraction of a per cent of the asymptotic value as long as the plate current is less than 80 per cent of the emission current and the plate voltage is greater than 2 volts for normal oxide operating temperatures.[3–5] Experimentally determined values of diode noise are 50 per cent higher than predicted by the theoretical expressions of Eqs. (12.13) and (12.15), so that θ assumes a value of unity for diodes. *The significance of Eq.* (12.15) *is that the noise power from a diode whose emission is space-charge-limited is the same as that from a resistor at the cathode temperature equal to the a-c diode resistance.*[6] The equivalent circuit for this case is given in Fig. 12.7.

---

[1] WILLIAMS, *op. cit.*

[2] BELL, *op. cit.*

[3] PEARSON, G. L., Shot Effect and Thermal Agitation in an Electron Current Limited by Space Charge, *Physics*, vol. 6, pp. 6–9, January, 1935.

[4] RACK, *op. cit.*

[5] NORTH, *op. cit.*

[6] Combination of Eqs. (12.14) and (12.15) and the use of the fact that $R_1 = \frac{2}{3}\frac{V_p}{I_0}$ for the diode with space charge suggests that Γ^2 has the value $(0.644)\,\frac{3kT_c}{eV_p}$, which is consistant with the observed relation of Eq. (12.13).

12.6. Reduced Shot Effect in Triodes with Space-charge-limited Current. Triodes, too, are noisy. In fact, the noise in triodes, as pointed out before, may be a limiting factor in the sensitivity of an electronic device. The noise is due to the effects just observed in diodes and shows as a fluctuation in plate current. It is convenient to interpret the noise in a triode as being due to a noisy resistor in series with the grid of the tube considered free from noise. The value of this noisy resistor in series with the grid is

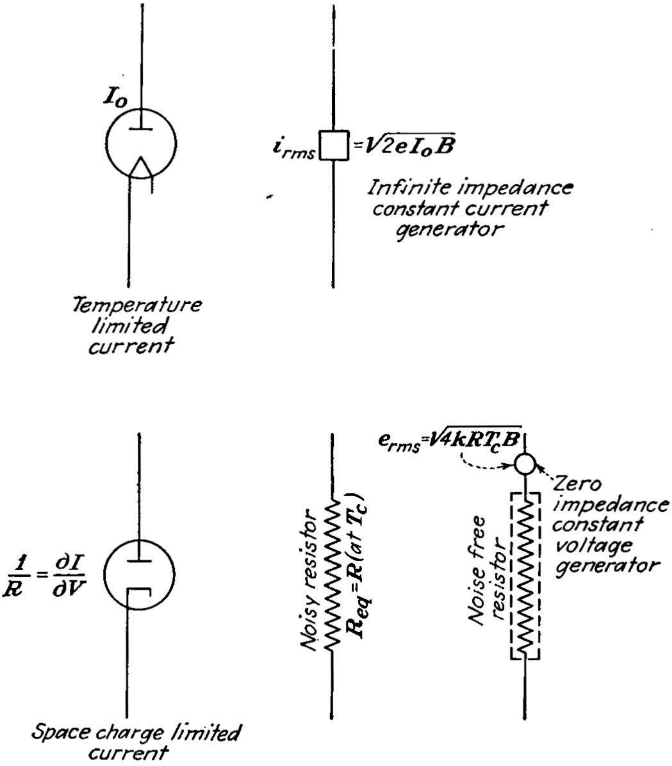


FIG. 12.7.—Resistor equivalent to a diode with space-charge-limited current.

$$R_{eq} = \frac{\theta T_c G}{G_m T_r G_m} \qquad \text{ohms} \tag{12.16}$$

where θ is the effective cathode-temperature ratio of approximate value two-thirds, G_m is the mutual conductance of the triode, G is the a-c conductance of the diode equivalent of the triode, T_c is absolute cathode temperature, and T_r is absolute room temperature. The ratio of equivalent-diode conductance to triode mutual conductance is the inverse-square ratio of equivalent-diode spacing to the cathode-grid spacing as discussed in Chaps. 7 and 8. For the plane-electrode diode, from Eq. (8.45), this ratio is

$$\frac{G}{G_m} = \left[1 + \frac{1}{\mu}\left(\frac{d_{cp}}{d_{cg}}\right)^{4/3}\right]^{3/2} \tag{12.17}$$

The corresponding ratio for cylindrical triodes is difficult to express exactly but is given approximately by

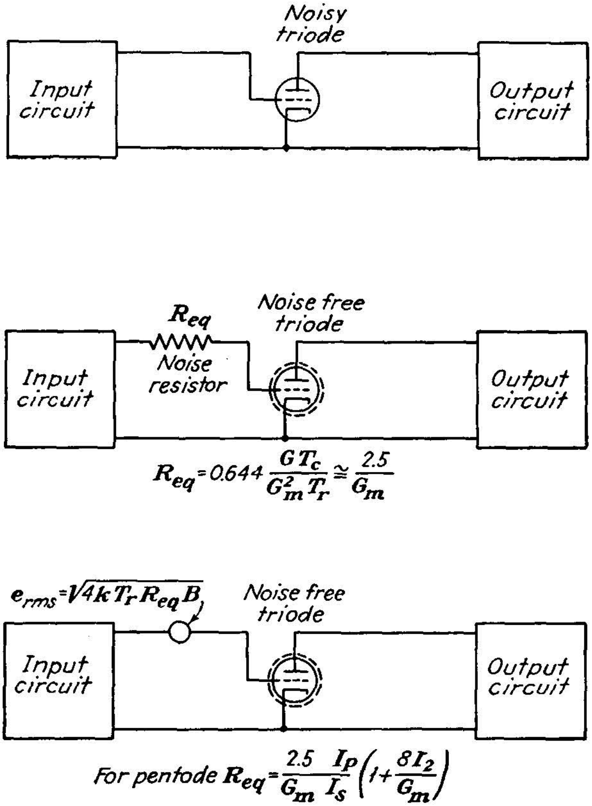


FIG. 12.8.—Equivalent circuit of a noisy triode.

$$\frac{G}{G_m} = 1 + \frac{1}{\mu}\left[1 + \frac{2}{3}\ln\frac{r_p}{r_g}\right] \tag{12.18}$$

The value of this ratio will generally lie between 1 and 2. Assuming that a typical value of the conductance ratio is 1.25 and that the cathode temperature is 3.33 times the room temperature, Eq. (12.16) reduces to

$$R_{eq} = \frac{2.5}{G_m} \qquad \text{ohms} \tag{12.19}$$

which is sufficiently accurate for a good many purposes.[1] The value of equivalent resistance thus given is that resistance which if inserted in series with the grid of the triode considered noise-free would cause as much noise current in the plate circuit of the tube as does the reduced shot effect. The equivalent circuits of triode in terms of a noise-generating resistor and a noise-free tube is given in Fig. 12.8. Observed values of noise in a triode agree very closely with the values predicted by the above equations, much more so than was the case with diodes.

12.7. Noise Due to Gas in Tubes. When there is gas in tubes, there is an extra component of noise due to the electrons and ions liberated by collision ionization. The electrons passed by the grid will collide with some of the gas molecules, forming positive ions and liberating more electrons. The liberated electrons will pass on to the plate and give rise to some extra noise. The positive ions will be attracted to the negative grid and flowing through the external impedance will cause a voltage in the grid that will also give rise to noise. The noise is proportional to the number of ions formed, which in turn is proportional to the normal space current and to the number of gas molecules, or gas pressure. Fortunately, the positive-ion grid current is a measure of the number of ions formed per second and can be used to determine the noise without knowing the gas pressure.[2,3] As with other components of noise, the noise can be described in terms of a resistor in series with the control grid of the tube considered noise-free. The equivalent noise-generating grid resistor is

$$R_{eq} = \left(20R_g^2 + 4 \times 10^4 \frac{I_p}{G_m^3}\right) I_1 \qquad \text{ohms} \tag{12.20}$$

where R_g is shunt resistance of the grid circuit, ohms
G_m is mutual conductance, mhos
I_p is plate current, amperes
I_1 is control-grid current, amperes

---

[1] HARRIS, W. A., Space Charge Limited Current Fluctuations in Vacuum Tube Amplifiers and Input Systems, *RCA Rev.*, vol. 5, pp. 505–524, April, 1941; vol. 6, pp. 114–124, July, 1941.

[2] *Ibid.*

[3] THOMPSON, B. J., and D. O. NORTH, Fluctuations in Space-charge-limited Currents Caused by Collision Ionization, *RCA Rev.*, vol. 5, pp. 371–388, January, 1941.

The first term in this expression is due to the flow of positive-ion current through the external grid impedance. The second term is due to the electrons liberated upon ionization that are attracted to the plate.

As an example, consider the case of a gassy tube for which the positive-ion control-grid current is 0.01 microampere. Let the mutual conductance of the tube be $5{,}000 \times 10^{-6}$ mho, the plate current 1 ma, and the shunt resistance of the grid circuit 100,000 ohms. Then the first term of Eq. (12.20) contributes a noise resistance of 2,000 ohms, and the second term contributes a noise resistance of 3.20 ohms. The second term of Eq. (12.20) is usually much smaller than the first term, as in this example, and can ordinarily be neglected.

12.8. Reduced Shot Effect in Multielectrode Tubes with Space-charge-limited Currents. Pentodes are even noisier than triodes—by a considerable factor. In fact, by comparison, triodes are relatively quiet. The additional noise in pentodes is due to the random division of the fluctuation noise between the electrodes. The individual groups of electrons that burst through the virtual cathode are quite local in their impingement upon electrodes in their subsequent travel, but the compensating currents due to the displacement of the virtual cathode are more or less uniformly distributed between the electrodes in the ratio of the direct currents. As a result, the smoothing, or compensating, action in a pentode is much less pronounced than in a triode.

As with the triode, the noise of a pentode can be expressed as being due to a noisy resistor in series with the control grid of the tube considered noise-free. The equivalent resistor in this case is[1,2]

$$R_{eq} = \frac{2.5}{G_m}\frac{I_p}{I_s}\left(1 + 8\,\frac{I_2}{G_m}\right) \tag{12.21}$$

where all the symbols have their customary significance. Currents must be expressed in amperes and conductances in mhos to yield equivalent resistance in ohms. The first factor of Eq. (12.21) is seen to be the equivalent resistance for a triode. The remainder of the expression generally increases the value of the triode equivalent resistance by a factor of three to five. For screen currents much smaller than the plate current the noise in both the plate and screen circuits is approximately equal to the true shot effect for a current equal to the screen current. The value of the mean-square noise current in the plate circuit is readily obtained from this by means of Eq. (12.12) and converted to an equivalent noisy resistor *in the plate circuit* by means of the nomograph of Fig. 12.3.

---

[1] NORTH, D. O., Fluctuations in Space Charge Limited Currents in Multi-collectors, *RCA Rev.*, vol. 5, pp. 244–260, October, 1940.

[2] HARRIS, *op. cit.*

12.9. Noise in Mixer Tubes. Mixers are also noisy. The factors that contribute to the noise of triodes and pentodes also contribute to mixer noise. In a mixer a large voltage from the local oscillator is applied to the tube so that the current and the mutual conductance swing over a large value. Since the noise of tubes is known as a function of current and mutual conductance, it is not too difficult to evaluate the mixer noise.

In a mixer it is the noise in the intermediate-frequency band that is of importance. The noise in this band will vary periodically over the local oscillator cycle as the mutual conductance and current vary with the voltage applied at the local oscillator frequency. The total intermediate-frequency noise can be obtained by summing the plate noise over the local oscillator cycle.

$$\bar{\imath}_{i-f}^2 = \frac{1}{2\pi} \int_0^{2\pi} \bar{\imath}_{pn}^2 \, d(\omega t) \tag{12.22}$$

where $\bar{\imath}_{pn}^2$ is the mean-square noise current in the plate at any instant t and ω is 2π times the local oscillator frequency. It is convenient (except in the case of the diode mixer) to express the mixer noise in terms of an equivalent noise resistance in series with the control grid of the tube considered as noise-free. Thus

$$\bar{e}_n^2 = \frac{\bar{\imath}_{i-f}^2}{G_c^2} \tag{12.23}$$

where G_c is the conversion transconductance of the tube. Corresponding to this value of input noise, the equivalent input noisy resistor is

$$R_{\text{eq}} = \frac{\bar{e}_n^2}{4kT_rB} \tag{12.24}$$

$$R_{\text{eq}} = \frac{\bar{\imath}_{i-f}^2}{4kT_rG_c^2B} \tag{12.25}$$

Upon applying the above ideas there are obtained the results shown in Fig. 12.9 for a fictitious pentode tube connected in accordance with the four most common mixer connections.[1] In this figure, G_0 is the maximum value which the mutual conductance assumes over the local oscillator cycle, generally that corresponding to zero grid voltage. The quantity G_x is the maximum value of the screen-plate transconductance. The quantity I_x is the maximum value of plate current when the signal is

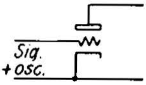
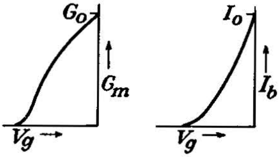
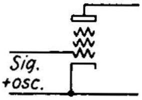
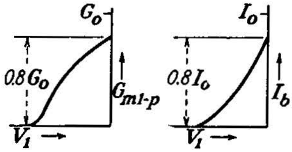
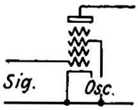
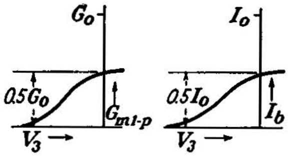
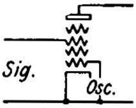
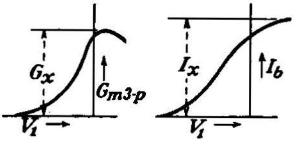
---

[1] Herold, E. W., Superheterodyne Converter Considerations in Television Receivers, *RCA Rev.*, vol. 4, pp. 324–337, January, 1940.

See also the summarizing articles by Herold, E. W., and L. Malter, Some Aspects of Radio Reception at Ultra-high Frequency, *Proc. I.R.E.*, vol. 31, pp. 423–438, 491–510, 567–582.

injected into the screen circuit and the local oscillator into the control-grid circuit.

It is seen that the triode is the best converter and the pentode with signal and local oscillator applied to the control grid is the next best. The others are too noisy for high-sensitivity applications.

TYPE OF FIRST DETECTOR	TYPICAL CHARACTERISTICS	APPROX. CONVERSION TRANSCOND.	APPROX. EQUIV. NOISE RESIST.	EQUIVALENT NOISE MICROVOLTS ON GRID*
		$0.28\,G_o$	$\frac{13}{G_o}$	8
		$0.23\,G_o$	$\frac{15+21\frac{I_o}{G_o}}{G_o}$	16
		$0.14\,G_o$	$\frac{120\frac{I_o}{G_o}}{G_o}$	32
		$0.28\,G_x$	$\frac{57\frac{I_x}{G_x}}{G_x}$	—

* Assuming $G_o = 15 \times 10^{-3}$ mhos
$I_o = 30 \times 10^{-3}$ amps
$\Delta f = 4$ mc

FIG. 12.9.—Comparison of fictitious modulators presumed to have similar cathode and first-grid structures. (*Herold.*)

When the values required in the tabulation of Fig. 12.9 are not available, the following formulas will be found sufficiently accurate for most purposes:

For triode mixers,

$$R_{eq} = \frac{4}{G_c} \tag{12.26}$$

or

$$R_{eq} = \frac{2.5\bar{G}_m}{G_c^{\,2}} \tag{12.27}$$

For pentode mixers

$$R_{\text{eq}} = \frac{I_p}{I_s}\left(\frac{4}{G_c} + \frac{20I_2}{G_c^2}\right) \tag{12.28}$$

or

$$R_{\text{eq}} = \frac{I_p}{I_s}\left(\frac{2.5\bar{G}_m}{G_c^2} + \frac{20I_2}{G_c^2}\right) \tag{12.29}$$

where currents must be expressed in amperes and conductances in mhos. $\bar{G}_m$ is the mean transconductance over the local oscillator cycle, and G_c is the conversion transconductance of the local oscillator. In addition, the following rules of thumb may be applied,

$$G_c \text{ (as converter)} = \tfrac{1}{4}G_m \text{ (as amplifier)} \tag{12.30}$$

$$I_p \text{ (as converter)} = \tfrac{1}{4}I_p \text{ (as amplifier)} \tag{12.31}$$

$$I_2 \text{ (as converter)} = \tfrac{1}{4}I_2 \text{ (as amplifier)} \tag{12.32}$$

and, for pentodes only,

$$R_{\text{eq}} \text{ (as converter)} = 4R_{\text{eq}} \text{ (as amplifier)} \tag{12.33}$$

in which the values "as amplifier" refer to the peak of the local oscillator cycle.[1]

12.10. Noise Induced at Ultra-high Frequencies by Random Emission. At ultra-high frequencies there is a conductive component to the input admittance of a tube. The finite transit time of the electrons makes it possible for the grid to transfer energy to the electrons as they are accelerated.[2] There is a separate component of noise associated with this effect.[3,4] The high-frequency components in the tube noise induce currents on the grid, which in turn influence the electron current, thus giving rise to an extra component of noise. The equivalent input noise conductance is found to be the same as the input conductance, but at approximately five times room temperature. This extra component of noise may be represented as a constant-current generator across the electronic component of grid conductance across the input circuit. To a first order of approximation the induced noise is independent of the normal noise component, which may be added in series with the grid as

---

[1] HARRIS, *op. cit.*

See also articles by HEROLD, *Proc. I.R.E.*, for some typical values.

[2] FERRIS, W. R., Input Resistance of Tubes as Ultra-high Frequency Amplifiers, *Proc. I.R.E.*, vol. 24, pp. 82–107, January, 1936.

[3] NORTH, D. O., and W. R. FERRIS, Fluctuations Induced in Vacuum Tube Grids at High Frequencies, *Proc. I.R.E.*, vol. 29, pp. 49–50, February, 1941.

[4] BALLANTINE, STUART, Schrot Effect in High Frequency Circuits, *Jour. Franklin Inst.*, vol. 206, pp. 159–168 August, 1928.

has been done before. The equivalent circuit for this effect is shown in Fig. 12.10. At ultra-high frequencies there is a component of grid-input conductance due to feedback through the cathode-lead inductance, as well as the component due to electron-transit-time effects. Both components vary as the square of frequency and thus are hard to separate. The feedback component has noise associated with it too, but as a resistor at room temperature.

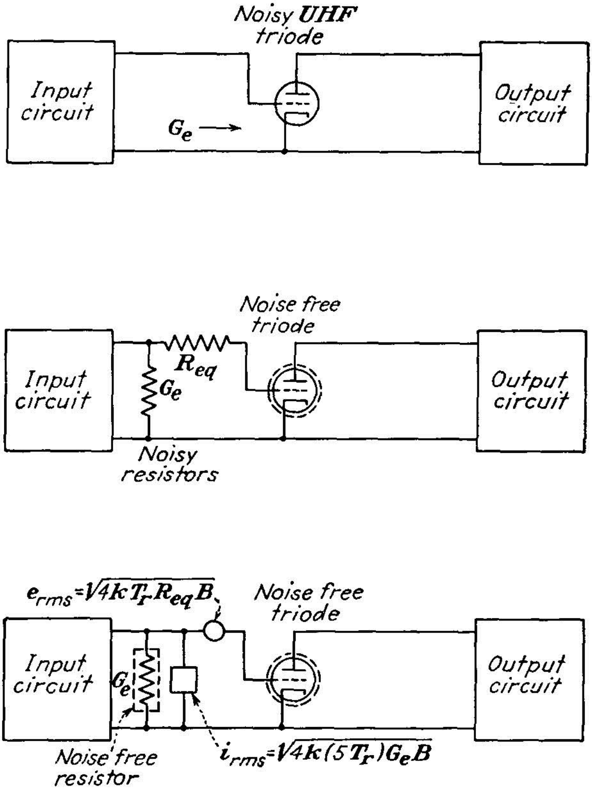


FIG. 12.10.—Equivalent circuit for induced noise in a tube at ultra-high frequencies.

In general, the noise of amplifiers does not change much with feedback. This is because the amplification and input impedance are ordinarily changed by the same factor.

12.11. Noise in Velocity-modulation Tubes. In a velocity-modulation tube a beam of electrons passes through the two grids of a resonator, between which there appears a high shunt resistance. Noise is generated, but there is evidently very little space-charge smoothing action in this case. This is expected from the fact that there is no virtual cathode between the grids of the resonator. As a result, the noise is nearly the full shot noise of the beam, and therefore the noise power delivered to the resonator is given closely by

$$N = 2eI_0BR_{sh} \qquad \text{watts} \tag{12.34}$$

where e is charge of the electron, 1.6020×10^{-19} coulomb
I_0 is beam current, amperes
B is frequency band width, cycles per sec
R_{sh} is shunt resistance of the resonator, ohms

The equivalent circuit is shown in Fig. 12.11. In many cases the electron transit time across the grids is an appreciable fraction of a cycle,

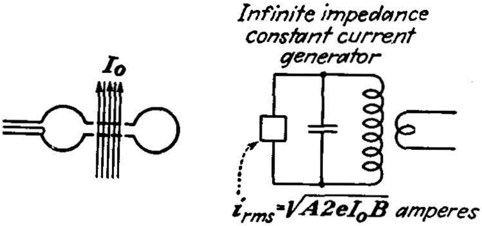


FIG. 12.11.—Noise-equivalent circuit of a velocity-modulated tube.

in which case the transfer of energy from the electrons to the resonator is not perfect and the noise power delivered above is reduced by the factor

$$A = \frac{\sin\left(\frac{\theta}{2}\right)}{\frac{\theta}{2}} \tag{12.35}$$

where θ is the transit angle of the electrons. This expression gives the efficiency of energy transfer between the electrons and the resonator and is developed in Chap. 17. It applies only to tubes with fine grids.

12.12. Noise in Phototubes. High-vacuum phototubes produce the same noise as do hot-cathode tubes with temperature-limited emission. This is true shot effect, giving rise to a mean-square fluctuation current of the value

$$\bar{i}_n{}^2 = 2eI_0B = 3.2040 \times 10^{-19}I_0B \qquad \text{amperes}^2 \tag{12.36}$$

This is the same value of current as is produced by a resistance at room temperature of value

$$R_{eq} = 1/20I_0 \qquad \text{ohms} \tag{12.37}$$

The noise may be represented by a constant-current generator in parallel with the tube considered noise-free. The noise from the tube will often be of the same order as that produced by the large-value resistor used to develop voltage. Thus, if a 10-megohm resistor is used to develop voltage, the noise from the tube will be the same as the noise from the resistor when the current is 0.005 microampere. At higher currents the phototube noise predominates. The equivalent circuit of a typical phototube input circuit is shown in Fig. 12.12.

Gas is sometimes used in phototubes to increase the plate current by cumulative ionization. When this is done, the noise is increased by about the same amount as the signal but the signal-to-noise ratio is in general improved because the contributions of noise from other sources becomes relatively less.

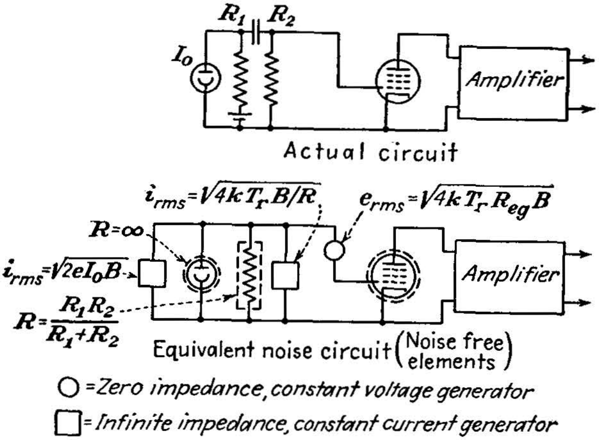


FIG. 12.12.—Noise-equivalent circuit of a phototube input circuit.

12.13. Noise in Secondary-emission Multipliers. Another type of electronic amplifier in which noise may be a consideration is the secondary-emission multiplier. In such a tube there is a series of electrodes at successively higher potentials, each coated with a material that emits a large ratio of secondary to primary electrons. Electrons that strike the first anode give rise to S times as many electrons, which are attracted to the second anode, where they give rise to S times as many as the striking electrons, or S^2 times as many as struck the first anode. After n such impacts the current is S^n times as great as it was originally, S being the ratio of secondary to primary electrons—a number that can be made as high as 9 or 10. It might be thought that with such a system tremendous amplifications could be obtained. True, they can, but no improvement in signal-to-noise ratio can be achieved.

Experimental results indicate that the following laws govern the noise associated with secondary emission in such a device:

1. Secondary emission from any anode follows the shot-effect law, $\bar{\imath}_n^2 = 2eBI_0$, where I_0 is the emitted secondary current.
2. Shot noise from any anode is multiplied by subsequent stages in the same way as the signal is.[1]

Consider the action in a few successive stages of secondary-emission amplification. Let the current from a first anode be I_0; then the mean-square noise current associated with this is

$$\bar{\imath}_n^2 = 2eBI_0 \qquad (n = 0) \tag{12.38}$$

When the direct current I_0 strikes the next diode, it gives rise to a direct current SI_0 and the corresponding mean-square noise current is multiplied by S^2, so that the mean-square noise current associated with the current SI_0 is

$$\bar{\imath}_n^2 = S^2 2eBI_0 + 2eBSI_0 \qquad (n = 1) \tag{12.39}$$

in which the first term is the amplifier noise power from the previous anode and the second is that associated with the liberated secondary current. The above expression is more simply written

$$\bar{\imath}_n^2 = 2eBI_0(S^2 + S) \qquad (n = 2) \tag{12.40}$$

At the next anode the liberated secondary current is S^2I_0, and the mean-square noise current is

$$\bar{\imath}_n^2 = S^2 2eBI_0(S^2 + S) + 2eBS^2I_0 \quad (n = 2) \tag{12.41}$$

which is equal to

$$\bar{\imath}_n^2 = 2eBI_0(S^4 + S^3 + S^2) \qquad (n = 2) \tag{12.42}$$

Extension of this process to n stages yields

$$\bar{\imath}_n^2 = 2eBI_0 \frac{S^n(S^{n+1} - 1)}{S - 1} \qquad (n = n) \tag{12.43}$$

This means that the ratio of output to input noise power is

$$\frac{\text{Noise power out}}{\text{Noise power in}} = \frac{S^n(S^{n+1} - 1)}{S - 1} \tag{12.44}$$

Correspondingly, the ratio of signal power out to signal power in is

$$\frac{\text{Signal power out}}{\text{Signal power in}} = S^{2n} \tag{12.45}$$

---

[1] ZWORYKIN, V. K., G. A. MORTON, and L. MALTER, The Secondary Emission Multiplier, *Proc. I.R.E.*, vol. 24, pp. 351–375, March, 1936.

Upon taking the quotient of the last two equations, the relative change in the ratio of signal to noise power is

$$\frac{\text{Signal-power-to-noise ratio in}}{\text{Signal-power-to-noise ratio out}} = \frac{S^n(S^{n+1} - 1)}{S^{2n}(S - 1)} \qquad (12.46)$$

This ratio is slightly greater than 1 but approaches this value as the secondary-emission ratio S is increased. This simply means that signal and noise are amplified about the same amount in a secondary-emission multiplier, and as a result there is no gain on the signal-to-noise ratio. There is, however, an advantage to using secondary-emission multiplication in that resistor noises are virtually eliminated. Thus a phototube with secondary-emission multiplication has a lower signal-to-noise ratio than a phototube-resistor-amplifier combination at low levels of illumination. A further discussion of this specific case is given in Sec. 19.20.

12.14. Definition of Noise Figure. From all the preceding sections it is seen that there are inherent limitations to electronic devices determined by unavoidable noise. Most electronic devices will, in fact, be noisier than simple theory predicts because of an accumulation of various effects. The smallest amount of noise that an electronic device can possibly exhibit is the available noise power from the thermal agitation of a resistor in the frequency band considered, as given by Eq. (12.6). Usually the noise will be more than this. It is therefore convenient to use as a figure of merit for an electronic device the ratio of the actual noise power at its output terminals to that which it would have if the noise were limited to the minimum noise from thermal agitation. This figure of merit is called the *noise figure* of the device. Basically, the *noise figure is an excess noise ratio.*

A rigorous definition of noise figure involves a consideration of the gain of the device and the available input noise power and the output power. The gain of a device, invariably a four-terminal network, is defined as the ratio of the available signal power at the output to the available signal power at the output of the signal generator,

$$\text{Power gain} = \frac{S_{\text{out}}}{S_{\text{in}}} = G \qquad (12.47)$$

This definition of gain is independent of the output impedance of the device, but it does depend upon the impedance of the signal generator, which is taken as the nominal input impedance of the device. In many applications there is no difficulty in using a signal-generator impedance that is equal to the input impedance of the device under consideration, but in some applications, such as extremely broad-band devices, it is extremely difficult to maintain a constant input impedance. Hence it is logical that the figure of merit should include nonconstancy of the input impedance. The available noise power between two terminals is defined as the noise power that would be absorbed by a matched output circuit. The available input noise power is simply that given from the Johnson noise formula of Eq. (12.6),

$$N_{\text{in}} = kTB \tag{12.48}$$

The noise figure is defined in terms of the factor of most importance in the ultimate sensitivity of an electronic device, the ratio of output signal to noise power. The maximum value this can have is the ratio of available input signal to noise power if there are no other noise sources in the device and if all impedances are properly matched. Some four-terminal networks that consist of passive elements only, say a transformer or a transmission line, have no noise sources present in them, but electronic devices always have some extra sources of noise. *The noise figure F of the device is defined as the ratio of the available signal-to-noise ratio at the signal-generator (input) terminals to the available signal-to-noise ratio at the output terminals*[1,2]

$$F = \frac{\dfrac{S_{\text{in}}}{kTB}}{\dfrac{S_{\text{out}}}{N_{\text{out}}}} \tag{12.49}$$

The noise figure of an electronic device is always greater than unity. The reference temperature is invariably taken as 290°K (63°F).

Some rearrangements of Eq. (12.49) are useful. Upon utilizing the power-gain definition of Eq. (12.47) the noise figure may be written

$$F = \frac{N_{\text{out}}}{GkTB} \tag{12.50}$$

From this the available noise output power is

$$N_{\text{out}} = FGkTB \qquad \text{watts} \tag{12.51}$$

The available output noise due only to noise sources in the network is

$$N_{\text{out}} - GkTB = (F - 1)GkTB \qquad \text{watts} \tag{12.52}$$

Example: It is desired to calculate the noise figure of a 250-ohm-input 30-mc intermediate-frequency amplifier having a band width of 2 mc and using 6AC7

---

[1] Friis, H. T., Noise Figures of Radio Receivers, *Proc. I.R.E.*, vol. 32, pp. 419–422, July, 1944.

[2] North, D. O., Absolute Sensitivity of Radio Receivers, *RCA Rev.*, vol. 6, pp. 332–343, January, 1942.

pentodes. Coupling between stages is provided with two inductively coupled tuned circuits. Let the tube constants be

$$\begin{aligned} I_p &= 10 \times 10^{-3} && \text{ampere} \\ I_2 &= 2.5 \times 10^{-3} && \text{ampere} \\ G_m &= 9 \times 10^{-3} && \text{mho} \end{aligned}$$

Then, from Eq. (12.21), the equivalent noisy resistor in series with the control grid is

$$\begin{aligned} R_{eq} &= \frac{2.5}{9 \times 10^{-3}} \frac{10 \times 10^{-3}}{12.5 \times 10^{-3}} \left(1 + \frac{8 \times 2.5 \times 10^{-3}}{9 \times 10^{-3}}\right) \\ &= 715 \qquad \text{ohms} \end{aligned}$$

From Fig. 12.1 this corresponds to an rms noise voltage of 4.85 microvolts. Since the noise is developed across the input impedance in series with the equivalent noise resistor, while the signal is developed only across the input impedance, the noise figure is

$$F = \frac{715 + 250}{250} = 3.86$$

In cases in which the stage amplification is low the effect of the noise from the following stage must be considered. In this case consider that the coupling network is a unity-ratio impedance transformer with an input impedance of 50,000 ohms. If the second stage is identical to the first in operating characteristics, then there is a noise voltage of 4.85 microvolts developed in series with its grid. This voltage corresponds to a current of 97.0 micromicroamperes through the plate load of the first stage. Upon referring this current back to the control grid of the first stage, it corresponds to a voltage of $\frac{97 \times 10^{-12}}{9 \times 10^{-3}}$, or 0.01075 microvolt. This is small compared with the 4.85 microvolts due to the first tube and so can be neglected in this case.

Noise Figure for Two Networks in Cascade. Usually the sources contributing to the excess noise in an electronic device are principally in the input circuit of the device. However, when the first stage of amplification has insufficient gain, the noise sources effectively located in the input circuit of the second stage of amplification contribute appreciably to the over-all noise as well. Noise from subsequent stages is generally so small compared with the amplified noise from the earlier extra sources that it may be neglected. When the condition cited above is the case, then the following relations hold: Let the first and second stages of amplification be designated by subscripts 1 and 2, respectively. Let over-all characteristics be designated by the subscript 12. The over-all power gain G_{12} is equal to the product of the gains of the first and second stages, G_1G_2. Let the band width be that of the over-all characteristic as defined by Eq. (12.9). The available noise power at the output of the first stage of amplification is

$$N_1 = F_1 G_1 kTB \qquad \text{watts} \tag{12.53}$$

from Eq. (12.51). The noise power in the output of the second stage due only to sources in that stage is, from Eq. (12.52),

$$N_2 - G_2 kTB = (F_2 - 1) G_2 kTB \tag{12.54}$$

The total available noise power at the output of the second stage is G_2 times the quantity in Eq. (12.53) plus the quantity in Eq. (12.54),

$$N_{12} = G_2 F_1 G_1 kTB + (F_2 - 1) G_2 kTB \tag{12.55}$$

or

$$N_{12} = \left[F_1 + \left(\frac{F_2 - 1}{G_1}\right)\right] G_1 G_2 kTB \tag{12.56}$$

Now, upon applying the general formula [Eq. (12.51)] to the over-all situation,

$$N_{12} = F_{12} G_{12} kTB \tag{12.57}$$

The values of available output noise over-all, from Eqs. (12.56) and (12.57), must be equal; therefore

$$F_{12} = F_1 + \frac{F_2 - 1}{G_1} \tag{12.58}$$

This is the important relation that has been sought. It gives the over-all noise figure of two amplifiers in terms of the separate noise figures and the gain of the first amplifier. Needless to say, the expression is not limited to amplifiers but may be applied to mixers and networks in general.

Sometimes noise figures of circuit elements are expressed in terms of equivalent temperatures, simply the room temperature multiplied by the noise figure. The symbol t is often used to designate the ratio of the actual noise to the available thermal-agitation noise of Eq. (12.6). This is most frequently done with passive elements that produce noise, such as crystal detectors. The noise figure of a crystal input receiver is

$$F = L_x(t_x - 1 + F_a) \tag{12.59}$$

where L_x is the conversion loss of the crystal detector corresponding to the reciprocal of the conversion gain of a tube mixer, t_x is the equivalent temperature ratio of the crystal, and F_a is the noise figure of the amplifier into which the output of the crystal detector is fed. If the crystal were noiseless, t_x would equal unity.

The design of electronic equipment with regard to obtaining a low noise figure is a rather complex problem, which will only be touched upon here.[1] The problem is a combination one, involving consideration of circuits and noise sources. It is possible in some cases that the minimum noise figure will be obtained with a condition of mismatched impedances owing to the fact that a mismatch will reduce the noise more than it will the signal. With receivers it is found practical to use r-f amplification before the mixer only up to a certain frequency. This frequency is of the order of 600 mc at this time. Beyond this frequency the noise figure of amplifiers is so great that the signal-to-noise ratio is increased rather than reduced upon amplification. There is also an upper frequency limit at which vacuum-tube mixers are practical. At present, this limit occurs at about 1,000 mc. It is quite possible that these frequency limits will be extended with time. The limits occur because the effective mutual conductance of all tubes decreases with increasing frequency owing to transit-time effects, thus raising the noise. The ultra-high-frequency induced noise effect also contributes to increasing noise.

12.15. Measurement of Noise and Noise Figure. It is not difficult to measure the noise and noise figure of an electronic device when its gain is sufficiently high so that an appreciable output noise power is obtained. Square-law devices such as thermocouples are preferred in noise measurements because the output indication is directly proportional to power. Crystal detectors may also be used at low levels of power. If the rectified crystal current is kept below a few microamperes, the crystal is almost certain to be a square-law device and as such is extremely sensitive. In this case the rectified crystal current is proportional to the input power. If a satisfactory power-output indicating device is used, the noise figure of a device can be measured by simply introducing signal input power until the output indication from the noise alone is doubled. The input power is then equal to that generated by the internal noise sources, and the noise figure is given by

$$F = \frac{S_{\text{in}}}{kTB} \tag{12.60}$$

where S_{in} is the signal power in. For any other adjustment of signal input power the noise figure is given directly from the defining relation of Eq. (12.49), it being remembered that the output-power indication is the sum of the noise and signal power out.

Diodes operating with temperature-limited emission may be used as a standard source of noise. Such diodes preferably have either a tungsten

---

[1] See the survey articles by HEROLD and MALTER, *op. cit.*

or a thoriated tungsten filament. It is difficult to keep the emitted current from an oxide-emitting surface constant under temperature-limited conditions of emission. The noise, of course, is due to shot effect, with the mean-square noise current given by Eq. (12.12). Diodes may successfully be operated as a standard noise source up to 100 mc. Beyond that frequency the impedance transformation introduced by the leads cannot be determined very accurately. Undoubtedly, special diodes can be built for noise measurements at higher frequencies. A standard noise source for measurements of the noise figure of an intermediate-frequency amplifier would consist of the diode in a shielded can with leads brought in through properly by-passed chokes. Across the diode there can be placed a coil that tunes the capacity of the diode to the center of the band of interest. The tuned circuit thus formed should be shunted by the nominal input resistance of the amplifier to be tested. Output leads are then brought from across the tuned circuit to the amplifier under test. Let the nominal input resistance of the amplifier be R; then the noise power delivered to the resistor of value R shunted across the diode is $2eI_0RB$, and the available power into the receiver is one-fourth of this. Let the diode current I_0 be adjusted until the normal noise output power of the receiver is doubled when the standard diode noise is connected to its input. Under these conditions the noise figure, from Eq. (12.60), is given by $\frac{eI_0RB}{2kTB}$; or since $\frac{e}{2kT}$ has the numerical value of 20, the noise figure is given simply by

$$F = 20I_0R \tag{12.61}$$

where I_0 is the diode current in amperes and R is the nominal resistance of the receiver in ohms.

12.16. Typical Tube-noise Values. In the table on the next page are given some typical operating conditions and associated noise values of representative triodes, pentodes, and mixers.

TABLE VI
TUBE-NOISE VALUES*

Type	Application	Voltages			Currents			Trans-conductance, micro-mhos	Noise-equivalent resistance		Noise-equivalent input, voltage** microvolts
		Plate volts	Screen volts	Bias volts	Plate ma	Screen ma	Cathode ma		Calculated ohms	Measured ohms	
6SK7	Pentode amplifier	250	100	−3	9.2	2.4	11.6	2,000	10,500	9,400–11,500	0.94
6SJ7	Pentode amplifier	250	100	−3	3	0.8	3.8	1,650	5,800	5,800	0.70
6SG7	Pentode amplifier	250	125	−1	11.8	4.4	16.2	4,700	3,300	....	0.53
6AC7/1852	Pentode amplifier	300	150	−2	10	2.5	12.5	9,000	720	600–760	0.25
956	Pentode amplifier	250	100	−3	5.5	1.8	7.3	1,800	9,400	....	0.90
1T4	Pentode amplifier	90	45	0	2.0	0.65	2.65	750	20,000	....	1.3
6SA7	Frequency converter	250	100	0	3.4	8.0	11.9	450§	240,000	210,000	4.5
6K8	Frequency converter	250	100	−3	2.5	6.0	8.5‡	350§	290,000	....	4.9
1R5	Frequency converter	90	45	0	0.8	1.8	2.75	250§	170,000	....	3.8
6L7	Pentagrid mixer	250	100	−3	2.4	7.1	9.5	375§	255,000	210,000	4.6
6J5	Triode amplifier	250	...	−8	9.0	...	.....	2,600	960	1,250	0.28
955	Triode amplifier	180	...	−5	4.5	...	.....	2,000	1,250	....	0.32
6AC7/1852	Triode amplifier	150	150	−2	...	...	12.5	11,200	220	200	0.14
6AC7/1852	Pentode mixer	300	150	−1†	5.2	1.3	6.5	3,400§	2,750	3,000	0.48
6SG7	Pentode mixer	250	125	−1†	3.0	1.1	4.1	1,180†	13,000	....	1.0
956	Pentode mixer	250	100	−1†	2.3	0.8	3.1	650†	33,000	....	1.7
6J5	Triode mixer	100	...	−1†	2.1	...	.....	620§	6,500	....	0.74
6AC7/1852	Triode mixer	150	150	−1†	...	...	6.5	4,200§	950	....	0.28
955	Triode mixer	150	...	−1†	2.8	...	.....	660§	6,100	....	0.72

* Reproduced from Harris, W. A., Space Charge Limited Current Fluctuations in Vacuum Tube Amplifiers and Input Systems, *RCA Rev.*, vol. 5, pp. 505–524, 6, pp. 114–124, April, July, 1941.

† At peak of oscillator cycle. ‡ Hexode section only. Triode section takes its current from a separate part of the cathode.

§ Conversion transconductance value. ** For effective bandwith of 5,000 cycles.

CHAPTER 13

ELECTROSTATIC ELECTRON OPTICS

13.1. Introduction. The term "electron optics" as applied to the behavior of electrons under the influence of electric and magnetic fields has been in use for some time. As the term implies, there is a close analogy between the behavior of light rays and electron beams, particularly when the fields through which the electron moves are purely electrostatic. Electrons move through an electric field just as do light rays through a medium of continuously variable index of refraction. Electrons can be reflected, refracted, and focused very much as can light rays.

Electron optics is a relatively new field of science, but already its study has led to the development of the cathode-ray tube, the high-intensity kinescope, the image-dissector tube, the iconoscope, the orthicon, the various forms of electron multiplier tubes, the electron microscope, and many other devices. The groundwork for the new science was laid more than a hundred years ago by Lagrange, Maupertius, and Hamilton, who recognized that the principle of least action as applied to particles was strictly analogous to the Fermat principle of least time, which holds for light rays. The modern phase of the subject was ushered in by Busch, who showed in 1926 that the action of a short axially symmetrical magnetic field on electron beams was similar to that of a glass lens on light rays. The science was given a firm foundation by the early workers in the field, among whom Davisson, Calbick, Brueche, Glaser, Knoll, Ruska, and Scherzer were outstanding. At this writing, the total literature includes hundreds of technical articles, and already a number of books completely devoted to the subject have been written.[1-5] Because of the extensive nature of the subject, it cannot

---

[1] Brueche, E., and O. Scherzer, "Geometrische Elektronenoptik," Springer, Berlin, 1934.

[2] Maloff, I. G., and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.

[3] Myers, L. M., "Electron Optics," Van Nostrand, New York, 1939. Contains excellent bibliography complete to 1939.

[4] Klemperer, O., "Electron Optics," Cambridge, London, 1939.

For other, more compact summaries see Zworykin, V. K., and G. A. Morton, "Television," McGraw-Hill, New York, 1940, and Gray, F., Electrostatic Electron be hoped that the present chapter will be more than an abstract of the most important aspects of the subject.

It has already been indicated that the electrostatic field between the wires of a grid constitutes a cylindrical lens which is capable of focusing electrons. A more useful type of lens is produced by any axially symmetric field, whether electric or magnetic. An example is the electrostatic field about a circular aperture. All the laws that exist for the lenses of physical optics apply as well to the lenses of electron optics. An analogy can be developed between the quantities of geometrical optics and the corresponding quantities of electron optics. In the treatment given here, the laws of electron optics will be developed from the mechanics of electron motion, and the analogy with those of geometrical optics will then be shown.

Snell's Law. The basic law of geometrical optics is Snell's law of refraction, from which all the properties of physical lenses can be deduced. This law has its exact counterpart in electron optics. Snell's law for optics is

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{13.1}$$

where n_1 and n_2 are the indices of refraction on two sides of a plane boundary and θ_1 and θ_2 are the angles of incidence and refraction of a light ray as measured from a normal to the boundary. The corresponding situation for electron optics is shown in Fig. 13.1. This shows the behavior of an electron moving in a region with a uniform potential V_1 and suddenly crossing into a region with a uniform potential V_2. This is approximately the situation that exists at the junction of the D's of a cyclotron, except that the region in which the potential changes from one value to the other has a small but finite dimension. In going from the region of one potential to the other, the component of velocity normal to the boundary is increased if the potential is increased, but the tangential component of velocity is unchanged. Equating the initial and final tangential components of velocity,

$$v_1 \sin \theta_1 = v_2 \sin \theta_2 \tag{13.2}$$

where $v = 5.93 \times 10^5 \sqrt{V}$ meters per sec when the potential is given in volts. Comparing Eqs. (13.1) and (13.2) *it is seen that the quantity in electron optics corresponding to index of refraction is electron velocity.*

Optics, *Bell Sys. Tech. Jour.*, vol. 18, pp. 1–31, January, 1939; *Zeit. für Tech. Phys.*, vol. 17 (No. 12), 1936.

[5] ZWORYKIN, V. K., and others, "Electron Optics and the Electron Microscope," McGraw-Hill, New York, 1946.

This in turn is proportional to the square root of potential if the electron starts from rest at a point of zero potential.

The Principle of Least Action. A further correspondence between electron and geometrical optics lies in the principle of least time and the principle of least action. The principle of least time states that a light ray will assume a path such that the time between any two points of its path will be a minimum compared with that for all other possible paths between the same two points. Thus

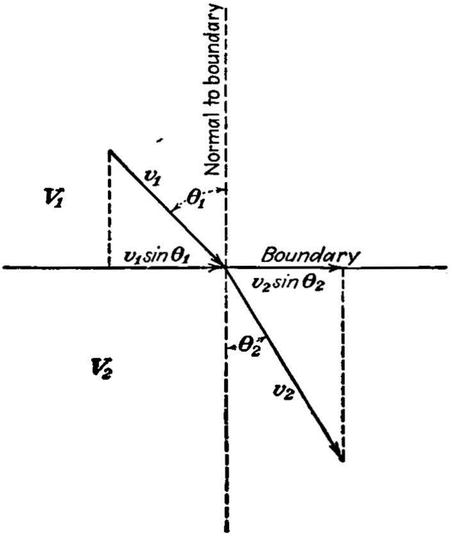


FIG. 13.1.—Electron refraction.

$$T = \int \frac{ds}{v} = \frac{1}{c} \int n\, ds = \text{min} \tag{13.3}$$

where s is distance, T is time, v is the velocity of light in a medium of index of refraction n, and c is the velocity of light in vacuum. In particle dynamics the corresponding law is that the integral of momentum with distance assumes a minimum value. The integral of momentum with distance is defined as action. The principle states that

$$\text{Action} = \int mv\, ds = \text{min} \tag{13.4}$$

The correspondence between the two principles is quite evident. Again it is seen that the counterpart of index of refraction is electron velocity. Note, however, that light velocity *does not* correspond to electron velocity.

Simple Lenses. The lenses encountered in electron optics are generally of a more complex type than the simple lenses of geometrical optics.

In general, it is more difficult to analyze and to represent their characteristics, on three distinct counts, as follows:

1. In light lenses a discrete number of refractions occur at surfaces between materials of different indices of refraction, whereas in electron lenses the refraction occurs continuously through the equivalent of a material of variable index of refraction.
2. The thin lenses of geometrical optics, *i.e.*, lenses whose axial dimension is short compared with their focal length, are often operated in air, *i.e.*, the light rays start and finish in a medium with the same index of refraction, with the result that the lens characteristic can be expressed in terms of a single parameter, the focal length. Electrostatic electron lenses on the other hand more often have initial and final potentials that are different, so that the equivalent initial and final indices of refraction are different, with the result that it takes two focal lengths, one for each direction, to describe the lens.
3. The lenses of electron optics are usually thick lenses, *i.e.*, the axial dimension of the lens is not short compared with the focal length. In such a lens it is not correct to measure the focal length from the center of the lens; rather, it must be measured from a reference plane known as the "principal plane," which may be outside the lens. This introduces another parameter for the thick lens. If in addition the initial and final potentials are different, four parameters are required to describe the lens: a focal length and a reference plane for each direction.

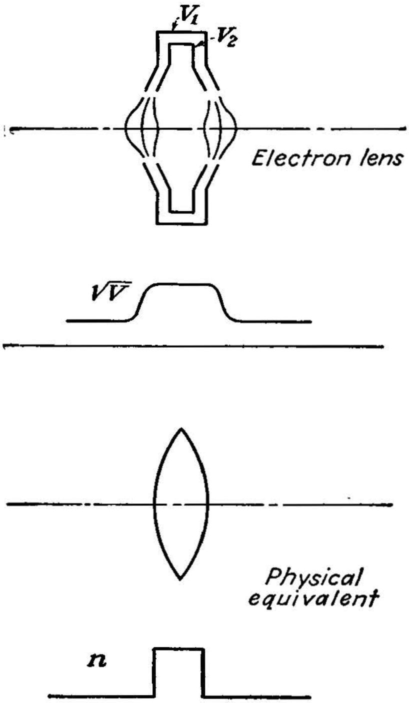


FIG. 13.2.—Electron lens equivalent of a thin physical lens.

Although the electron-optical equivalent of thin light lenses is seldom used, it is interesting for a first consideration to describe the equivalent of such thin lenses. In Fig. 13.2 is shown a simple, thin, convex lens and its approximate electrostatic equivalent. This latter consists of two concentric electrodes of revolution, as shown, with the inner electrode at a higher potential than the outer. The corresponding axial variation of the index of refraction is shown for both cases. The focusing property of the lenses is derived from the combination of the variation of the index of refraction and the curvature of the bounding surface. In the case of the light lens the bounding surface is sharply defined, while for the electrostatic lens it is not so sharply defined. Both lenses have a convergent action. In the case of the electrostatic lens the convergent action results because the radial component of the gradient of potential pushes the electron toward the axis on both sides of the lens. The electron path shown in the figure may be used to define the term "focal length." An electron entering the lens parallel to the axis is deflected toward the axis while passing through the lens and emerges headed toward it. The subsequent path is a straight line because the electron is in a field-free region after passing through the lens. If the initial and final straight-line portions of the path be extended so that they intersect in the lens, then the axial distance between the plane of this intersection and the plane at which the electron crosses the axis is known as the *focal length.* The point at which the electron crosses the axis is known as the *focal point,* and the plane through this point normal to the axis is known as the *focal plane.* For the lenses of Fig. 13.2 the focal lengths in the two directions are the same since the initial and final indices of refraction are the same. Further, any electron ray entering the lens parallel to the one shown will cross the axis at the focal point in the absence of aberrations.

The case in which the initial and final indices of refraction are not equal seldom occurs in geometrical optics, but it is the most common case in electron optics. A fictitious example to illustrate the light case may be assumed to consist of a thin, convex glass lens in the side of a tank of oil, so that the light rays start in air and end in oil. (It is further assumed that the index of refraction of glass lies between those of oil and air.) This situation is shown in Fig. 13.3 along with the equivalent electrostatic lens. The equivalent electron lens in this case consists of circular apertures in two parallel-plane conductors maintained at different potentials. It is interesting to note that the focusing action of the electron lens is derived from the curvature of the equipotential surfaces shown rather than from the shape of the electrodes—this is usually the case. Also shown are the electron and light rays entering the lenses from both directions parallel to the axis. These rays are known as the *principal rays* of the lens, and their intersection with the axis defines the focal lengths as indicated above. The ray passing from right to left is known as the "first principal ray," while that passing from left to right is known as the "second principal ray." In this case the focal lengths in the two directions are not equal. It will be shown later that the two focal lengths are in the ratio of the initial and final indices of refraction, the larger index of refraction being associated with the longer focal length.

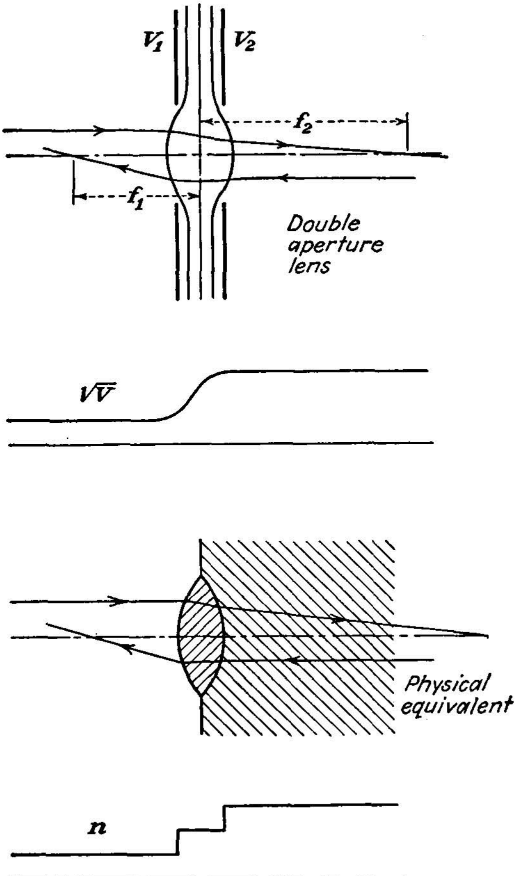


FIG. 13.3.—Double-aperture lens and physical equivalent.

The lens action in the case of Fig. 13.3 is more subtle than it appears at first glance. Referring to the light lens and considering a ray passing through it from left to right, it is seen that the action at the left face of the lens is convergent while that at the right face of the lens is divergent. The net action for the assumed indices of refraction is, however, always convergent, as will presently be shown. The ray passing from right to left, the first principal ray, experiences first a divergent and then a convergent action. The same type of action occurs in the electrostatic lens except that the electron path is smoothly curved instead of consisting of straight-line segments. The second principal ray, originating at the left, experiences a convergent action in passing through the left part of the lens because the gradient of potential has a radial component that is directed toward the axis. In passing through the right part of the lens the second principal ray experiences a divergent action because the gradient of potential has a radial component that is directed outward from the axis. It may be noticed that when the electron is traveling in the direction of increasing potential the curvature of the equipotential surfaces corresponds to the curvature of the equivalent optical-lens system. Thus, when the equipotential surface as approached by the electron is convex, the action is convergent and the equivalent physical lens surface is also convex. When the equipotential surface approached by the electron moving in the direction of increasing potential is concave, then the equivalent optical surface is also concave and the lens action is divergent. When the electron is moving in the direction of decreasing

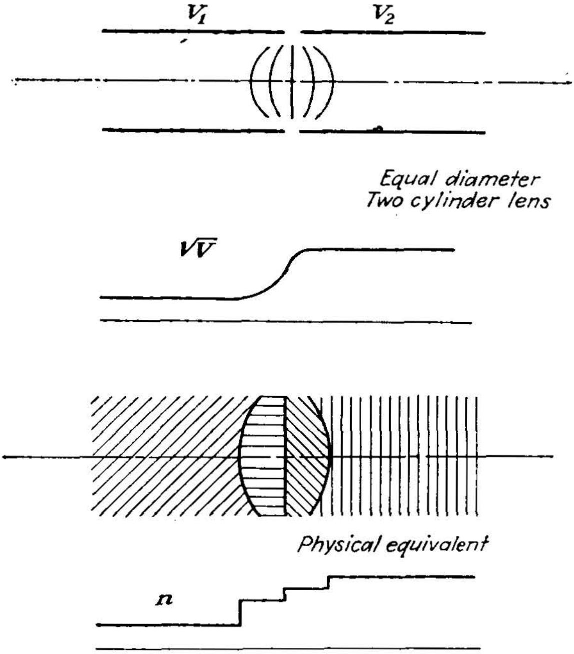


FIG. 13.4.—Two-cylinder lens and physical equivalent.

potential, then a convex equipotential corresponds to a divergent action and a concave equipotential corresponds to a convergent action (as on the right face of the lens of Fig. 13.2).

It remains only to consider the case of a thick lens. Such a lens is produced by the field of two equal-diameter coaxial cylinders at different potentials as shown in Fig. 13.4. Such a lens is characterized by having a long region in which the potential variation occurs. The corresponding lens dimensions will generally not be short compared with the focal length. Also shown in Fig. 13.4 is the equivalent physical lens. Principal rays for this case have the form shown later in Fig. 13.20. The action here is very similar to that in the preceding example except that the region in which the lens action occurs is longer. In all the examples given the electron-lens action has been shown for only one set of potentials. If the electrode-potential ratio is increased, then the difference between the initial and final index of refraction is increased and the lens becomes stronger and the focal lengths become shorter. Thus, in effect, every electron lens corresponds to a whole set of physical lenses, one for every possible voltage ratio. This property makes the electron lens a much more versatile instrument than the physical lens because the lens strength can be changed by simply changing electrode potentials instead of having to move lens components relative to one another.

Lens Formulas. For the simple thin lens of Fig. 13.2 having the same initial and final index of refraction the formula relating distance from the lens to object and image and the focal length is

$$\frac{-1}{l_1} + \frac{1}{l_2} = \frac{1}{f} \tag{13.5}$$

where l_1 = distance from lens center to object
l_2 = distance from lens center to image
f = focal length

and the minus sign occurs because l_1 is measured to the left from the lens center. This is well known to photographers and students of physical optics and will not be proved here. The general lens formula, of which this is a special case, will be developed for electron lenses from the differential equation of the electron paths. The geometry of the arrangement is shown in Fig. 13.5.

For the thin lens of Fig. 13.3, which operates between two different indices of refraction, there are two focal lengths, and the lens formula is given by

$$\frac{f_1}{l_1} + \frac{f_2}{l_2} = 1 \tag{13.6}$$

where f_1 is the so-called "first focal length" associated with a ray entering the lens parallel to the axis from the right and f_2 is the "second foca distance" associated with a ray entering the lens parallel to the axis from the left, object distance is l_1, and image distance is l_2. The geometrical relations for this lens are shown in Fig. 13.6. Note that a ray passing through the center of the lens does not make equal angles with the axis before and after passage in this case. Note, however, that the principal rays can still be used to construct an image; in fact, principal rays can so be used to construct the image in general.

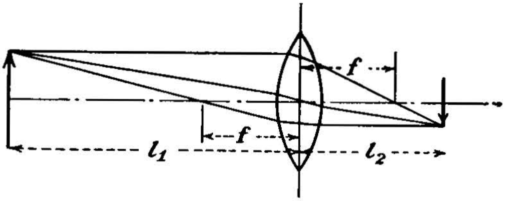


FIG. 13.5.—Focal relations in a thin lens.

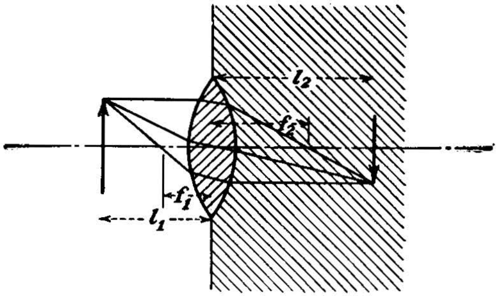


FIG. 13.6.—Focal relations in a lens operating between two different media.

For the thick lens of Fig. 13.4 the situation is somewhat more complicated. The lens formula in this case has the form

$$\frac{f_1}{l_1 - P_1} + \frac{f_2}{l_2 - P_2} = 1 \tag{13.7}$$

where the symbols have the significance shown in Fig. 13.7. Distances measured to the right are positive, to the left negative. The focal lengths f_1 and f_2 are measured from reference planes H_1 and H_2, which are designated as first and second principal planes, respectively. These are located at the intersection of the extension of the initial and final straight-line portions of the respective principal rays. The distances P_1 and P_2 measure the distance from the lens center to the principal planes in the direction of the principal rays. The distance P_2 is negative in this case. The focal lengths f_1 and f_2 are measured from the principal planes to the intersection of the principal ray with the axis, f_1 being negative. These intersections of the principal rays with the axis are known as "focal points" and are located at a distance F_1 and F_2 from the lens center. The significance of Eq. (13.7) is that the object and image distances must be measured with respect to the principal planes rather than with respect to the lens center. In effect, the behavior of a thick lens is the same as if the space between the principal planes did not exist, making them coincident, and a thin lens were located at the plane of coincidence.

The method of constructing an image from an object is evident from Fig. 13.7. To find the point on the image corresponding to any point on the object, draw a first principal ray through that point on the object and through the first focal point until it intersects the first principal plane. From this point of intersection draw a ray parallel to the axis extending to the right. Through the same point of the object draw a second principal ray parallel to the axis, and extend it until it intersects the second principal plane. From this point of intersection draw a ray through the second focal point, and extend it until it intersects the ray first constructed. The intersection of the two principal rays gives the point on the image corresponding to the point on the object. Thus, if the two focal lengths and the location of the two principal planes of a thick lens are known, the image corresponding to any object can easily be constructed.

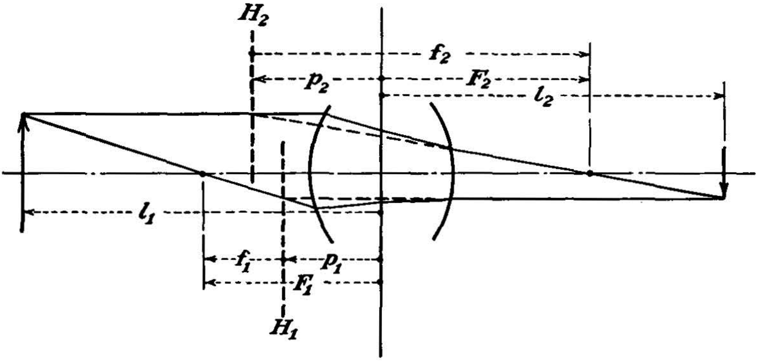


FIG. 13.7.—Thick-lens terminology.

13.2. Electrostatic-lens Fields. The analytical treatment of electron lenses has not been very completely developed. Ideally, it would be possible to obtain expressions for the potential fields associated with any given set of electrodes and then solve for the path of an electron through this field. Actually, the fields of electron lenses are not simple of determination, and the solution of electron trajectories through them is even less so. However, by studying all aspects of lens fields and electron paths it is possible to accumulate enough fragments of information about electron lenses so that the whole picture can be pieced together rather well. The sum total of information that can be gathered is still small enough so that frequent recourse is had to model determination of fields and numerical solution of path equations. Even this procedure has its limitations and in the end gives way to experimental determination of lens characteristics. Nevertheless, no complete understanding of electrostatic lenses is possible without a fairly complete examination of the nature of the fields of electron lenses.

Virtually all the electron-lens fields are two-dimensional fields having a symmetry about an axis of rotation. Cylindrical coordinates are best suited for describing such fields, radial distance being indicated by the symbol r and axial distance by the symbol z. Because of the symmetry of rotation the angular coordinate θ is not involved. Laplace's equation in the above two-dimensional cylindrical coordinates takes the form

$$\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r\,\partial V}{\partial r}\right) + \frac{\partial^2 V}{\partial z^2} = 0 \tag{13.8}$$

All expressions for fields of rotational symmetry must be solutions of this equation.

General Form of Fields with Rotational Symmetry. One solution of Laplace's equation as given above is

$$V(r,z) = \sum_{n=1}^{\infty} (a_n\epsilon^{k_n z} + b_n\epsilon^{-k_n z})J_0(k_n r) \tag{13.9}$$

where the k's are values of the separation constant encountered in solving the Laplace equation, the a's and b's result from fitting the potential to the electrodes, and J_0 is the zero-order Bessel function of the first kind. The second kind of Bessel function, $N_0(k_n r)$, does not appear because the potential along the axis is finite. The k's can be either real or imaginary. If imaginary values of k are used, then an integral form of the expression for the potential field may be written

$$V(r,z) = \int_0^\infty [A(k)\cos kz + B(k)\sin kz]J_0(ikr)\,dk \tag{13.10}$$

where A and B are functions of k that are determined from the shape of the electrodes. The function $J_0(ikr)$ has real values and is something like the function ϵ^r. The parameter k disappears in the integration.

Probably the most useful form of the solution of Laplace's equation in cylindrical coordinates is obtained by expressing the potential as a power series in r. This is done by assuming that the solution is of the form

$$V(r,z) = \sum_{n=0,1,2,\cdot}^{\infty} a_n(z)r^n \tag{13.11}$$

If this expression is substituted in Eq. (13.8), the values of the coefficients a_n, functions of z, may be determined and the series is thus established. Consider the step-by-step operations upon the nth term of the series in this determination.

$$\frac{\partial V}{\partial r} = na_n(z)r^{n-1} \tag{13.12}$$

$$r\frac{\partial V}{\partial r} = na_n(z)r^n \tag{13.13}$$

$$\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) = n^2a_n(z)r^{n-1} \tag{13.14}$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) = n^2a_n(z)r^{n-2} \tag{13.15}$$

The corresponding term for the nth power of r is $(n + 2)^2a_{n+2}r^n$. The other term of Laplace's equation yields

$$\frac{\partial^2 V}{\partial z^2} = a_n''(z)r^n \tag{13.16}$$

where the primes indicate derivatives with respect to z. Adding terms involving the nth power of r, a process that takes care of all powers because Eq. (13.8) is an identity,

$$(n + 2)^2a_{n+2}r^n + a_n''(z)r^n = 0 \tag{13.17a}$$

which gives

$$a_{n+2} = -\frac{a_n''(z)}{(n + 2)^2} \tag{13.17b}$$

This is a recurrence formula that gives the coefficient of any power of r in terms of the coefficient of the second term proceeding.

The symmetry of the field about the axis allows only even powers of r since values of the field for any positive and negative value of r must be the same. This requires that

$$a_1 = a_3 = a_5 = \cdot\cdot\cdot = 0 \tag{13.18}$$

With this restriction the entire series can be expressed in terms of the coefficient a_0.

$$V(r,z) = a_0r^0 - \frac{a_0''(z)r^2}{2^2} + \frac{a_0^{(4)}(z)r^4}{2^2 \cdot 4^2} + \cdots + \frac{(-1)^n}{(n!)^2}\left(\frac{r}{2}\right)^{2n} a_0^{(2n)}(z) + \cdots \qquad (13.19)$$

where the superscripts in parentheses indicate the order of the derivative and n is no longer the n of Eqs. (13.11) to (13.17) but assumes integral values as before. If now the value of a_0 can be determined in terms of the potential, the series will be given in its simplest form. The value of a_0 is fixed by the fact that, when $r = 0$,

$$V(0,z) = a_0(z) = V_0(z) \qquad (13.20)$$

In other words, a_0 is the value of the potential along the axis. This axial potential will be denoted by the function $V_0(z)$ hereafter, to simplify the notation and to indicate that it is a function of a single variable. The subscript zero will further serve as a constant reminder that the potential along the axis only is involved. Expressing the coefficients of Eq. (13.19) in terms of the axial potential,

$$V(r,z) = V_0(z) - \frac{V_0''(z)r^2}{2^2} + \frac{V_0^{(4)}(z)r^4}{2^2 \cdot 4^2} + \cdots + \frac{(-1)^n V_0^{(2n)}(z)}{(n!)^2}\left(\frac{r}{2}\right)^{2n} + \cdots \qquad (13.21)$$

This is the expression that has been sought. It is one of the most useful and most extensively used relations in electrostatic electron optics. The significance of this expression is that if the variation of potential along the axis of a field of rotational symmetry is known then the potential at any point in the field can be calculated. It follows that if the axial variation of potential is known then the derivatives of the axial potential with axial distance are determined. The derivatives can always be determined numerically or graphically if not analytically. In fact, the axial potential need not be and frequently is not capable of analytical expression.[1]

---

[1] Scherzer has given another expression by which the potential at any point in a field of rotational symmetry may be determined from the axial potential. The value of $V(r,z)$ is given by the real part of the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi} V_0(z + ir\sin\alpha)\,d\alpha$, in which the expression in the integral is the axial potential function of the argument $(z + ir\sin\alpha)$, α being a parameter that disappears upon integration. This expression converts to the series of Eq. (13.21) upon series expansion and term-by-term integration. It is of somewhat limited use because it generally requires that the axial potential be capable of analytical expression.

See SCHERZER, O., Zur Theorie der Elektronenoptischen Linsen Fehler, *Zeit. für Phys.*, vol. 8, pp. 183–202, January, 1933.

Many important properties of rotational fields can be deduced from the series of Eq. (13.21). Let the series be expanded further in terms of $z_1 = z - z_0$ in the vicinity of z_0. Then

$$V_0(z_1) = V_0(z_0) + V_0'(z_0)z_1 + \tfrac{1}{2}V_0''(z_0)z_1^2 + \cdots \qquad (13.22)$$

by Maclaurin series. The corresponding expression for potential, by Eq. (13.21), becomes

$$V(r,z) = V_0(z_0) + z_1V_0'(z_0) + \frac{1}{2}z_1^2V_0''(z_0) - \frac{r^2V_0''(z_0)}{4} + \cdots \qquad (13.23)$$

The equipotential $V(r,z) = V_0(z_0)$ in the vicinity of the axis reduces to the hyperbola

$$\frac{r^2V_0''}{4} = z_1V_0' + \frac{z_1^2V_0''}{2} \qquad (13.24)$$

Upon applying Eq. (5.26), the radius of curvature of an equipotential at the axis is found to be

$$R = \frac{2V_0'}{V_0''} \qquad (13.25)$$

The radius of curvature generally assumes its smallest value when the second derivative of the axial potential is greatest.

At a saddle point of potential as shown in Fig. 13.14 at the aperture center $V_0' = 0$, the radius of curvature tends to become zero, and the equipotentials are straight lines intersecting and forming a branch point at the axis. From Eq. (13.24), for the conditions stated it is seen that

$$\tan^2\gamma = \frac{r^2}{z_1^2} = 2 \qquad \gamma = 54°44' \qquad (13.26)$$

where γ is the angle between one of the equipotential branch lines and the axis. Equipotential lines at a saddle point will always intersect the axis as straight lines, making an angle of 54°44′ with it.[1]

---

[1] It is of interest to record the properties of two-dimensional fields expressible in the rectangular coordinates x and y and having no variation in the z direction. Let the x axis coincide with a line of symmetry; then

$$V(x,y) = V_0(x) - \frac{y^2V_0''(x)}{2!} + \frac{y^4V_0^{(4)}(x)}{4!} + \cdots + \frac{(-1)^ny^{2n}V_0^{(2n)}(x)}{(2n)!} + \cdots$$

The radius of curvature of an equipotential at a point along the line of symmetry is given by $R = \dfrac{V_0'}{V_0''}$. At a saddle point on a line of symmetry the equipotentials are straight lines making an angle of 90 deg with each other and 45 deg with the axis. These relations apply in cases such as the line of symmetry midway between the grid wires of an ideal plane triode.

The Equal-diameter Two-cylinder Lens. The equal-diameter two-cylinder lens is very extensively used in electron optics. The field of such a lens is shown in Fig. 13.8. Here are shown the equipotential lines within the coaxial cylinders. All the equipotential lines pass through the gap between the two cylinders. They also all intersect the axis at right angles. The plot is further seen to be symmetrical about the axis and about the midplane. The shape of the field is nearly independent of the gap spacing, provided that this is small.

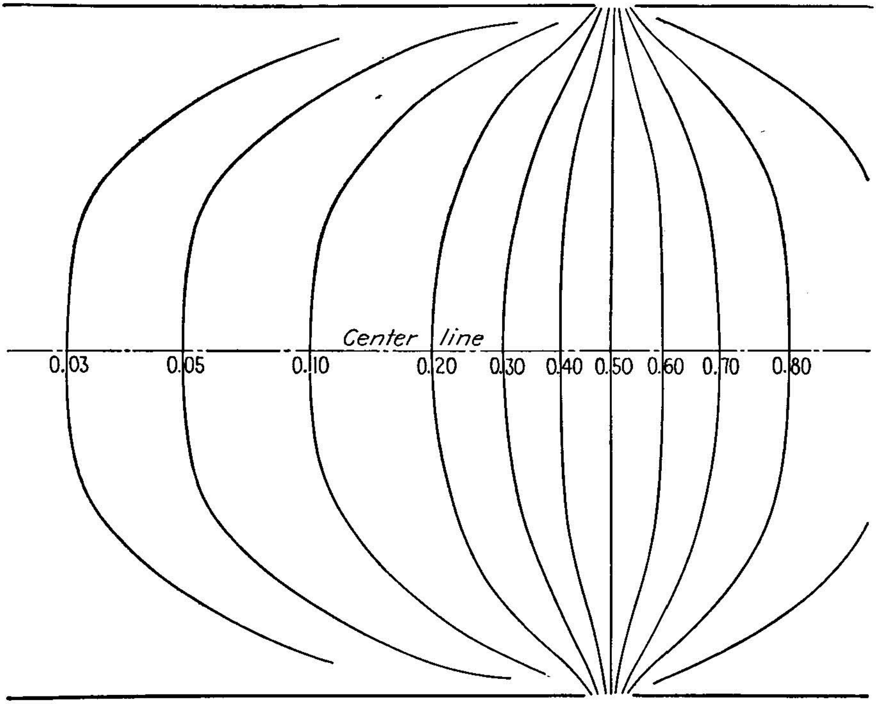


FIG. 13.8.—Field plot of an equal-diameter two-cylinder lens.

The axial-potential distribution of the equal-diameter two-cylinder lens with small gap spacing has been found to be

$$V_0(z) = V_1 \tanh \frac{1.32z}{R} \tag{13.27}$$

when the two cylinders have potentials of $-V_1$ and $+V_1$, respectively, and R is the radius of the cylinders.[1,2] If the cylinder potentials are not

---

[1] GRAY, *op. cit.*, p. 25.

[2] BERTRAM, S., Determination of Axial Potential Distribution in Axially Symmetric Fields, *Proc. I.R.E.*, vol. 28, pp. 418–421, September, 1940. See also BERTRAM, S., Calculation of Axially Symmetric Fields, *Jour. App. Phys.*, vol. 13, pp. 496–502, August, 1942, for general material on this subject.

equal, then the axial potential takes the form

$$V_0(z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{2} \tanh \frac{1.32z}{R} \tag{13.28}$$

where V_1 is the first-cylinder potential and V_2 is the second-cylinder potential. The derivatives of the axial potential are readily found to be

$$V_0'(z) = \frac{1.32}{R}\left(\frac{V_2 - V_1}{2}\right)\frac{1}{\cosh^2\left(\frac{1.32z}{R}\right)} \tag{13.29}$$

and

$$V_0''(z) = \frac{-3.48}{R^2}\,\frac{V_2 - V_1}{2}\,\frac{\tanh\left(\frac{1.32z}{R}\right)}{\cosh^2\left(\frac{1.32z}{R}\right)} \tag{13.30}$$

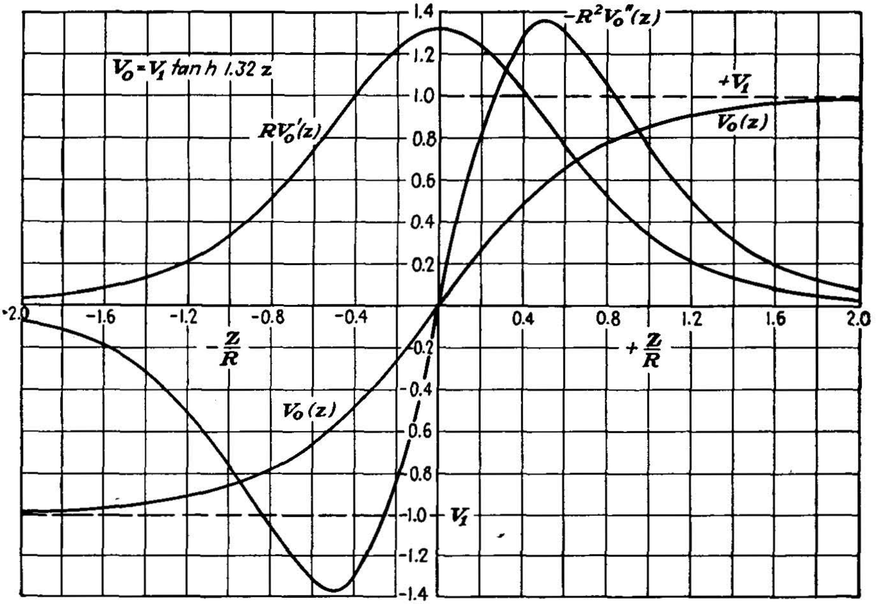


FIG. 13.9.—Axial potential and its derivatives—equal-diameter two-cylinder lens.

The axial potential and its first two derivatives for the case in which $V_1 = -V_2$ and $R = 1$ are plotted in Fig. 13.9. Examination of the curve for the axial potential shows that, at a distance of one radius from the midplane, the potential is within 8 per cent of its final value. At a distance of one diameter from the midplane, the potential is within 1 per cent of its final value. The entire region of variation of potential is therefore virtually confined to a region within one diameter of the midplane.

Equal-diameter Spaced Cylinders. Another electrostatic lens frequently encountered is that of two coaxial equal-diameter cylinders spaced an appreciable distance. A simple approximate formula for the axial potential in this case takes the form

$$V_0(z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\dfrac{2.64s}{R}} \ln \left\{ \frac{\cosh\left(\dfrac{1.32z}{R}\right)}{\cosh\left[\dfrac{1.32(z - s)}{R}\right]} \right\} \tag{13.31}$$

where s is the axial spacing between the cylinders.[1] This expression reduces properly to Eq. (13.28) for $s = 0$. The expression given was derived on the assumption that the potential variation between the two cylinders at a radial distance equal to the cylinder radius is linear. This is a moderately good approximation, but not exact.

For this same case, an empirical approximation to measured axial-potential distributions takes the form

$$V_0(z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{b} \int_0^z \epsilon^{-\frac{\pi}{b^2}z^2}\, dz \tag{13.32}$$

where b is an experimental parameter equal to the reciprocal of the slope of the potential curve at $z = 0$ and having the value[2]

$$b = 2R \left[0.73 + 0.53 \left(\frac{s}{2R} \right)^2 \right] \tag{13.33}$$

For large values of z the value of axial potential assumes the correct value of the electrode potential by virtue of the fact that the integral assumes the value $\frac{b}{2}$ for $z = \pm\infty$. The formula is also approximately correct for small values of z as may be seen by setting $s = 0$ and expanding the integral in series. The first few terms give

$$V_0(z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{2} \left[1.37 \frac{z}{R} - 0.673 \left(\frac{z}{R} \right)^3 + \cdots \right] \tag{13.34}$$

whereas the expression of Eq. (13.28) involving the hyperbolic tangent gives

$$V_0(z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{2} \left[1.32 \frac{z}{R} - 0.790 \left(\frac{z}{R} \right)^3 + \cdots \right] \tag{13.35}$$

---

[1] BERTRAM, S., Determination of Axial Potential Distribution in Axially Symmetric Fields, *Proc. I.R.E.*, vol. 28, p. 420, September, 1940.

[2] KIRKPATRICK, PAUL, and J. G. BECKERLY, Ion Optics of Equal Coaxial Cylinders, *Rev. Sci. Instr.*, vol. 7, pp. 24–26, January, 1936.

The slope of the empirical function thus agrees within 5 per cent with the correct value for $z = 0$. The general formula is probably well within 10 per cent as long as the cylinder spacing is less than 1.75 cylinder diameters. The integral form of Eq. (13.32) is very convenient for some lens calculations because of the fact that the function is readily differentiated and integrated and because numerical evaluations of the functions involved are extensively tabulated. The formula given applies strictly to the case of electrodes that have toroidal corona rings attached to the edge of the cylinders that are tangent to the cylinder edges at their outside diameter and have a radius one-tenth of the cylinder radius.

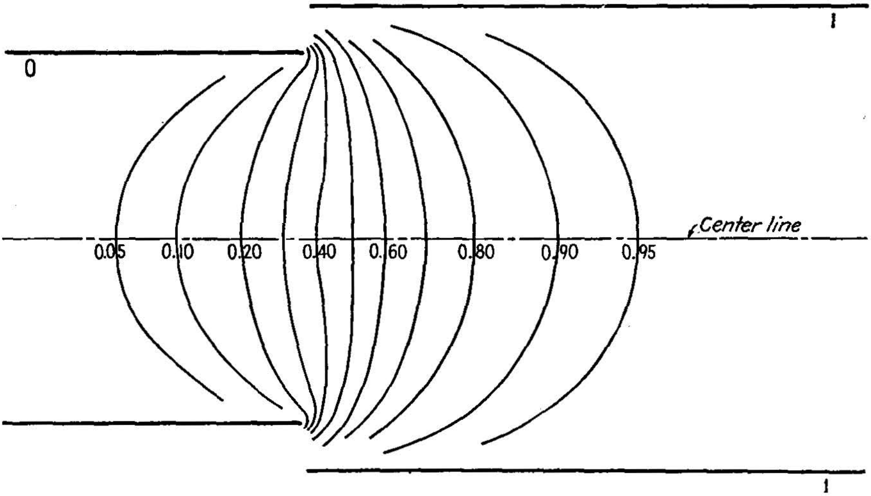


FIG. 13.10.—Potential field of a two-cylinder lens, $D_2/D_1 = 1.25$.

Two-diameter-cylinder Lenses. No exact analytical expressions are available for electrostatic lens made of coaxial cylinders of different diameters. The fields for such lenses are easily measured by means of an electrolytic tank. Results of such measurements are given in Figs. 13.10 to 13.12 for diameter ratios of 1.25, 1.50, and 2.0, respectively. All these electrode arrangements perform about equally well as lenses so that there is not much choice between them. A comparison of their characteristics is had by plotting their axial-potential variations on the same graph, as is done in Fig. 13.13. The differences between these curves are not of great practical interest. All exhibit the same general characteristics. They differ only in the amounts and position of their maximum slopes and curvatures.

Aperture Lenses. Another lens of great interest is that associated with a circular aperture in a plate perpendicular to an applied field.

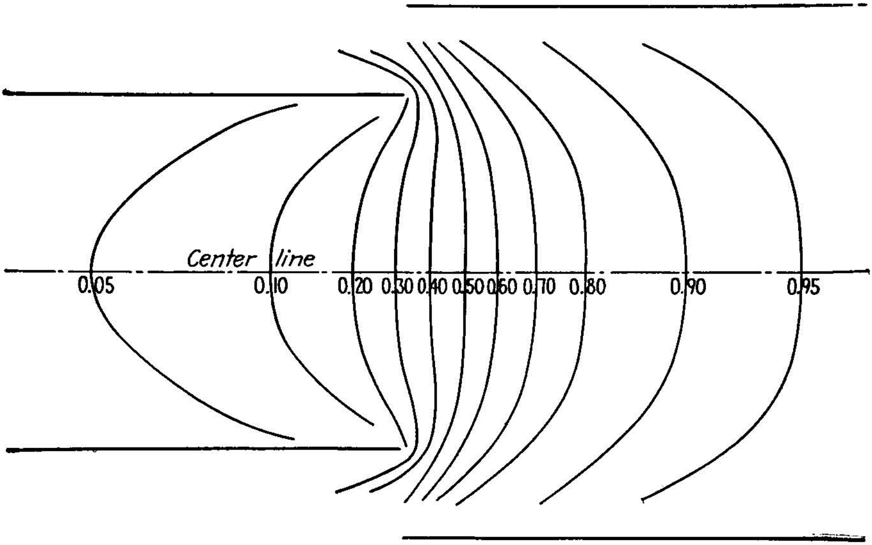


FIG. 13.11.—Potential field of a two-cylinder lens, $D_2/D_1 = 1.5$.

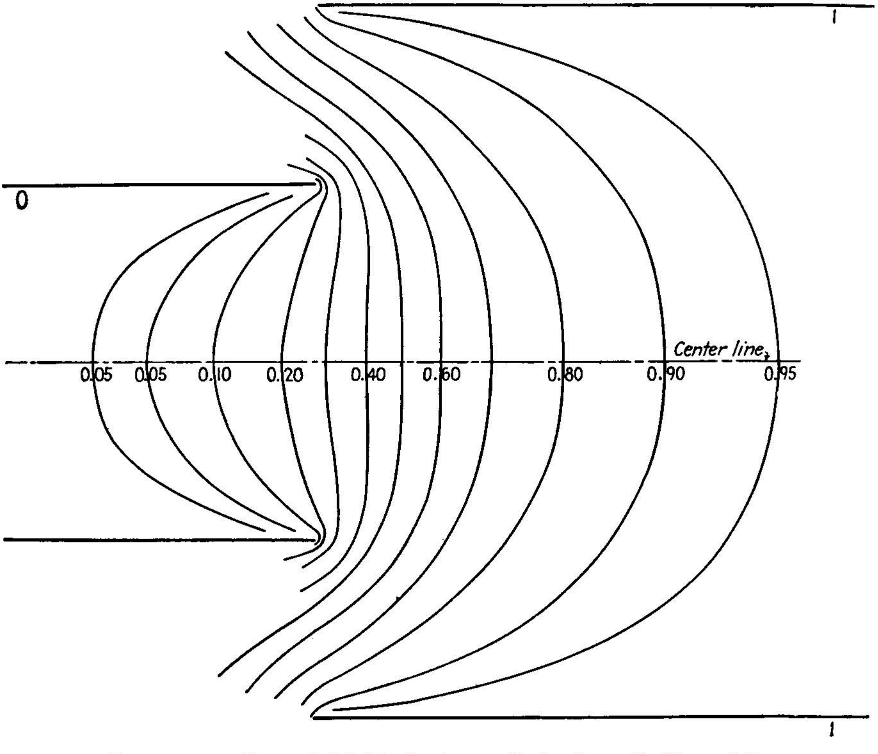


FIG. 13.12.—Potential field of a two-cylinder lens, $D_2/D_1 = 2.0$.

Such a lens will have potential variations that are essentially the same as those between parallel plates except in the immediate vicinity of the aperture. The potential expressions for this case have been worked out by fitting equipotential surfaces which are hyperboloids of revolution to the circular aperture, the edges of which in any plane through the axis are the foci of the hyperboloids.[1] For the case of a plane electrode containing a circular aperture of radius R and operated at zero potential midway between two planes at potential V and spaced a distance d large compared with R, the expression for the axial potential is

$$V_0(z) = V\frac{|z|}{d} - V\frac{R}{d}\frac{2}{\pi}\left[\frac{z}{R}\arctan\left(\frac{R}{z}\right) - 1\right] \tag{13.36}$$

where z is the distance measured from the center plate containing the circular aperture. The resulting potential field is shown in Fig. 13.14. This is a case that exhibits a saddle point at the center of the aperture. Here the potential profiles parallel to the axis and to the center plate curve in opposite direc-

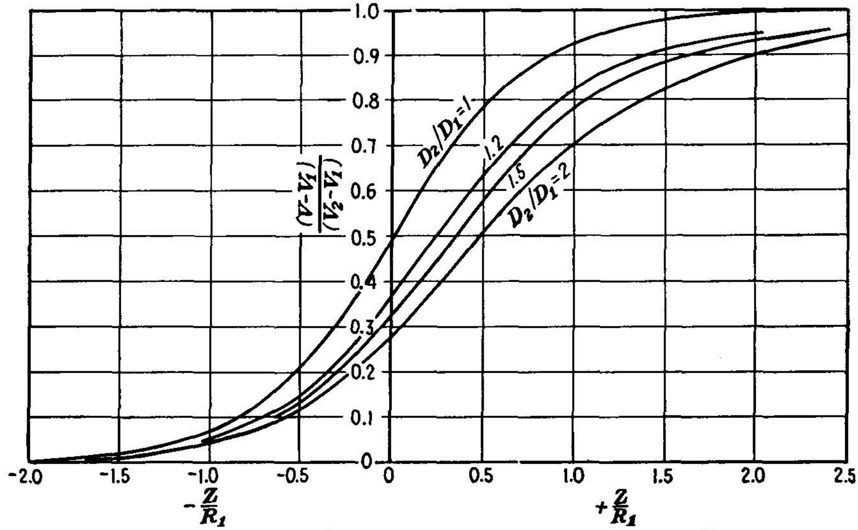


FIG. 13.13.—Axial potentials of two-cylinder lenses for different diameter ratios.

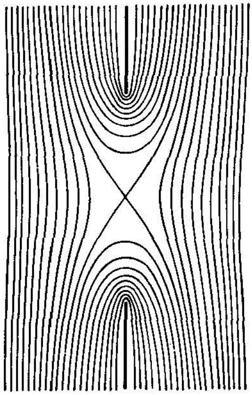


FIG. 13.14.—Aperture midway between plates at the same potential.

---

[1] OLLENDORF, F., "Potential Felder der Elektrotechnik," pp. 295–297, Springer, Berlin, 1932.

tions. The equipotentials through this point are seen to be straight lines making an angle of 54°44′ with the axis.

A case of more general interest is that of a circular aperture in a plate between two plates of different potentials and at different spacings. Let the plates and their potentials be numbered in order from left to right, the circular aperture being in plate 2. The axial potential is then given by

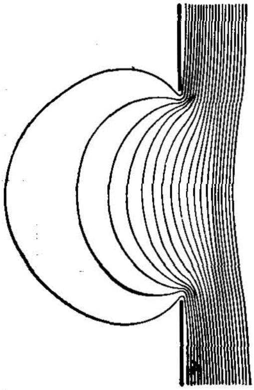


FIG. 13.15.—Potential field of a single-aperture lens.

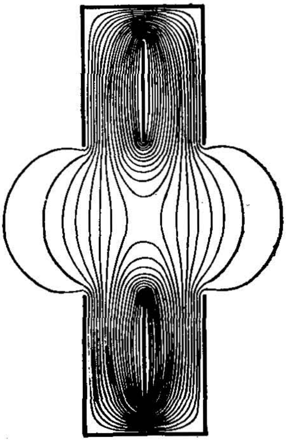


FIG. 13.16.—Potential field of an *Einzel* lens.

$$V_0(z) = \frac{(V_1 - V_2)d_{23} + (V_3 - V_2)d_{12}}{2d_{12}d_{23}} \left\{ |z| - \frac{2R}{\pi}\left[\frac{z}{R}\arctan\left(\frac{R}{z}\right) - 1\right]\right\} + \frac{(V_3 - V_2)d_{12} - (V_1 - V_2)d_{23}}{2d_{12}d_{23}} z + V_2 \quad (13.37)$$

where V_1 is potential of first plate
V_2 is potential of plate containing aperture, the second plate
V_3 is potential of third plate
d_{12} is distance from first to second plate
d_{23} is distance from second to third plate
z is axial distance measured from plate containing aperture
R is aperture radius

The resultant potential field for the case of V_1 and V_2 having a value of zero is shown in Fig. 13.15. The penetration of the equipotential lines into the region of zero potential gradient is seen to be quite small. At one aperture diameter the potential gradient falls to about 5 per cent of the gradient on the other side of the aperture.

A type of electrostatic lens using apertures that has proved very useful in some electron microscopes is shown in Fig. 13.16. This lens is called an "*Einzel* lens" (single lens) because it initiates and terminates in a single value of potential. It has the advantage that it may be placed anywhere along an electron stream without disturbing adjacent potential relations. The electrons leave this lens at the same potential at which they enter. The axial potential forms a symmetrical hill with a saddle point at the center.

13.3. Electron Paths. The general differential equation for the path of an electron in an electrostatic field of rotational symmetry is a little too complex to be generally useful. If, however, the considerations are restricted to electrons that move close to the axis and make a small angle with it, the so-called "paraxial rays," then the differential equation of motion becomes relatively simple. In electron optics as in physical optics it is found that most of the properties of lenses can be determined from the behavior of the paraxial rays.

The general differential equation in two-dimensional cylindrical coordinates is the same as that for two-dimensional rectangular coordinates as given in Eq. (6.59), with z and r substituted for x and y,

$$2V\frac{d^2r}{dz^2} = \left(\frac{\partial V}{\partial r} - \frac{\partial V}{\partial z}\frac{dr}{dz}\right)\left[1 + \left(\frac{dr}{dz}\right)^2\right] \tag{13.38}$$

where the potential V is understood to be a function of r and z. If attention is restricted to paraxial electrons, then the angle that these make with the axis is small and hence the term $\left(\frac{dr}{dz}\right)^2$ is small compared with unity and can be dropped. Further, if the radial distance of an electron from the axis is small, use can be made of the small value approximations derived from the series expansion for potential as given in Eq. (13.21). Thus

$$\frac{\partial V}{\partial r} = 0 - \frac{2rV_0''}{4} + \frac{4V_0^{(4)}r^3}{64} + \cdots \tag{13.39}$$

or, for small r, approximately

$$\frac{\partial V}{\partial r} \cong -\frac{rV_0''}{2} \tag{13.40}$$

Likewise,

$$\frac{\partial V}{\partial z} = V_0' - \frac{r^2V_0'''}{4} + \cdots \tag{13.41}$$

or, for small r, approximately

$$\frac{\partial V}{\partial z} \cong V_0' \tag{13.42}$$

Upon making these substitutions into Eq. (13.38) and letting $V = V_0$ for the conditions of small radial distance imposed, the *differential equation of motion of a paraxial electron becomes*

$$\frac{d^2r}{dz^2} + \frac{V_0'}{2V_0}\frac{dr}{dz} + \frac{V_0''}{4V_0}r = 0 \tag{13.43}$$

In all the above, the argument z has been understood to be associated with axial potential V_0. Equations (13.43) and (13.21) are probably the two most important equations in electrostatic electron optics. From these all the important relations regarding lenses may be derived. The above equation may be reduced to several alternative forms that are sometimes more useful. By combining the first and second derivatives there results

$$\sqrt{V_0}\frac{d}{dz}\left(\sqrt{V_0}\frac{dr}{dz}\right) = -\frac{V_0''r}{4} \tag{13.44}$$

The first-derivative terms may be eliminated from Eq. (13.43) by making the substitution:

$$\rho = rV_0^{1/4} \tag{13.45}$$

The differential equation of motion then becomes

$$\frac{d^2\rho}{dz^2} + \frac{3}{16}\left(\frac{V_0'}{V_0}\right)^2\rho = 0 \tag{13.46}$$

All the above differential equations confirm the observations previously made on the properties of the electron paths. *The path is seen to be independent of the charge and mass of the electron. The path depends only upon the shape of the potential field and not upon the magnitude of the potential. If the electrode configuration is enlarged, the electron path is correspondingly enlarged.*

In general, the expressions for axial potential are sufficiently complex in even the simplest cases so that it is not possible to solve explicitly for the electron paths. It is, however, possible to solve the differential equation of the electron path numerically in all cases. In spite of the fact that the differential equations of motion are in general insoluble, most of the important properties of lenses may be deduced from them.

13.4. General Lens Properties. *Thin Lenses.* A thin lens is one in which the lens dimensions are short compared with the focal length. The focal length of such a lens may be determined from Eq. (13.44) by studying the path of an electron that enters a region of potential variation parallel to the axis. If the angle at which this electron emerges from the lens can be determined, the focal length will be known without solving for the path completely. Let the lens under consideration be one similar to that shown in Fig. 13.3. Let the initial and final values of potential relative to a path from left to right be V_1 and V_2, respectively. A first integration of Eq. (13.44) gives

$$\left[\sqrt{V_0}\frac{dr}{dz}\right]_1^2 = -\frac{1}{4}\int_{z_1}^{z_2} \frac{rV_0''}{\sqrt{V_0}}\,dz \tag{13.47}$$

If the lens is very short, then the value of r will not be greatly changed in passing through the region of potential variation though the direction of the electron and hence the value of $\frac{dr}{dz}$ will be. The coordinate r may accordingly be treated as a constant and removed from within the integral sign. If, in addition, attention be restricted to the second principal ray, *i.e.*, the ray entering the lens parallel to the axis from the left, then the lower limit of the left-hand term of the equation is zero and the equation reduces to

$$\sqrt{V_2}\left(\frac{dr}{dz}\right)_2 = \frac{-r}{4}\int_{z_1}^{z_2} \frac{V_0''}{\sqrt{V_0}}\,dz \tag{13.48}$$

In passing through the lens the electron is bent toward the axis. As soon as the electron is a short distance beyond the lens, it is in a field-free region and hence its path is subsequently a straight line. From simple geometry

$$f_2 = \frac{-r}{\left(\frac{dr}{dz}\right)_2} \tag{13.49}$$

where r is the radial position of the electron on passing through the lens. From this the formula for focal length becomes

$$\frac{1}{f_2} = \frac{1}{4\sqrt{V_2}}\int_{z_1}^{z_2} \frac{V_0''}{\sqrt{V_0}}\,dz \tag{13.50}$$

A similar treatment of the case of an electron entering the lens parallel to the axis from the right yields

$$\frac{1}{f_1} = \frac{-1}{4\sqrt{V_1}}\int_{z_1}^{z_2} \frac{V_0''}{\sqrt{V_0}}\,dz \tag{13.51}$$

When the axial potential of a lens is known, it is necessary only to measure the area under the curve of $\frac{V_0''}{\sqrt{V_0}}$ and then multiply by the reciprocal of four times the square root of external potential. Comparing Eqs. (13.50) and (13.51), it is seen that the two focal lengths of the lens are in the ratio of the square root of the limiting values of potential,

$$\frac{f_2}{f_1} = \frac{-\sqrt{V_2}}{\sqrt{V_1}} \tag{13.52}$$

This is exactly analogous to the law for light lenses, which says that the focal lengths are in the ratio of the indices of refraction on the two sides of the lens.

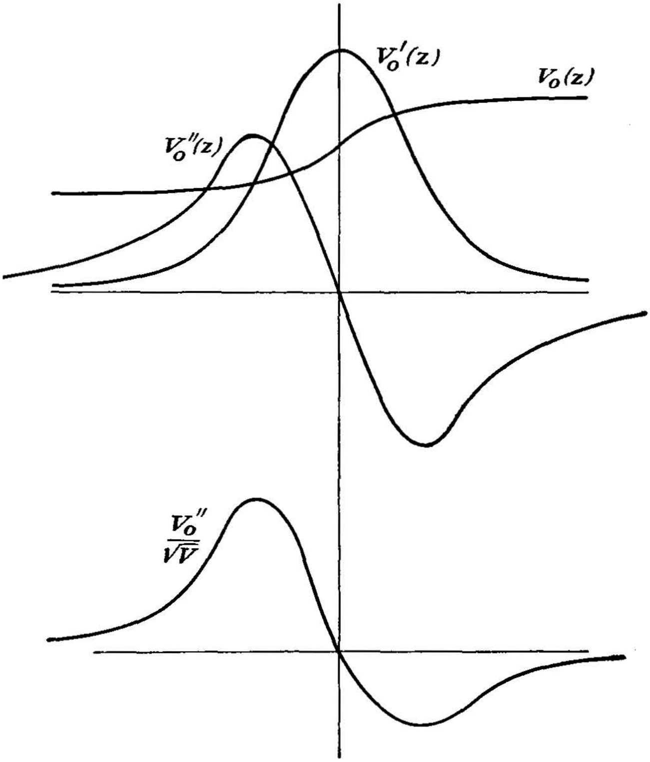


FIG. 13.17.—Axial potential functions of a double-aperture lens.

Some typical curves of axial potential and the integrand of Eqs. (13.50) and (13.51) for the lens of Fig. 13.3 are shown in Fig. 13.17. The first part of the lens has a convergent action, and this is associated with a positive value of the second derivative of the axial potential. The second part of the lens has a divergent action, which is, however, weaker because of the higher velocity of the electron, and this is associated with a negative value of the second derivative of the axial potential. The reason for the association of the sign is evident from Eq.

(13.40), from which it is seen that the radial component of the gradient of potential is directly proportional to the second derivative of the axial potential as long as the distance from the axis is not too great. As a result of this, the radial force on an electron is directed toward the axis when the second derivative of the axial potential is positive, and vice versa. It may be stated as a general rule that *the action of a lens segment is convergent whenever the second derivative of the axial potential is positive and divergent whenever the second derivative of the axial potential is negative.* In the case of a symmetrical lens such as is shown in Fig. 13.3, the convergent and divergent forces in the two halves of the lens are the same, but the deflection that results is always greater on the low potential side; for here the velocity of the electron is less, and the deflection for a given force is greater.

An alternative form of Eqs. (13.50) and (13.51) that yields much useful information is obtained by evaluating the integral by parts. Let

$$u = V_0^{-1/2} \qquad \text{and} \qquad dv = V_0'' \, dz$$

Then

$$du = -\tfrac{1}{2} V_0^{-3/2} V_0' \, dz \qquad \text{and} \qquad v = V_0'$$

Making use of these substitutions in the well-known formula for integration by parts,

$$\int u \, dv = uv - \int v \, du \tag{13.53}$$

there results

$$\frac{1}{f_2} = \frac{V_0'(z_2) - V_0'(z_1)}{4V_0(z_2)} + \frac{1}{8\sqrt{V_0(z_2)}} \int_{z_1}^{z_2} \frac{(V_0')^2}{V_0^{3/2}} \, dz \tag{13.54}$$

where z_1 refers to a point to the left of the lens just outside of the region of appreciable potential variation and z_2 refers to a corresponding point to the right of the lens. The corresponding formula for the first focal length is had by simply interchanging the subscripts 1 and 2 in the above equation.

For the case of lenses whose initial and final gradients of potential are zero, the first term of the right-hand side above becomes zero, and the integral alone gives the focal length.

$$\frac{1}{f_2} = \frac{1}{8\sqrt{V_0(z_2)}} \int_{z_1}^{z_2} \frac{(V_0')^2}{V_0^{3/2}} \, dz \tag{13.55}$$

The form of the integral in this case is particularly revealing. It is apparent that the integrand is always positive because the first derivative of the axial potential, which may be negative, is squared and hence the focal length is positive. The interpretation of this is that *the lens is* *convergent in all cases in which the initial and final gradients of potential are zero.*

When the initial and final gradients of potential are not zero, as is the case with most single-aperture lens, the first term of Eq. (13.54) will usually make the major contribution to the focal length.

$$\frac{1}{f_2} = \frac{V_0'(z_2) - V_0'(z_1)}{4V_0(z_2)} \tag{13.56}$$

This formula is generally accepted as a sufficiently accurate one for single apertures. The lens action of such an aperture may either be convergent, f positive, or divergent, f negative. In the simple case of an aperture at a positive potential in front of a plane cathode and having a field-free region beyond, the lens action is divergent and the focal length is four times the cathode-aperture spacing, as may be seen by substitution into Eq. (13.56).[1]

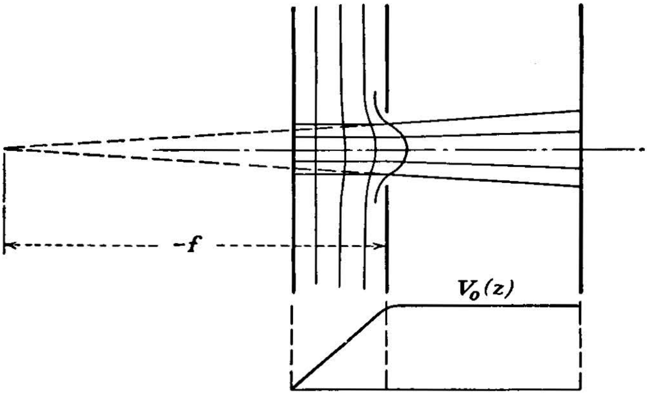


FIG. 13.18.—Divergent action of a single-aperture lens.

The focusing properties of single apertures are illustrated in Figs. 13.18 and 13.19. Figure 13.18 shows the case of an aperture with a positive gradient of potential on its left and a zero potential gradient on its right. The difference of the gradients is therefore negative, and the lens is divergent. Figure 13.19 shows the case of an aperture with a zero gradient of potential on its left and a positive one on its right. The difference of gradients in this case is positive, and the lens is convergent.

---

[1] The corresponding formula for the focal length of a lens consisting of a straight slit in a plane electrode is

$$\frac{1}{f_2} = \frac{V_0'(z_2) - V_0'(z_1)}{2V_0(z_2)}$$

This means that the cylindrical lens of a slot is twice as strong as the circular lens of an aperture.

Because of the positive gradient to the right of the lens the electron paths upon emergence from the aperture are slightly curved, being parabolic rather than straight. This results because the aperture imparts a crosswise component of velocity which is proportional to the distance from the axis at which the electron crosses the aperture plane. The subsequent field adds a constant axial component of acceleration to this constant crosswise component of velocity.

Thick Lenses. No simple formulas exist for the parameters of thick lenses. In order to treat this subject it is first necessary to define the lens parameters. Then a number of basic relations between the parameters can be pointed out. It can later be shown how the lens parameters may be calculated or measured. After that it is desirable to present the resultant lens characteristics in some simple compact form. These steps will now be taken up one at a time. Attention will be restricted to those lenses whose initial and final gradients of potential are zero.

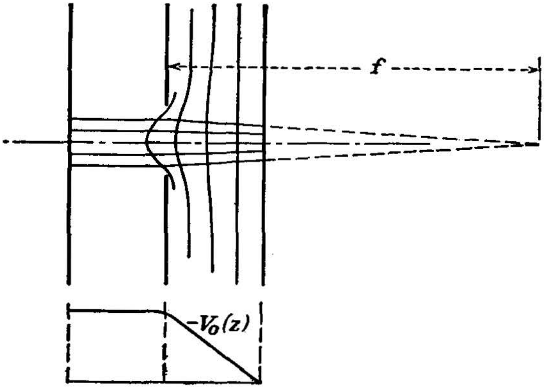


FIG. 13.19.—Convergent action of a single-aperture lens.

The differential equation of motion of the paraxial electron given in Eq. (13.43) is a second-order linear differential equation. As such, it has two linearly independent solutions, and any general solution can be expressed as a linear combination of these two independent solutions. It is convenient to take as the independent solutions of the equation the ray that leaves the lens parallel to the axis and the ray that enters the lens parallel to the axis. *The two rays that leave and enter the lens parallel to the axis, respectively, are known as the "principal rays" of the lens.* The ray that is parallel to the axis to the right of the lens is known as the "first principal ray." It is usually considered to be moving from right to left, but it may just as well be considered as moving from left to right. The ray that is parallel to the axis on the left side of the lens is known as the "second principal ray." For lenses with initial and final gradients of potential that are zero the initial and final portions of the rays will be straight lines. The principal rays of an equal-diameter two-cylinder lens are shown in Fig. 13.20. Any general ray may be expressed as a combination of these two rays.

As mentioned before, the left portion of a lens such as that of Fig. 13.20 has a convergent action, while the right portion has divergent action. The strength of these two portions is such that the convergent action always dominates. The first principal ray, taken as moving from right to left, first experiences divergent action and then a stronger convergent action. The second principal ray, taken as moving from left to right, first experiences a convergent action and then a weaker divergent action.

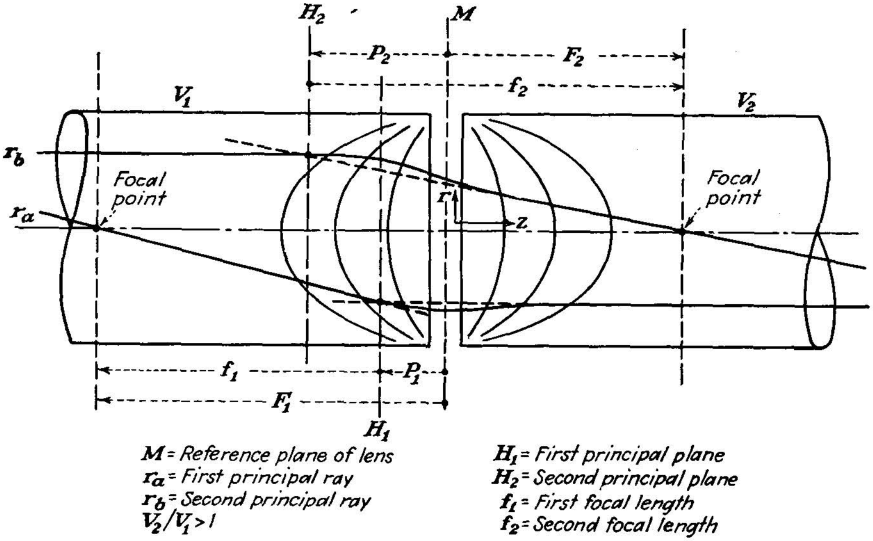


FIG. 13.20.—Thick-lens terminology.

The principal rays serve to define the four thick-lens parameters. *If the initial and final straight-line portions of the principal rays are extended until they intersect, the intersections locate what are known as the "principal planes."* The principal planes are shown as H_1 and H_2 in Fig. 13.20. The location of the principal planes relative to the reference plane, usually the midplane or electrode junction, is given by the distances P_1 and P_2. Almost without exception, the relative location of the principal planes is as shown in Fig. 13.20. Both principal planes lie on the foreside of the lens. Furthermore, the principal planes are crossed, *i.e.*, the second principal plane lies before the first principal plane. Although this is the usual relative disposition for electrostatic electron lenses, it is not for light lenses. A thick double-convex light lens, for instance, has its two principal planes on opposite side of the lens center and not crossed.

A focal length of a thick lens is defined as the distance from the principal plane of the lens to the point at which the corresponding principal ray crosses the axis of the lens. There are two focal lengths, one associated with each axis. These are designated by the symbol f as shown in Fig. 13.20.

The intersections of the principal rays with the lens axis are known as "focal points." The distance from the lens center to a focal point is indicated by the symbol F. The above definitions are sufficient to describe completely the characteristics of a thick lens.

Let the two principal rays of a lens be $r_a(z)$, the first principal ray, and $r_b(z)$, the second principal ray. Then any general ray may be expressed as a linear combination of these two principal rays,

$$r(z) = c_a r_a(z) + c_b r_b(z) \tag{13.57}$$

Although it is not ordinarily possible to write the expressions for the complete principal rays, it is possible to write expressions for the initial and final straight-line portions and the general lens formula can be derived from these partial expressions.

Let the radial offset of the portion of the principal rays parallel to the axis be unity. Then, to the left of the lens, as in Fig. 13.20 the straight-line portions of the principal rays are given by

$$r_a(z_1) = \frac{z_1 - P_1 - f_1}{f_1} \tag{13.58}$$

and

$$r_b(z_1) = 1 \tag{13.59}$$

Assume that the general ray starts at a point on the axis of the lens to the left of the first focal point. The general ray will then pass through the lens, be deflected toward it, and cross the axis again at a point to the right of the second focal point. At the point where the general ray crosses the axis to the left of the lens, from Eq. (13.57),

$$\frac{c_a}{c_b} = -\frac{r_b}{r_a} = \frac{-f_1}{z_1 - P_1 - f_1} \tag{13.60}$$

To the right of the lens the principal rays are given by

$$r_a(z_2) = -1 \tag{13.61}$$

and

$$r_b(z_2) = \frac{-z_2 + P_2 + f_2}{f_2} \tag{13.62}$$

At the point on the right of the lens where the general ray crosses the axis, r is zero, and from Eq. (13.57)

$$\frac{c_a}{c_b} = -\frac{r_b}{r_a} = \frac{-z_2 + P_2 + f_2}{f_2} \tag{13.63}$$

Equating the ratio of constants at the two axial crossings of the general ray,

$$-f_1 f_2 = (z_1 - P_1 - f_1)(P_2 + f_2 - z_2) \tag{13.64}$$

This is readily rearranged to give

$$\frac{-f_1}{z_1 - P_1} + \frac{f_2}{z_2 - P_2} = 1 \tag{13.65}$$

This is the lens formula of a thick lens of the type shown in Fig. 13.20. It is the counterpart of Eq. (13.7) for light lenses. The sign convention used here has been that all quantities measured to the right from a reference plane are positive, while those measured to the left from a reference plane are negative.

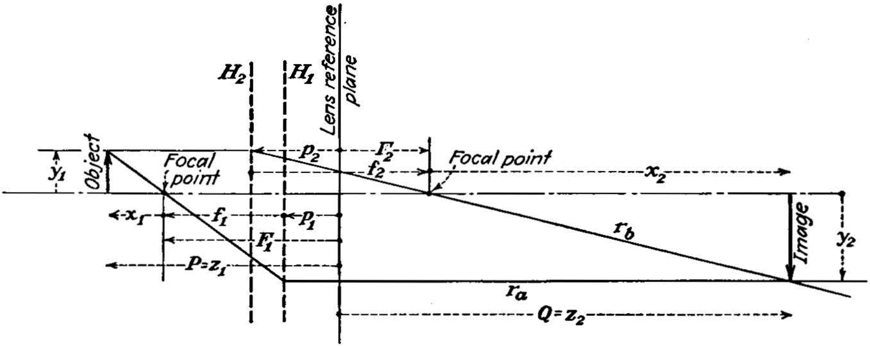


FIG. 13.21.—Graphical construction of a thick-lens image.

The significance of Eq. (13.65) is that the focal lengths are measured not from the lens center but from the corresponding principal planes.

The lens parameters defined above are known as the "cardinal characteristics" of the lens. It takes four of these to describe the thick lens. The quantities that are usually given are the two focal lengths and the distance of the focal points from the reference plane. In electron lenses the parameters change with voltage ratio so that it is necessary to present curves of these four quantities as a function of the voltage ratio.

Knowing the focal lengths and the position of the principal planes makes it possible to construct an image corresponding to any object. The construction involved is shown in Fig. 13.21. Through a point on the object draw a line through the first focal point until it intersects the first principal plane. From this point of intersection draw a line to the right parallel to the axis. These two segments of straight line correspond to the first principal ray. From the same point on the object draw a line parallel to the axis to the right until it intersects the second principal plane. From the point of intersection draw a straight line through the second focal point until it intersects the first principal ray. This last point of intersection defines the point on the image corresponding to the point on the object from which the two principal rays originated.

From Fig. 13.21 the lateral magnification of the lens can be defined as

$$\begin{aligned} M &= \frac{y_2}{y_1} \\ &= -\frac{x_2}{f_2} \\ &= -\frac{f_1}{x_1} \end{aligned} \tag{13.66}$$

where the y's are the radial coordinates of corresponding points on object and image and the x's are the distances from object and image to the nearest focal point. From the above relations there results *Newton's law,*

$$x_1x_2 = f_1f_2 \tag{13.67}$$

From Fig. 13.21 it is also seen that object and image distances from the lens reference plane are given in terms of the cardinal lens parameters by

$$z_1 = P = F_1 - \frac{f_1}{M} \tag{13.68}$$

and

$$z_2 = Q = F_2 - Mf_2 \tag{13.69}$$

To determine the remaining laws of importance applying to thick lenses, reference is again made to the differential equation of motion of the paraxial electron [Eq. (13.44)]. Consider the linearly independent principal rays $r_a(z)$ and $r_b(z)$. Substitute r_a into Eq. (13.44), and multiply by r_b. Then substitute r_b into the same equation, multiply by r_a, and subtract from the first equation. Indicating derivatives with respect to z by primes,

$$r_b(\sqrt{V_0}\,r_a')' - r_a(\sqrt{V_0}\,r_b')' = 0 \tag{13.70}$$

Add and subtract the quantity $\sqrt{V_0}\,r_a'r_b'$; then

$$\frac{d}{dz}\left(r_b\sqrt{V_0}\,r_a' - r_a\sqrt{V_0}\,r_b'\right) = 0 \tag{13.71}$$

Integrate between limits z_1 and z_2 far enough to the left and right of the lens so that the potential variation is negligible. This allows $r_a'(z_2)$ and $r_b'(z_1)$ to be considered zero, and the result of the integration reduces to

$$r_a(z_2)\sqrt{V_0(z_2)}\,r_b'(z_2) + r_b(z_1)\sqrt{V_0(z_1)}\,r_a'(z_1) = 0 \qquad (13.72)$$

From this equation two very important conclusions may be drawn. Observe that

$$f_1 = -\frac{r_a(z_2)}{r_a'(z_1)} \qquad (13.73)$$

and

$$f_2' = -\frac{r_b(z_1)}{r_b'(z_2)} \qquad (13.74)$$

Upon making these substitutions into Eq. (13.72), it follows that

$$\frac{f_2}{f_1} = -\frac{\sqrt{V_0(z_2)}}{\sqrt{V_0(z_1)}} \qquad (13.75)$$

This is a perfectly general proof that the ratio of the focal lengths of a lens is the same as the ratio of the corresponding indices of refraction. The relation is valid, however, only if the focal points of the lens lie outside of the region of appreciable potential variation.

Returning to Eq. (13.72) again and identifying the ratio $-\dfrac{r_a(z_2)}{r_b(z_1)}$ as the lateral magnification M, and $-\dfrac{r_b'(z_2)}{r_a'(z_1)}$ as the angular magnification m_a,

$$Mm_a\frac{\sqrt{V_0(z_2)}}{\sqrt{V_0(z_1)}} = 1 \qquad (13.76)$$

which is *Lagrange's law*. The angular magnification is the ratio of the tangents of the angles that the second and first principal rays make with the axis. For small angles, the tangent is approximately equal to the angle. The above law states that the product of the lateral magnification, the angular magnification, and the ratio of the final and initial indices of refraction is unity. This law has its exact counterpart in geometrical optics.

13.5. Calculation of Lens Characteristics. Since analytical methods fail in general in determining the characteristics of thick lenses, recourse is frequently had to numerical computation. From the previous discussion it is known that, if the potential along the axis of an electrostatic lens is known, then the potential anywhere in the lens is determined and can be calculated. Further, the differential equation for an electron moving close to the axis and making a small angle with the axis, a so-called "paraxial electron," can be written in terms of the axial potential. It is necessary only to solve such an equation numerically for rays entering and leaving the lens parallel to the axis, the principal rays, to obtain the cardinal lens characteristics, *i.e.*, the focal lengths and the location of the principal planes.

Numerous methods for calculating the principal rays of electrostatic lenses have been proposed. The most important of these will be briefly described, and then two of the simplest methods will be given in more detail.

Klemperer and Wright have proposed an application of the trigonometric ray-tracing method of physical optics.[1] The electrostatic field is broken up into a succession of thin lenses having a constant ratio of equivalent index of refraction for adjacent lenses. Formulas are given for calculating the effect of every refraction at a lens surface upon the angle of a ray and the point at which it crosses the axis. Lens surfaces are assumed to be spherical, and their radius of curvature must be determined either graphically or from the axial potential. This method requires a large number of equivalent thin lenses, at least 20 for an accurate determination, and the results converge slowly as the number of segments taken is increased.

Maloff and Epstein have proposed several methods based upon a step-by-step solution of the differential equation of motion of the paraxial electron. The methods give the electron path as an exponential of the axial distance in any increment and join the paths in successive increments both in magnitude and in slope. The methods are capable of good accuracy, but the tabulations are very numerous.[2]

A method of joined circular segments based upon Salinger's formula for the radius of curvature of an electron path has also been proposed. Increments of radial and axial displacement are expressed in terms of axial potential and associated factors.[3] This method likewise requires rather extensive tabulation.

Method of Linear Axial-potential Segments. One of the simplest methods proposed is based upon the differential equation of motion of

---

[1] KLEMPERER, and W. D. WRIGHT, Investigations of Electron Lenses, *Proc. Phys. Soc.* (*London*), vol. 51, Part II, pp. 296–317, March, 1939.

[2] MALOFF and EPSTEIN, *op. cit.*, pp. 81–89.

See also SCHLESINGER, KURT, A Mechanical Theory of Electron-image Formation, *Proc. I.R.E.*, vol. 32, pp. 483–493, August, 1944.

[3] SPANGENBERG, KARL, and L. M. FIELD, Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses, *Proc. I.R.E.*, vol. 3, pp. 138–144, March, 1942.

the paraxial electrons.[1] The axial potential is replaced by a number of straight-line segments that approximate it as closely as possible, as shown in Fig. 13.22. The differential equation is then solved for the successive regions in which the potential is linear and the gradient is constant. At each boundary between segments there is a jump in the slope of the electron path because of the jump in the gradient of potential. The final path as determined by this method consists of a number of curved segments of path connected together, giving a path that is continuous but that has discontinuities in slope at the corners of the segmented approximation to the axial distribution of potential. Such a path cannot represent accurately the true nature of the path within the lens, but it can be used to obtain relations between initial and final values with considerable accuracy. The method is relatively easy to apply and gives fair accuracy for as few as six straight-line segments in the approximation to the axial-potential curve.

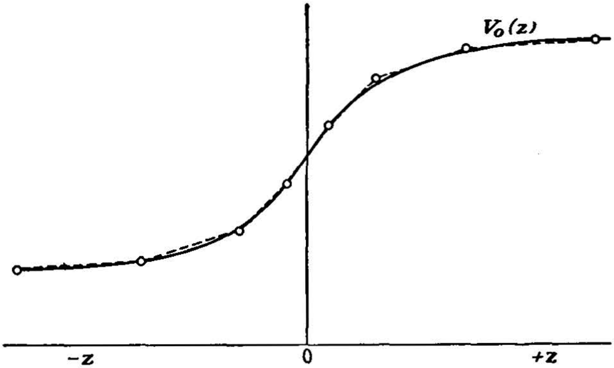


FIG. 13.22.—Approximation of axial potential by linear segments of potential.

If the axial potential is assumed to be made up of straight-line segments, then the second derivative of the axial potential is zero and the differential equation of motion of paraxial electrons of Eq. (13.43) reduces to

$$r'' + \frac{1}{2}\frac{r'}{V_0}V_0' = 0 \tag{13.77}$$

where both r and V_0 are functions of axial distance z and the primes denote derivatives with respect to z. A first integration of this equation gives

$$r' = CV_0^{-\frac{1}{2}} \tag{13.78}$$

---

[1] GANS, R., Electron Paths in Electron Optics, *Zeit. für Tech. Phys.*, vol. 18, pp. 41–48, February, 1937.

A second integration gives

$$r(z_2) - r(z_1) = \frac{2C[V_0^{1/2}(z_2) - V_0^{1/2}(z_1)]}{V_0'} \tag{13.79}$$

where subscripts 1 and 2 refer to the left and right extremities of a segment. These two general equations give the electron path along any segment of axis. At the junction of two segments there is a discontinuity in the slope of the axial-potential function. Upon integrating the first and last terms of Eq. (13.43), the difference of the slopes of the electron path on the two sides of the junction is proportional to the difference of the first derivative of axial potential on the two sides of the junction,

$$r'(z_b) - r'(z_a) = -r\,\frac{[V_0'(z_b) - V_0'(z_a)]}{4V_0} \tag{13.80}$$

where subscripts a and b refer to values on the left and right side of the junction, respectively. In the particular case where the derivative of the axial potential is zero, integration of Eq. (13.77) gives

$$r(z_2) - r(z_1) = (z_2 - z_1)[r'(z_1)] \tag{13.81}$$

The above set of equations suffices to calculate approximate principal rays. By alternate use of Eqs. (13.80) and (13.79) and the occasional use of other equations where necessary, the focal lengths and focal points of a lens may be obtained.

Method of Equivalent Thin Lenses. The usual electron lens has a convergent behavior on the low-potential side and a divergent behavior on the high-potential side, the net lens behavior being convergent. The behavior is convergent when the second derivative of the axial potential is positive and divergent when the second derivative is negative. It is reasonable, therefore, to consider that the lens is made up of two thin lenses, a convergent lens followed by a divergent lens.[1] If the strength and location of these lenses are known, the cardinal points of the equivalent thick lens may be determined.

The focal lengths of the convergent lens as shown in Fig. 13.23 are given by

$$\frac{1}{F} = \int_{z_1}^{z_m} \frac{V''}{2\sqrt{2V}}\,dz \tag{13.82}$$

and

$$f_1 = -F\sqrt{2V_1} \tag{13.83}$$

$$f_2 = -f_1\sqrt{\frac{V_m}{V_1}} \tag{13.84}$$

---

[1] MYERS, *op. cit.*, p. 131.

where F is a focal term from which the focal lengths are derived, V_1 is the lowest potential on the lens axis, and V_m is the potential at the point at which the second derivative assumes a value of zero, changing sign. The integration of (13.82) is carried over the region in which the second derivative is positive.

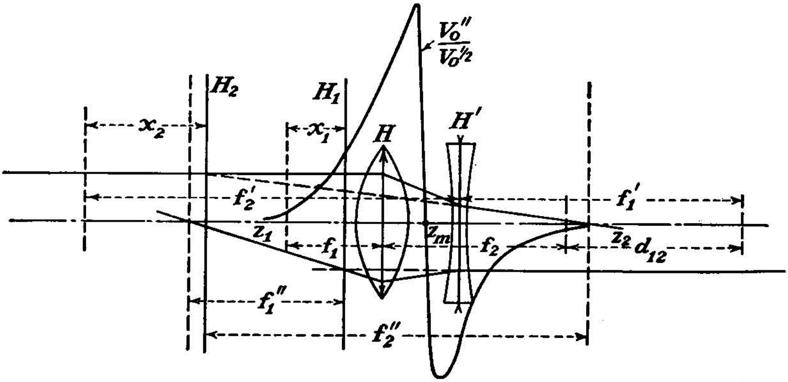


FIG. 13.23.—Thin-lens components of a thick lens.

Similarly, the focal distances for the divergent component of the lens are given by

$$\frac{1}{F'} = \int_{z_m}^{z_2} \frac{V''\,dz}{2\sqrt{2V}} \qquad (13.85)$$

and

$$f_1' = -F'\sqrt{2V_m} \qquad (13.86)$$

$$f_2' = -f_1'\sqrt{\frac{V_2}{V_m}} \qquad (13.87)$$

where V_2 is the highest value of potential reached on the axis on passing through the lens.

When the focal lengths of the convergent and divergent components of the lens are known, the focal characteristics of the entire lens are readily determined, this being a simple problem in the combination of lenses. When the distance between the second focal point of the convergent component and the first focal point of the divergent component is d_{12}, then the focal lengths of the entire lens are[1]

$$f_1'' = -\frac{f_1 f_1'}{d_{12}} \qquad (13.88)$$

$$f_2'' = \frac{f_2 f_2'}{d_{12}} \qquad (13.89)$$

---

[1] ROSIN, S., and O. H. CLARK, Combinations of Optical Systems, *Jour. Opt. Soc. Amer.*, vol. 31, pp. 198–201, March, 1941.

The location of the first principal plane measured from the first focal point of the convergent component is

$$x_1 = \left(\frac{f_1 f_2}{d_{12}}\right) + f_1'' \tag{13.90}$$

and the location of the second principal plane as measured from the second focal point of the divergent component of the lens is

$$x_2 = \left(\frac{f_1' f_2'}{d_{12}}\right) - f_2'' \tag{13.91}$$

The method is extremely rapid in application. The location of the lens components is best taken as being at the center of gravity of the area represented by the integrals of Eqs. (13.82) and (13.85), as shown in Fig. 13.23.

13.6. Measurement of Lens Characteristics. All the computational methods referred to in the previous section are subject to some error that is difficult to determine except by extensive calculations. In general, it may be said that, although computational methods are adequate, experimental methods are preferable and usually more dependable.

As with computation so with experimental determination, several methods are available. One method involves construction of a special electron gun, which generates filamentary rays parallel to the axis that are put through the lens being measured.[1] Lens characteristics are obtained from the voltages required to produce a focus.

Another method makes use of an ordinary electron gun followed by a movable mesh grid and then by the lens under test. Data are taken on the voltage ratio necessary to apply to the lens to focus an image of the mesh on a fluorescent screen for all positions of the mesh. Magnifications are also noted and lens characteristics are deduced from these data.[2]

Another experimental method used in determining the lens characteristics is based upon observed magnifications of measuring grids placed before and after the lens structure.[3] This method will be described in some detail.

Double-grid Method of Measuring Lens Characteristics. The experimental method used in determining the lens characteristics is based upon observed magnifications of measuring grids placed before and after the lens structure.

A grid of closely spaced parallel wires (for measurement purposes only and not for control of the beams) is placed in the fore part of the lens. This grid casts a shadow upon a fluorescent screen following the

---

[1] Klemperer and Wright, *op. cit.*

[2] Maloff and Epstein, *op. cit.*

[3] Spangenberg and Field, *op. cit.*

lens. In order to avoid the need of a tube having parts that can be moved relative to one another while in a vacuum, another measuring grid is used between the end of the gun and the fluorescent screen. This arrangement is shown schematically in Fig. 13.24, in which the measuring grid in the fore part of the lens is indicated by a vertical row of dots. With this arrangement of measuring grids, it is necessary to make observations on the magnifications of two grids, as the voltage ratio of the main lens electrodes is varied for each of two distances of the lens from a point source of electrons. Hence two complete runs must be made to obtain the data from which the complete lens characteristics can be measured.

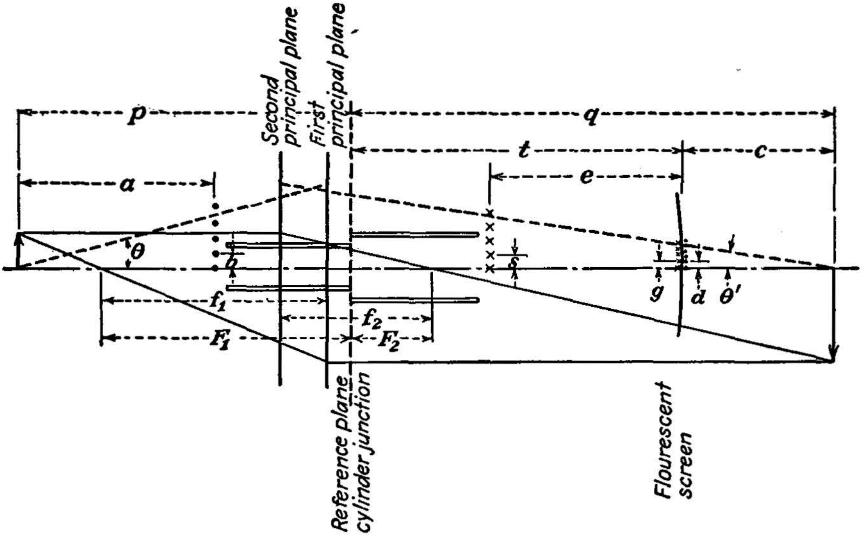


FIG. 13.24.—Experimental determination of electrostatic lens characteristics.

The details of the mathematical relations involved can be seen from Fig. 13.24. The cathode-lens structure gives the effect of a point source of electrons at a known point near the cathode. The location of this point and the constancy of its position under varying conditions of lens voltage ratio are determined by placing two measuring grids in the fore part of the lens and observing the ratio of their magnifications. The constancy of the ratio of magnifications indicates that the location of the point source changes very little with lens voltage ratio and also over the normal range of control-grid voltages used. The location of the point source is very nearly at the control-grid aperture in front of the cathode. When these facts have been checked from a test run, it is no longer necessary to use two measuring grids in the fore part of the lens.

With the point source of electrons available the following general method is applied: The angular magnification of the bundle of rays is determined from screen patterns obtained on the fluorescent screen, such as that shown in Fig. 13.25. Here the lines in one direction are the shadows of one measuring grid, and the lines in the other direction are the shadow of the other measuring grid. When the angular magnification is known, then for any given voltage ratio the lateral magnification can be determined from Lagrange's law, which states that the product of the internal magnification and the angular magnification is equal to the square root of the ratio of the final and initial potentials. Image distances at each of the two object distances used are given for various voltage ratios from magnifications of the second grid alone. The object distances are known from physical measurements on the gun assembly. When lateral magnification, object distance, and image distance are known as a function of voltage ratio for two different values of the object distance, then the cardinal quantities f_1, f_2, F_1, and F_2 of the lens may be calculated readily.

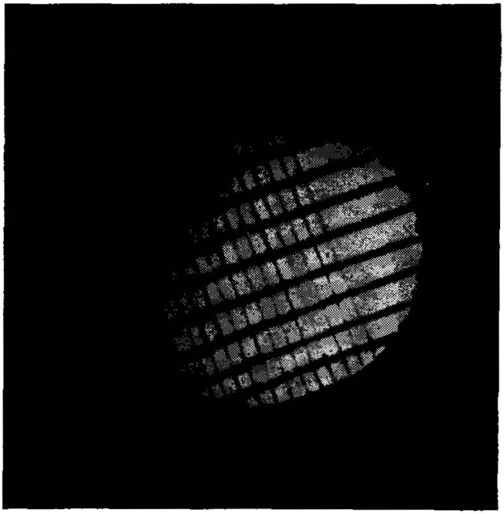


Fig. 13.25.—Shadows of measuring grids on a fluorescent screen.

The method by which this calculation is made will be briefly indicated. Object and image distances can be expressed in terms of the lateral magnification and focal distances as

$$P = -\frac{f_1}{M} + F_1 \tag{13.92}$$

$$Q = -Mf_2 + F_2 \tag{13.93}$$

These two equations involve the four quantities f_1, f_2, F_1, and F_2 as unknowns. In order to determine them, it is necessary to know two sets of associated values of P, Q, and M for the same voltage ratio. When subscripts 1 and 2 are used to indicate values of P, Q, and M for two different values of P at a given voltage ratio, then there may be obtained from the above relations the following expressions for the cardinal focal distances:

$$f_1 = \frac{P_1 - P_2}{\dfrac{1}{M_2} - \dfrac{1}{M_1}} \tag{13.94}$$

$$f_2 = \frac{Q_1 - Q_2}{M_2 - M_1} \tag{13.95}$$

$$F_1 = \frac{P_1 M_2 - P_2 M_1}{M_2 - M_1} \tag{13.96}$$

$$F_2 = \frac{\dfrac{Q_1}{M_2} - \dfrac{Q_2}{M_1}}{\dfrac{1}{M_2} - \dfrac{1}{M_1}} \tag{13.97}$$

Up to this point the relations are the same as those used by Maloff and Epstein. It is now necessary only to show how the lateral magnification may be deduced from the screen patterns to complete the collection of necessary relations. In Fig. 13.24 it is seen that the angular magnification is given by

$$M_a = \frac{\theta'}{\theta} \tag{13.98}$$

For small angles such as are encountered in the gun the angular magnification in terms of the dimensions is given very closely by

$$\frac{\theta'}{\theta} = \frac{ad}{bc} \tag{13.99}$$

in which c is the distance beyond the fluorescent screen to the point at which the ray would focus. This distance is determined from the spacings of the grid images as follows:

For focus beyond fluorescent screen,

$$c = \frac{e}{1 - \dfrac{s}{g}} \tag{13.100}$$

where the symbols have the significance given in Fig. 13.24.

For focus between second measuring grid (crosses) and fluorescent screen

$$-c = \frac{e}{1 + \frac{s}{g}} \tag{13.101}$$

When the angular magnification is known, then the lateral magnification may be calculated from Lagrange's law as previously indicated.

With the above relations the cardinal quantities are readily calculated. In practice, this is most easily done by plotting curves of the various quantities involved against voltage ratio, for the same voltage-ratio observations may not have been taken on one run as on the other. There is a small hole in each curve at the point where the beam focus is at the fluorescent screen, for the image becomes so small here that it is not possible to measure the spacings of the wires on the images. However, there is no trouble in drawing smooth and continuous curves through these holes if the data are taken with care.

The accuracy achieved by this method is of the order of 10 per cent for lenses with small openings and 20 per cent for lenses with large spacings.

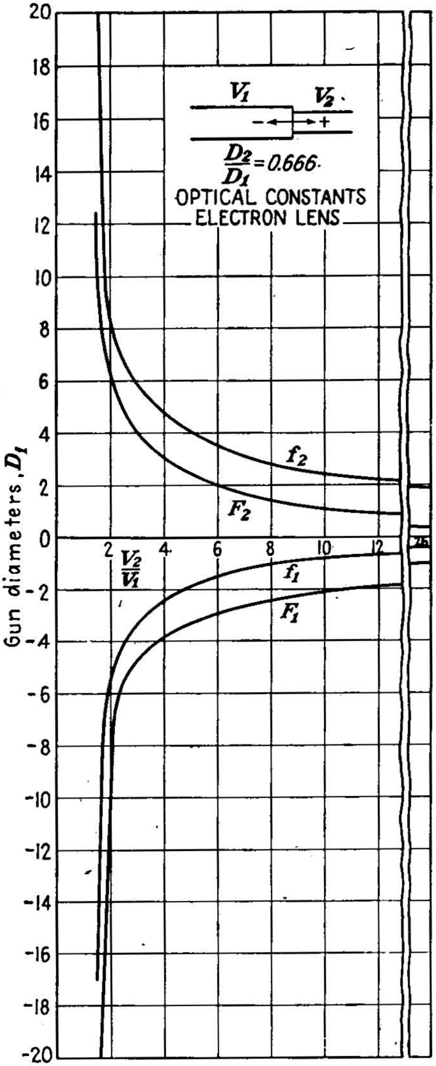


FIG. 13.26.—Optical characteristics of a two-cylinder lens, $D_2/D_1 = 2/3$.

13.7. Optical Characteristics of Lenses. By means of the method of double grids just described in the previous section it is possible to determine experimentally the optical characteristics of lenses over a wide range of voltage ratios. The lens characteristics are completely prescribed

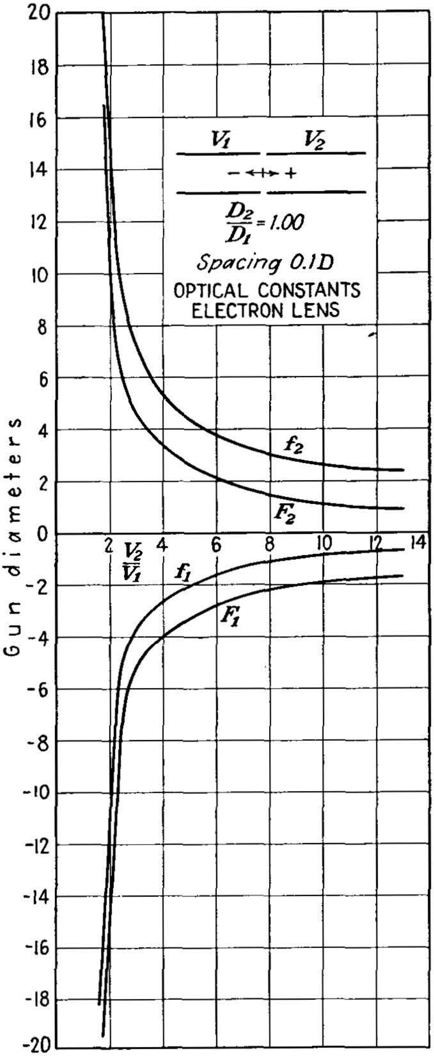


FIG. 13.27.—Optical characteristics of a two-cylinder lens, $D_2/D_1 = 1$.

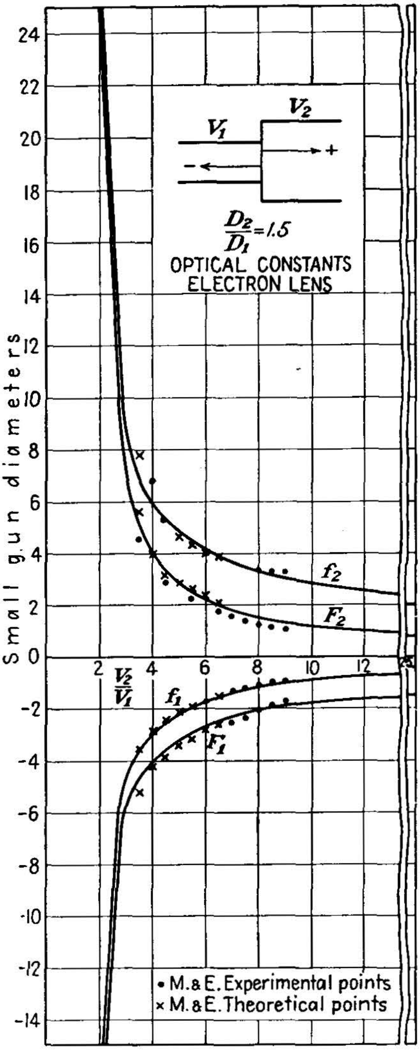


FIG. 13.28.—Optical characteristics of a two-cylinder lens, $D_2/D_1 = 1.5$.

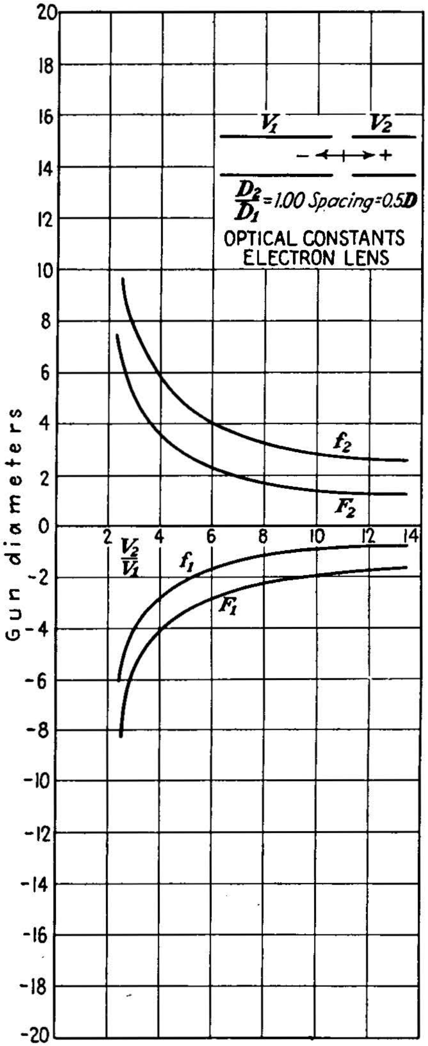


FIG. 13.29.—Optical characteristics of an equal-diameter two-cylinder lens, $S = 0.5D$.

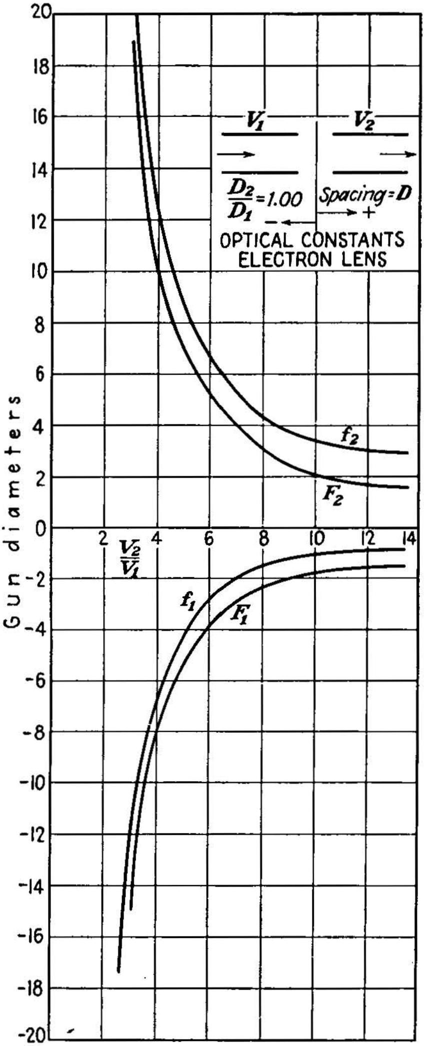


FIG. 13.30.—Optical characteristics of an equal-diameter two-cylinder lens, $S = 1.0D$.

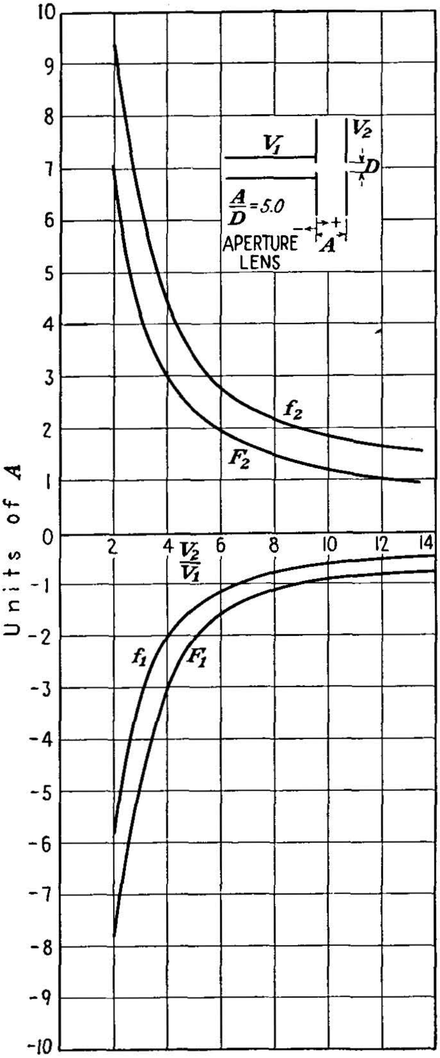


FIG. 13.31.—Optical characteristics of a double-aperture lens, $A/D = 5$.

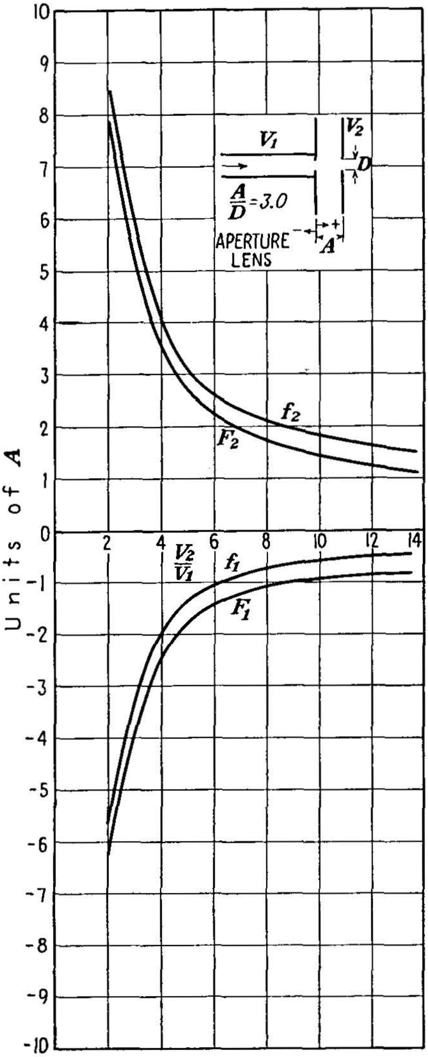


FIG. 13.32.—Optical characteristics of a double-aperture lens, $A/D = 3$.

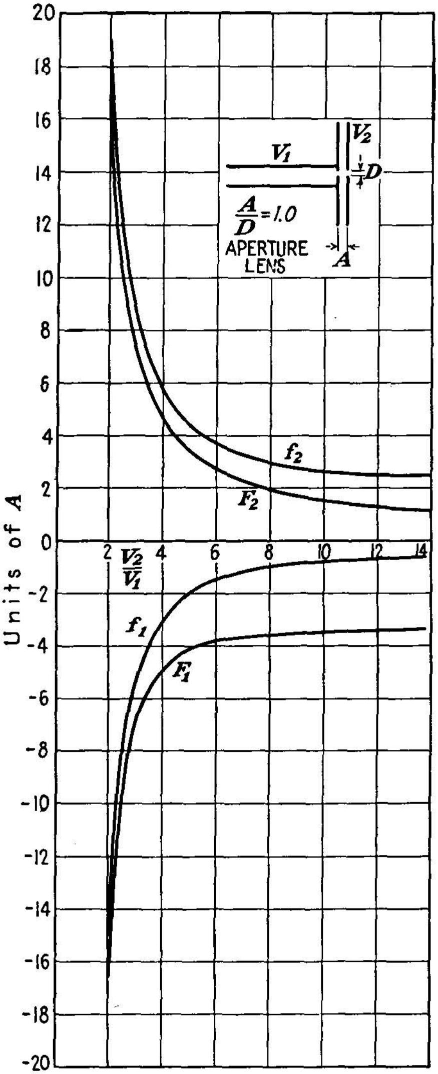


FIG. 13.33.—Optical characteristics of a double-aperture lens, $A/D = 1$.

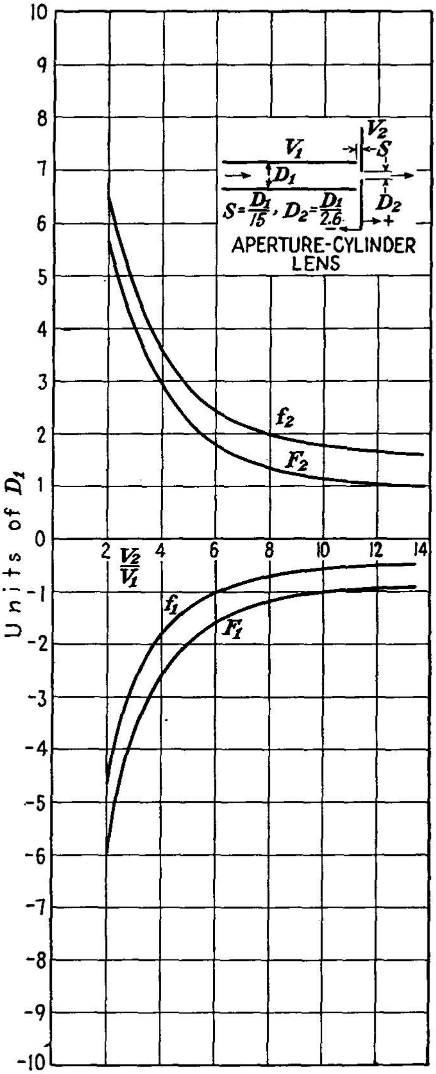


FIG. 13.34.—Optical characteristics of an aperture-cylinder lens.

if the two focal lengths and the location of the two principal planes are given as a function of the voltage ratio. In Figs. 13.26 to 13.34 there are plotted the focal lengths and location of the focal points of nine of the commonest electrostatic electron lenses.[1]

Examination of the lens-characteristic curves of Figs. 13.26 to 13.34 reveals that all *these* lenses have the following characteristics in common:

1. Focal lengths are always uniformly decreasing functions of voltage ratio.
2. Principal planes always lie on the low-voltage side of the lens.
3. Principal planes are crossed, with the exception of the large-diameter aperture lens, *i.e.*, the first principal plane lies between the second principal plane and the lens center on the low-voltage side of the lens.
4. Focal length in the direction of increasing potential is always greater than the focal length in the other direction.
5. The position of the principal planes does not change much with voltage ratio except at very low values.

A comparison of the focal properties of the specific lenses yields the following observations:

1. The focal length of two-diameter cylinder lenses increases, *i.e.*, the lens grows weaker for all but the highest voltage ratios, as the ratio of second to first cylinder diameter increases.
2. The focal length of equal-diameter cylinder lenses increases, *i.e.*, the lens grows weaker, as the axial spacing of the cylinders increases. The change is small for small spacings but increases rapidly as the spacing is increased.
3. The focal length of aperture lenses increases, *i.e.*, the lens grows weaker, as the aperture diameter increases. The change is small for small diameters but increases rapidly as the diameter increases.
4. Aperture lenses have for the most part shorter focal lengths than cylinder lenses if aperture spacing be taken equal to first cylinder diameter as a unit of length.
5. The cylinder-aperture lens has the shortest focal length of all lenses tested.
6. The equal-diameter lens with axial spacing of one diameter has the longest focal length of all the lenses in this collection.

After all the comparisons between lenses have been made, it must be admitted that there is not much choice between them, for the focal

---

[1] SPANGENBERG, KARL, and L. M. FIELD, The Measured Characteristics of Some Electrostatic Electron Lenses, *Elec. Commun.*, vol. 21 (No. 3), pp. 194–204, 1943.

length of any of the lenses can be adjusted at will by simply changing the voltage ratio applied to the electrodes. Cylinder lenses are usually preferred to aperture lenses as objective lenses because the electron beam is shielded from the effect of any charges that may accumulate on the glass walls of the vacuum envelope. They also permit the use of limiting apertures within the cylinders to reduce the beam diameter.

13.8. Calculation of Lens Characteristics. In general, electrostatic lenses are not amenable to extensive analytical treatment. The lenses that can be calculated cannot readily be built, and vice versa. It is of interest, however, to confirm the results observed in the previous section by noting the results of such cases as have been completely solved. A complete solution of the lens for which the axial potential is of the form

$$V_0(z) = A \exp\left(\frac{4\sqrt{3}}{3} R \arctan \frac{z}{a}\right) \tag{13.102}$$

has been given.[1] In this expression for axial potential the constants A and R are related to the initial and final values of potential by

$$R = \frac{3}{4\sqrt{3}\,\pi} \ln \frac{V_2}{V_1} \tag{13.103}$$

and

$$A = V_1 \exp\left(\frac{2\sqrt{3}\,\pi}{3}\right) \tag{13.104}$$

This axial-potential distribution is not greatly different from that found in two-diameter cylinder lenses, as may be seen in Fig. 13.35, in which there is plotted the potential distribution $V_0(z) = \exp(\arctan z)$.

The general solution of Eq. (13.46) with the axial-potential distribution of Eq. (13.102) is

$$r(z) = \sqrt{1 + \left(\frac{z}{a}\right)^2} \exp\left(-\frac{\sqrt{3}\,R}{3} \arctan \frac{z}{a}\right) \left[C_1 \sin\left(w \operatorname{arc\,cot} \frac{z}{a}\right) + C_2 \left(w \operatorname{arc\,cot} \frac{z}{a}\right)\right] \tag{13.105}$$

where $w = \sqrt{1 + R^2}$. From this the focal lengths are found to be

$$f_1 = \frac{a\epsilon^{\frac{\sqrt{3}R\pi}{3w}}}{\sin\left(\frac{\pi}{w}\right)} \tag{13.106}$$

---

[1] HUTTER, R. G. E., Rigorous Treatment of the Electrostatic Immersion Lens Whose Axial Potential Distribution Is Given by $\phi(z) = \phi_2 \exp(\arctan z)$, *Jour. Appl. Phys.*, vol. 16, pp. 678–699, November, 1945.

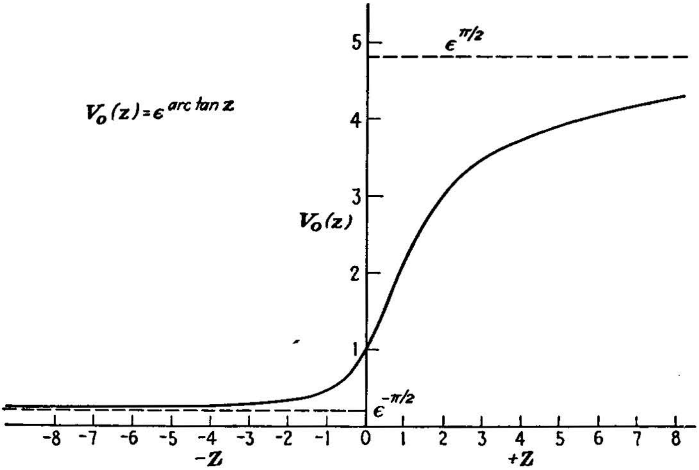


FIG. 13.35.—Plot of the axial potential, $V_0(z) = \exp$ (arctan).

$$f_2 = \frac{-a\epsilon^{\frac{\sqrt{3}R\pi}{3w}}}{\sin\left(\frac{\pi}{w}\right)} \tag{13.107}$$

The focal lengths as a function of voltage ratio are given in Fig. 13.36. The position of the principal planes is given by

$$P_1 = a\,\frac{\epsilon^{\frac{-\sqrt{3}R\pi}{3w}} + \cos\left(\frac{\pi}{w}\right)}{\sin\left(\frac{\pi}{w}\right)} \tag{13.108}$$

$$P_2 = a\,\frac{\epsilon^{\frac{\sqrt{3}R\pi}{3w}} - \cos\left(\frac{\pi}{w}\right)}{\sin\left(\frac{\pi}{w}\right)} \tag{13.109}$$

The position of these is also plotted in Fig. 13.36. All the properties of lenses observed experimentally are confirmed by this example. It is of interest to note that the ratio of focal lengths as given by the square root of the electrode potential ratio holds only to a ratio of about 6 in this case. At a voltage ratio of 16 the ratio of the focal lengths is 3.7 instead of 4. This departure from the theoretical value occurs because the lens is a very strong one and for moderately large electrode-potential ratios the principal rays cross the axis within the region of potential variation, whereas in the derivation of Eq. (13.75) it was assumed that the rays crossed the axis outside the region of appreciable potential variation.

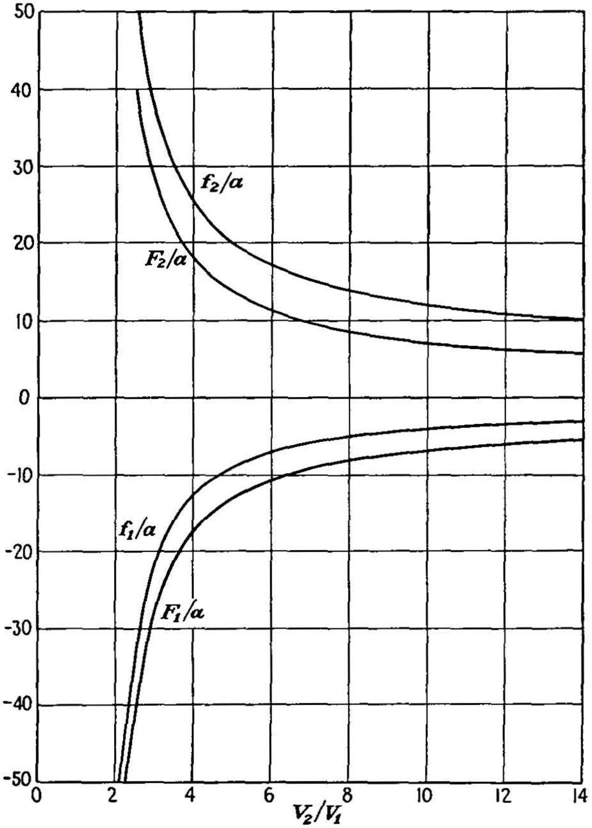


FIG. 13.36.—Optical characteristics of the Hutter lens.

13.9. ***P-Q*** **Curves of Lenses.** The optical characteristics of lenses presented in the previous section do not tell a great deal directly about the lens performance. These optical characteristics are parameters that enter into the calculation of image distance corresponding to object distance for any voltage ratio. The lens parameters disappear in the calculation, and only the associated object and image distance and corresponding magnification remain. It would therefore seem logical to present lens characteristics in such a way that the resultant properties and not the construction parameters were revealed. This has been done in a type of curve that will be referred to as the P-Q curves of a lens. The significance of the letters is that the curves present associated object distance P and image distance Q, as in Fig. 13.21, and corresponding lateral magnification M for any voltage ratio. The object distance, image distance, and lateral magnification are calculated by means of Eqs. (13.68) and (13.69).

The *P*-*Q* curves of the nine common lenses discussed before are shown in Figs. 13.37 to 13.45. In these curves there are shown contours of constant lateral magnification and constant voltage ratio against axes of object and image distance. The *P*-*Q* curves are in effect a graphical presentation of the solution to all the first-order image problems associated with the lens. The advantage of this presentation is that it gives design data immediately, without calculation. The presentation is further sufficiently explanatory so that it can be used without a complete understanding of the theory of electron optics.

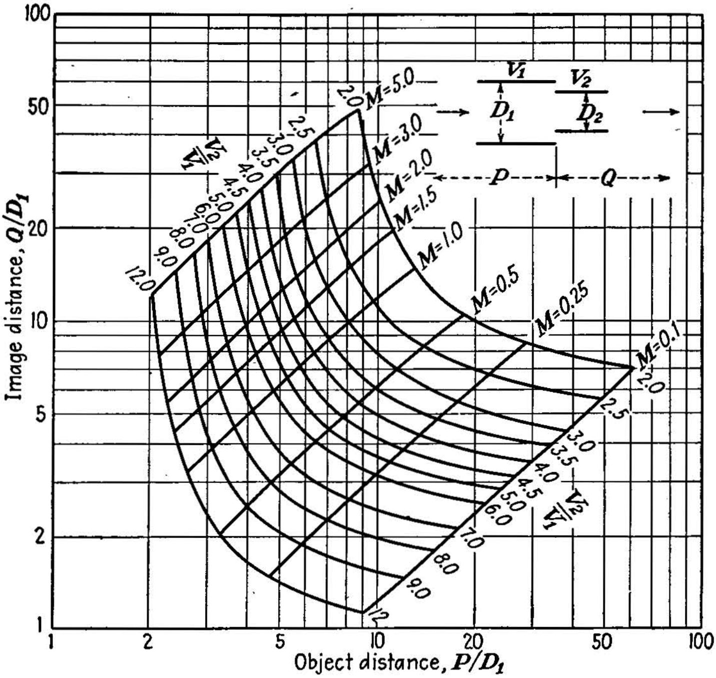


FIG. 13.37.—*P*-*Q* curves of a two-cylinder lens, $D_2/D_1 = 2/3$.

A study of the *P*-*Q* curves of Figs. 13.37 to 13.45 reveals the following characteristics as being common to all lenses:

1. As object distance is increased at a given voltage ratio, the corresponding image distance decreases, as does also the magnification.
2. For a given object distance the image distance and magnification decrease as the voltage ratio is increased.
3. In any lens there is a minimum object distance that can be used at any given voltage ratio. This minimum object distance is the

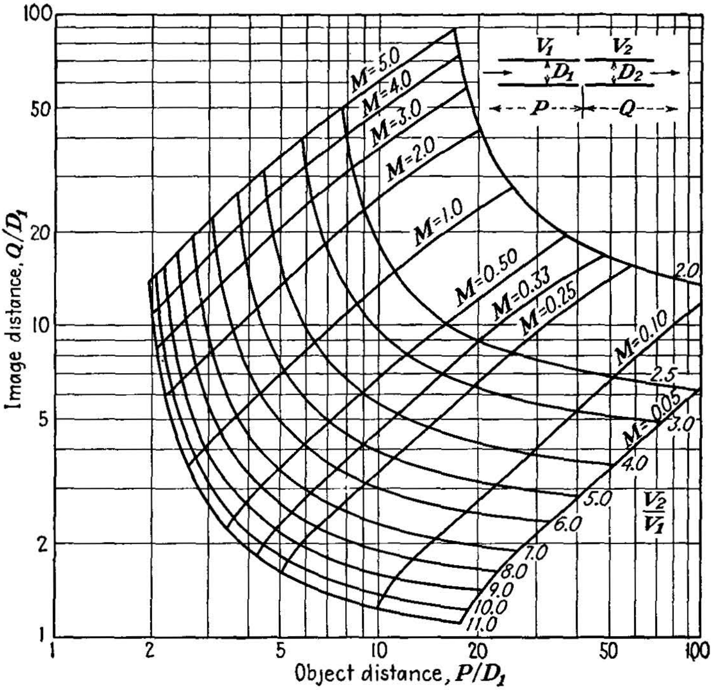


FIG. 13.38.—*P-Q* curves of a two-cylinder lens, $D_2/D_1 = 1$, $S = 0.1D_1$.

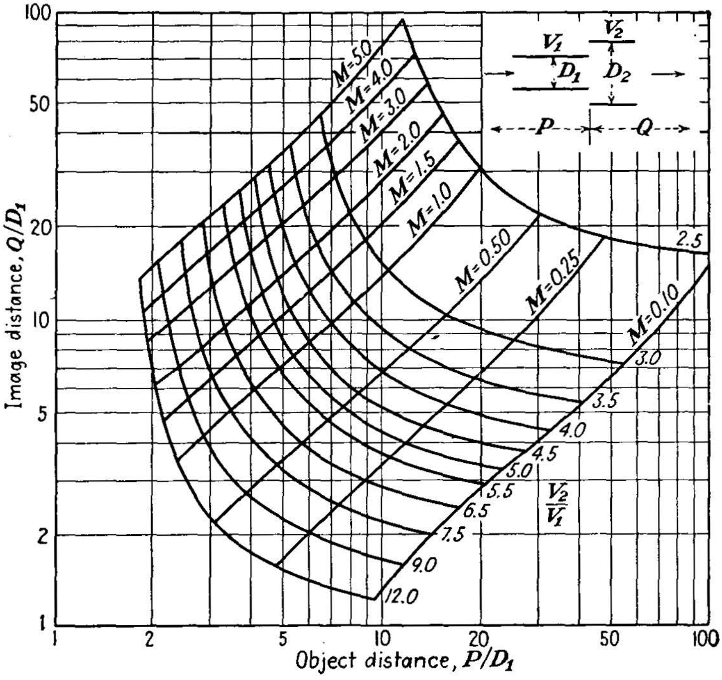


FIG. 13.39.—*P-Q* curves of a two-cylinder lens, $D_2/D_1 = 1.5$.

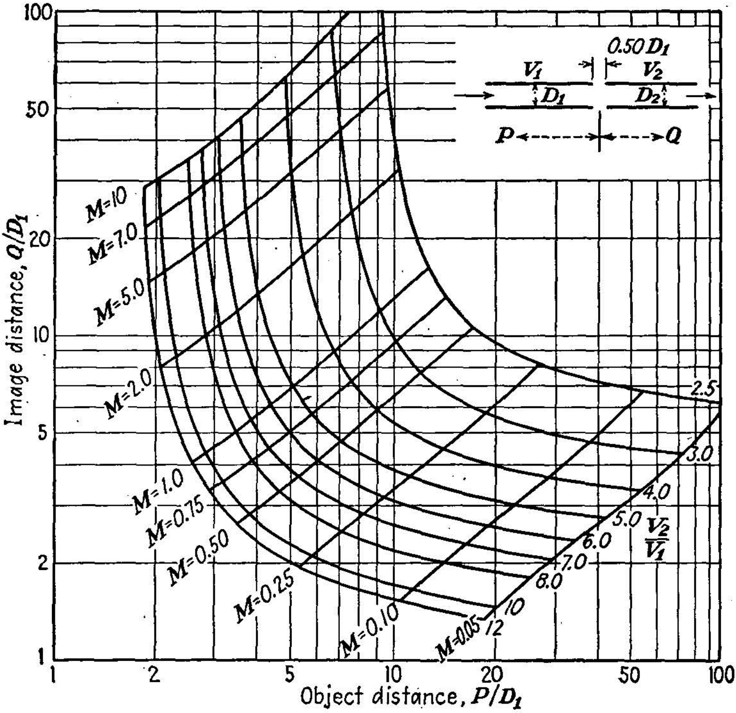


FIG. 13.40.—*P-Q* curves of a two-cylinder lens, $D_2/D_1 = 1$, $S = 0.5D_1$.

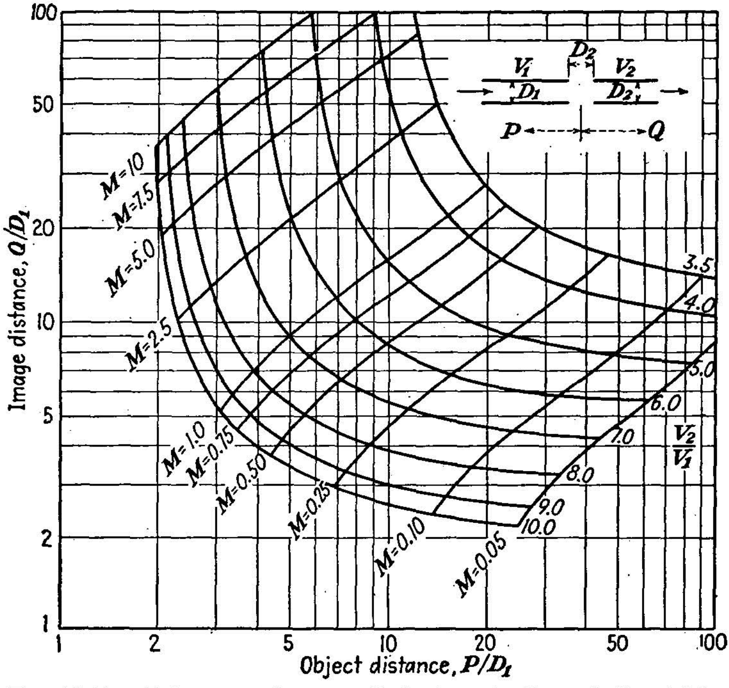


FIG. 13.41.—*P-Q* curves of a two-cylinder lens, $D_2/D_1 = 1$, $S = 1.0D_1$.

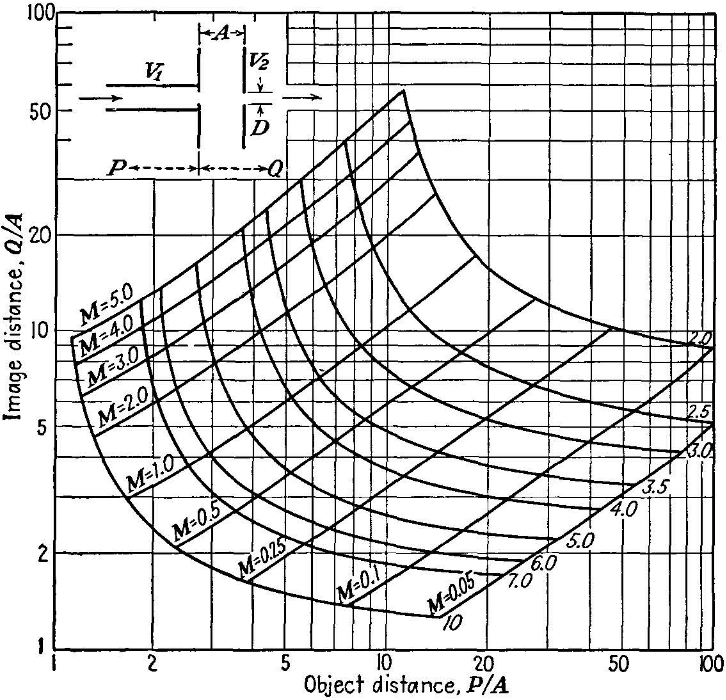


FIG. 13.42.—*P*-*Q* curves of a double-aperture lens, $A/D = 5$.

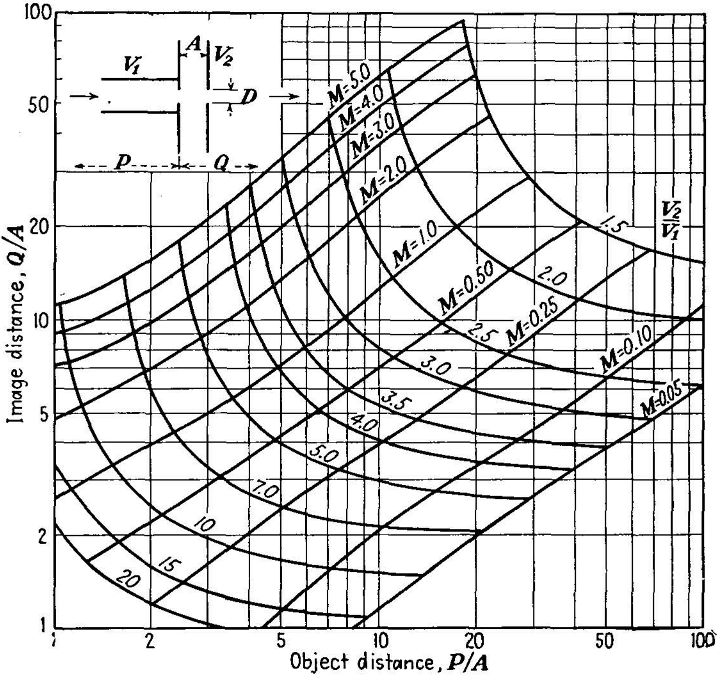


FIG. 13.43.—*P*-*Q* curves of a double-aperture lens, $A/D = 3$.

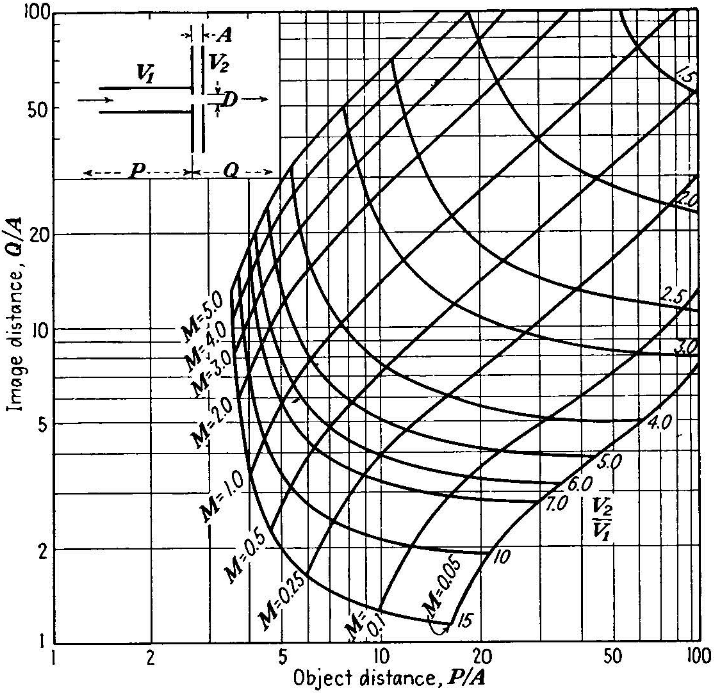


FIG. 13.44.—*P*-*Q* curves of a double-aperture lens, $A = D = 1$.

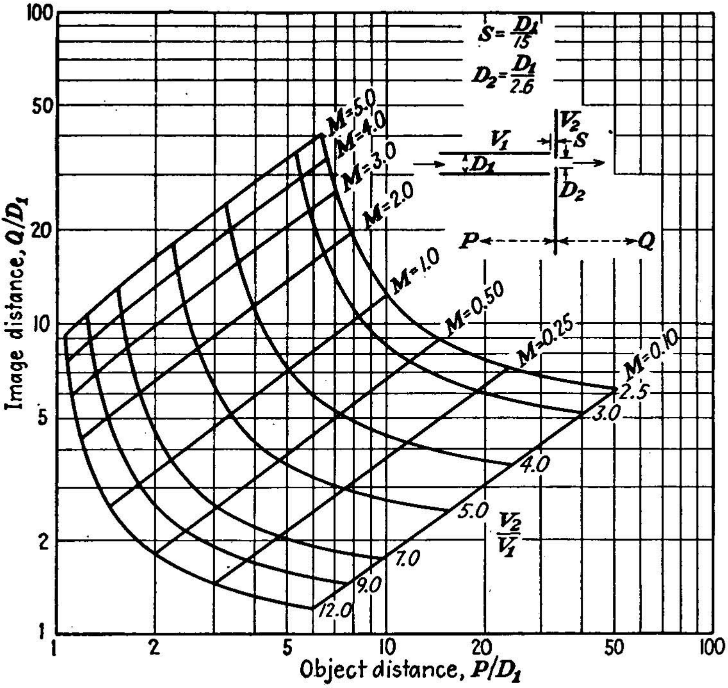


FIG. 13.45.—*P*-*Q* curves of a cylinder-aperture lens.

first focal length of the lens, plus the displacement of the first principal plane from the lens center.

The outstanding magnification characteristics of electron lenses as observed from the P-Q curves are as follows:

1. The contours of constant magnification are approximately straight lines, with a slope of 1. This is exactly the case for thin lenses.
2. An approximate universal magnification formula that fits all lenses shown is

   $$M = k\frac{Q}{P} \tag{13.110}$$

   where P and Q are the object and image distance, respectively. Values of the constant for the lenses tested are

   Cylinder lens: $\frac{d_2}{d_1} = 0.667 \qquad k = 0.82$
   $\frac{d_2}{d_1} = 1 \qquad S = 0.1 \qquad k = 0.78$
   $\frac{d_2}{d_1} = 1.5 \qquad k = 0.76$
   $\frac{d_2}{d_1} = 1 \qquad S = 0.5 \qquad k = 0.80$
   $\frac{d_2}{d_1} = 1 \qquad S = 1 \qquad k = 0.60$
   Aperture lens: $\frac{a}{d} = 5 \qquad k = 0.95$
   $\frac{a}{d} = 3 \qquad k = 0.80$
   $\frac{a}{d} = 1 \qquad k = 0.78$
   Cylinder-aperture lens: $k = 0.82$

   It is seen that, with only two exceptions, the value of the constant is within a few per cent of 0.8.

The observed magnification property is strictly in accordance with theoretical expectations, though the agreement is not at all apparent. From Lagrange's law [Eq. (13.76)] it is expected that the lateral magnification will equal the product of the ratio of image to object distance multiplied by the square root of the reciprocal of the voltage ratio, namely, $M = \frac{Q}{P}\sqrt{\frac{V_1}{V_2}}$. This follows from the fact that the angular magnification is nearly equal to the ratio of object to image distance. The actual contours of constant lateral magnification do not seem to follow this law, but the discrepancy is only an apparent one, not a real one. The apparent discrepancy is due to the fact that the object and image distances are measured from an arbitrary point in the lens, whereas they should be measured from an equivalent thin lens located between the principal planes. If in the above modification of Lagrange's law the distances P and Q are measured from a point midway between the principal planes, then the calculated contours of constant lateral magnification are almost indistinguishable from the measured ones in sample cases that have been tested.

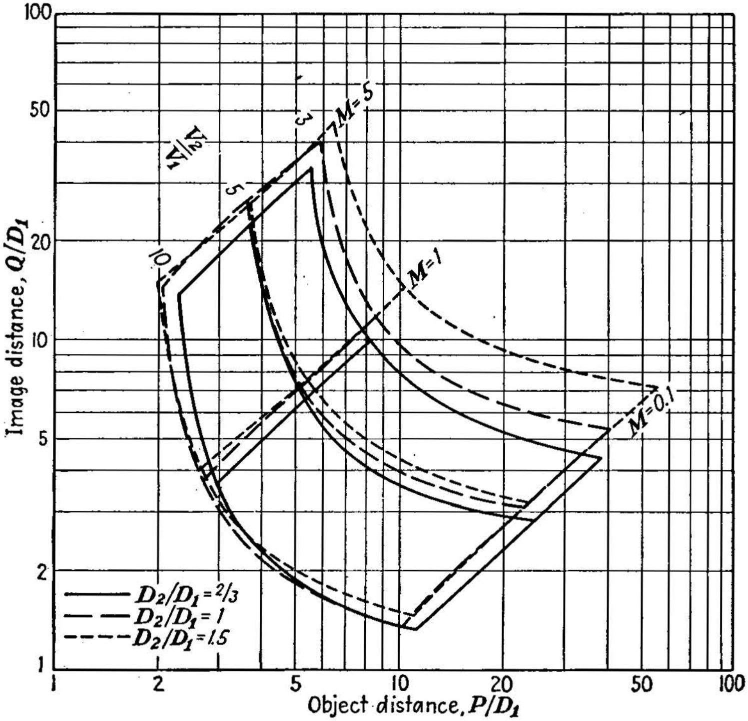


FIG. 13.46.—Comparison of the P-Q curves of two-diameter lenses.

Comparison of Lenses. An important feature of the P-Q curves is that they make possible a comparison of the focal lengths and magnifications of various types of lenses over the whole range of voltage ratios and object distances. In Figs. 13.46 to 13.48 are drawn, for comparison, parts of the complete curves of similar types of lenses.

In Fig. 13.46 are shown the effects of changing the ratio of diameters in a two-diameter cylinder lens. The curves show that, for ratios of diameters between 1 and 1.5, there is not much difference in magnification. For the smaller ratio of diameters the magnification is distinctly more (object distance and voltage ratio held constant). In the vicinity of useful application, say $P = 3$ and $Q = 20$, the voltage ratio required for any ratio of cylinder diameters is about the same.

In Fig. 13.47 is given a similar comparison of equal-diameter cylinder lenses for different spacings between cylinders. This comparison reveals that the magnification of such lenses is about the same for small axial spacings up to about 0.5 diameter and then increases considerably as the spacing is increased (object distance and voltage ratio held constant). For the most part the lens becomes weaker as the axial spacing between cylinders increases.

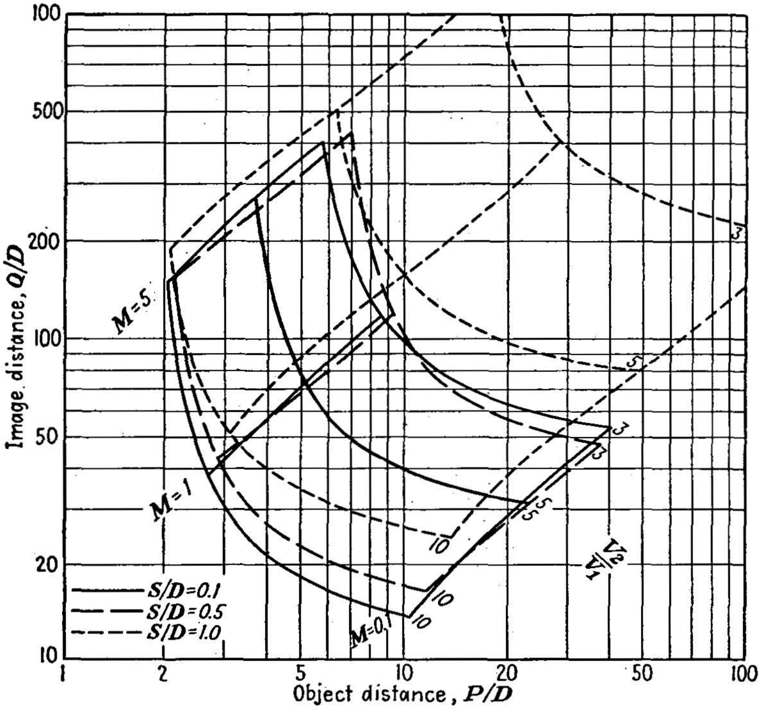


FIG. 13.47.—Comparison of the P-Q curves of equal-diameter lenses.

A comparison of aperture lenses is given in Fig. 13.48. It is a little hard to draw any general comparisons because of the pronounced crossovers in the P-Q characteristics for different lens dimensions. In the vicinity of a short object distance and a long image distance the magnification is not greatly different for different ratios of aperture spacing to diameter. For a fixed object and image distance the voltage ratio necessary to obtain a focus increases as the ratio of aperture spacing to diameter increases.

The Einzel Lens. Another lens which is more or less in a class by itself is the so-called "*Einzel* lens" (after the German word "single," indicating that there is a single value of the limiting potential). The *Einzel* lens consists of three apertures equally spaced, the outer two of which are maintained at the beam potential and the inner of which may

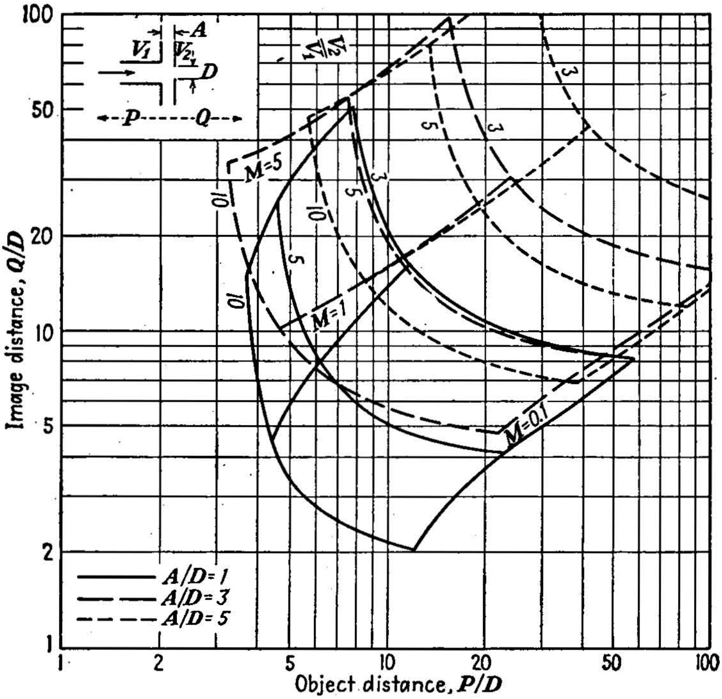


FIG. 13.48.—Comparison of the *P*-*Q* curves of double-aperture lenses.

be at either a higher or a lower potential. The electrode arrangement and potential field of such a lens have already been shown in Fig. 13.16. Such a lens exhibits a convergent action whether the center electrode is more or less positive than the outer electrode. The nature of the focusing characteristics of a special type of *Einzel* lens are shown in Fig. 13.49. The lens exhibits a focal length which decreases as the ratio $\frac{V_2 - V_1}{V_1}$ increases, where V_2 is the inner-electrode potential and V_1 is the outer-electrode potential. For negative values of the same potential ratio the focal length decreases until the center electrode is sufficiently negative to make the saddle point at the center of the lens assume a negative potential. Beyond this point the electrons cannot penetrate the lens but are reflected back. The reflection is of such a nature that the action as the center-electrode potential is made still further negative is first that of a concave mirror and then that of a convex mirror. This change in the nature of the reflection occurs because at first the electrons can penetrate to a point within the lens where the equipotential lines are concave and then as the center electrode becomes more negative they are only able to penetrate slightly into a region where the equipotential lines are convex.[1]

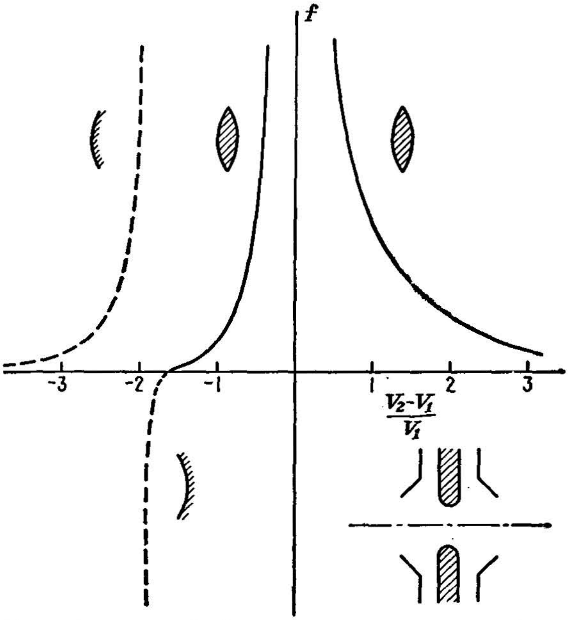


FIG. 13.49.—Focal characteristics of an *Einzel* lens.

13.10. Aberrations. As is the case with physical lenses, electron lenses are not perfect image-forming devices but are subject to a number of distortions or lens errors known as "aberrations." Because of the almost exact analogy that exists between geometrical and electron optics, every one of the aberrations found in light lenses is also found in electrostatic lenses. Thus the terminology of light lenses is directly transferable to electron lenses.

All the lens theory that has been given so far has been a first-order theory. This is the so-called "Gaussian optics." The differential

---

[1] JOHANNSON, H., and O. SCHERZER, Uber die elektrische Elektronen Sammellinse, *Zeit. für Phys.*, vol. 80 (No. 3, 4), pp. 183–192, 1933.

equation of the paraxial electron was obtained by dropping all terms in r and $\frac{dr}{dz}$ of order higher than the first. If the restrictions on the electron under consideration are made more liberal and only terms in r and $\frac{dr}{dz}$ higher than the third are neglected, then the theory of so-called "third-order imagery" is obtained. This third-order theory reveals all the defects in image formation that are encountered.

The differential equation of third-order imagery is obtained by starting again with the general differential equation of motion of Eq. (13.38), then using the first *two terms* of Eqs. (13.21), (13.39), and (13.41), and then neglecting any terms of order higher than the third. The resulting differential equation of motion of an electron of third-order imagery is

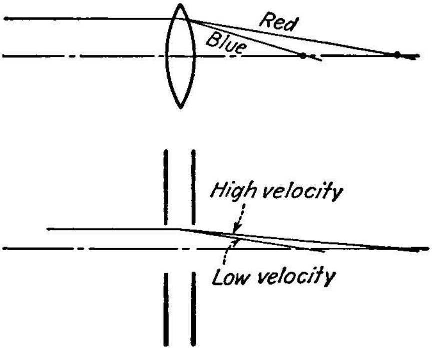


FIG. 13.50.—Chromatic aberration.

$$r'' + \frac{V_0'r'}{2V_0}\left[1 + r^2\left(\frac{V_0''}{4V_0} - \frac{V_0'''}{4V_0'}\right) + r'^2\right] + \frac{V_0''r}{4V}\left[1 + r^2\left(\frac{V_0''}{4V_0} - \frac{V_0^{(4)}}{8V_0''}\right) + r'^2\right] = 0 \quad (13.111)$$

where the primes indicate derivatives with respect to z.

A study of Eq. (13.111) reveals five distinct types of monochromatic aberration possible in electrostatic lenses. The five types are generally classified as coma, astigmatism, curvature of field, distortion of field, and spherical aberration. In addition to these types, chromatic aberration may be present. This makes six defects that are possible with perfect structures and low currents. Distortions due to space charge and imperfections in the electrode structure may also be present. Each of the above defects will now be briefly described.

Chromatic Aberration. This is the well-known effect in geometrical optics that causes light of different wave lengths to have different focal lengths as shown in Fig. 13.50. In electron optics the analogous effect is that electrons with different velocities will focus at different points. In electron lenses the velocity of the electrons varies only to the extent that the velocity of emission is different for different electrons. Since this variation in emission velocity is generally small compared with the accelerating potentials used, the error is not a serious one.

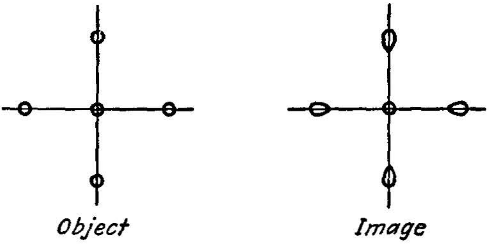


FIG. 13.51.—Coma.

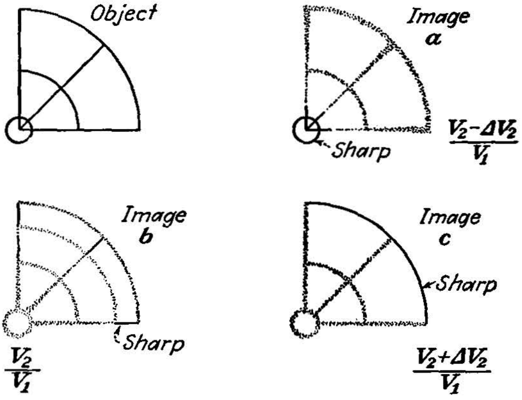


FIG. 13.52.—Astigmatism.

As an example of the effect of chromatic aberration, consider the case of a single-aperture lens for which the focal length is given by

$$f = \frac{4V_0}{V_0'(z_2) - V_0'(z_1)} \tag{13.56}$$

If the different electrons have energies corresponding to different values of V_0, then

$$\frac{df}{dV_0} = \frac{f}{V_0} \tag{13.112}$$

which is really just another way of saying that the focal length depends upon the electron energy; the higher the energy, the lower the change in focal length.

Coma. This is an extraaxial aberration, *i.e.*, one that appears only for images and objects not lying on the axis of the lens system. The effect is due to the fact that different circular zones about the axis have different magnification. As a result, a set of concentric circles off the axis is imaged as a set of slightly distorted circles that are not concentric but that have a drop-shaped envelope with the tail pointed away from the axis. The type of distortion resulting is shown in Fig. 13.51. The effect is lessened if a smaller portion of the lens center is used, but this reduces the amount of light or beam current and may not always be desirable.

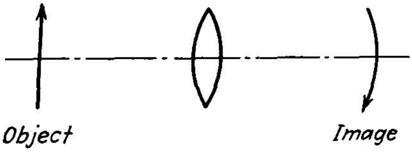


FIG. 13.53.—Curvature of field.

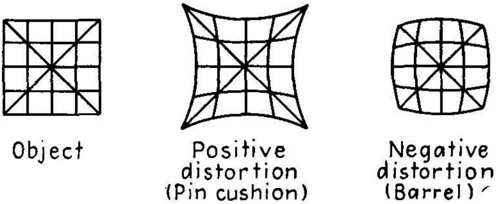


FIG. 13.54.—Distortion of field.

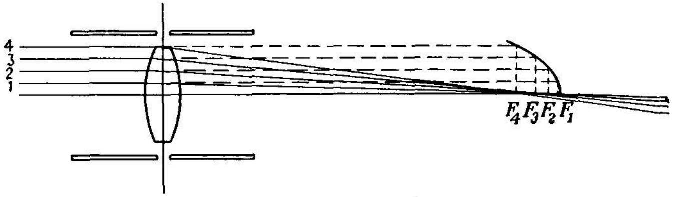


FIG. 13.55.—Spherical aberration.

Astigmatism. This is a well-known effect in geometrical optics. The effect is that, in any object off the axis, lines directed toward the axis have a different focal length from those at right angles to these. A compromise focus gives an image of least diffusion in which neither of the lines is clear. This effect is illustrated in Fig. 13.52. As focusing voltage is changed, a focus is first obtained at the center of the image, then along a radial line, and then along circumferential lines. This is another of the extraaxial effects. In cylinder lenses, if the electrodes are slightly elliptical the beam focus will be a short line instead of a spot. As the focusing voltage is adjusted, the short line will become a fuzzy spot and then a sharp short line again, but at right angles to its former position.

Curvature of Field. This lens defect usually accompanies but is more pronounced than astigmatism. The effect evidences itself by the fact that an object lying in a plane perpendicular to the axis has an image which does not lie on a plane but which lies on a slightly curved surface, a surface of revolution about the axis approximately spherical which is concave toward the lens. The result of this form of aberration is that an object consisting of a set of circles concentric about the axis gives an image which is sharp at only one radial distance. If the image plane is adjusted to make the center sharp, then the outside circle will be fuzzy, and vice versa. This lens defect is illustrated in Fig. 13.53.

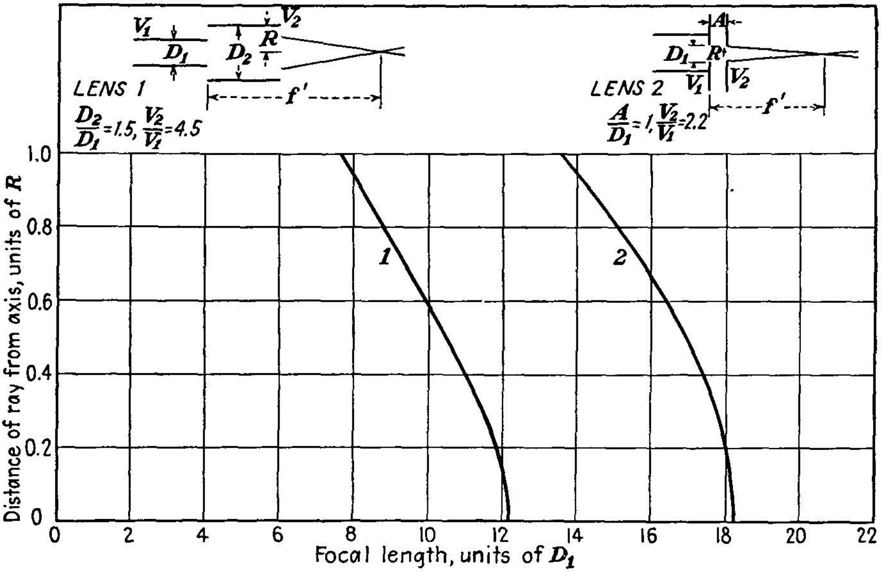


FIG. 13.56.—Focal length as a function of the radial coordinate in the lens.

Distortion of Field. This defect is due to variations of the linear magnification with radial distance in the lens. If the object is a small checkerboard, then the distortion evidences itself by giving rise to pincushion- and barrel-shaped images shown in Fig. 13.54. If the magnification increases with radial distance, it is considered positive and the pincushion-shaped image results. If the magnification decreases with radial distance, it is considered negative and the barrel-shaped image results.

Spherical Aberration. This is another lens defect well known in geometrical optics. Basically the effect is that rays entering the lens parallel to the axis have a focal length which changes with the radial distance at which they pass through the lens, as shown in Fig. 13.55. The focal length as a function of radial position in the lens can be measured by any of the experimental methods previously described and yields curves such as those of Fig. 13.56. In the curves shown the focal length reduces as the radial distance increases. This is known as "positive" spherical aberration and is the kind invariably encountered in electron lenses. The focal length is seen to decrease slowly at first and then more rapidly. This is an axial aberration that has the effect of giving a spot focus instead of a point focus. The minimum size of spot that can be obtained for any lens aperture increases with the radius of the aperture. A typical curve illustrating this effect is shown in Fig. 13.57.

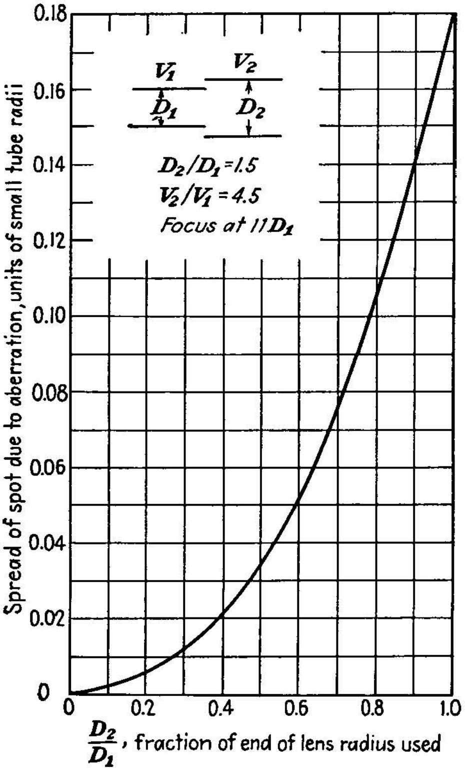


FIG. 13.57.—Minimum spot size as a function of aperture radius.

Spherical aberration is one of the most serious of the various aberrations. It is always present and in electron lenses is invariably positive. In physical lenses it is possible to combine elements with equal positive and negative spherical aberration to obtain a lens that is free of this effect. In electrostatic lenses it is possible only to reduce this effect, as, for instance, by using a two-diameter cylinder lens with the high-potential cylinder having a smaller radius than the lower.[1] It is also possible to reduce spherical aberration in aperture lenses by the use of specially shaped thick electrodes with curved surfaces corresponding to the equipotential[2]

$$V(r,z) = (a \sin kz + b \cos kz)J_0(ikr) \tag{13.113}$$

It has also been shown that a symmetrical lens with an axial-potential variation given by[3]

$$V_0(z) = V_1(1 + A\epsilon^{-Bz^2}) \tag{13.114}$$

has minimum spherical aberration. Unfortunately, it is very difficult to build a lens having such an axial variation of potential since the electrode structure required is not a practical structure at all. In general, the spherical aberration associated with a given lens structure may be reduced by simply eliminating sharp corners and edges on the electrodes. A rounding of corners and edges eliminates large gradients, which seem to contribute considerably to the lens defects.

Other Lens Defects. In addition to the above optical defects, electrostatic lenses are subject to a few ills to which physical lenses do not fall heir. The space-charge mutual repulsion between electrons prevents electron beams from coming to a point focus and in general exhibits the same effects as spherical aberration. This subject will be given an analytical treatment in the chapter on Cathode-ray Tubes. In addition, imperfections in the electrode structure will give rise to some remarkable distortions. In lenses with small apertures, if the plane of the apertures is not perpendicular to the axis, the beam will focus into a tadpole-shaped figure. Modern techniques are, however, sufficiently good so that distortions resulting from electrode imperfections seldom appear in commercial tubes.

---

[1] KLEMPERER and WRIGHT, *op. cit.*

[2] GRAY, *op. cit.*

[3] SCHERZER, O., Die Schwache elektrische Einsellinse geringster spharischer Aberration, *Zeit. für Phys.*, vol. 1, pp. 23–26, June, 1936.

CHAPTER 14

MAGNETIC LENSES

14.1. Focusing Action of Axial Magnetic Fields. Electron beams can be focused with magnetic as well as with electric fields, though the analogy with optics is not so readily established. Reference has already been made to one type of magnetic focusing In Sec. 6.6 there was discussed the case of a long uniform magnetic field parallel to an axis. Electrons leaving a point on the axis with their principal component of velocity directed parallel to the axis move out with helical paths of approximately the same pitch and come to a focus farther along the axis. This action is shown in Fig. 14.1. The motion of the individual electrons is a combination of a linear translation parallel to the axis and a circular motion in a plane perpendicular to the axis, giving rise to a helical path. The radius of the circular component of travel is given by Eq. (6.67). It is proportional to the radial component of velocity and inversely to the magnetic-flux density at the starting point. The focal length (pitch of the helices) is given by Eq. (6 68). The focal length depends upon the axial component of velocity directly and upon the magnetic-flux density inversely. Thus the focal lengths of the different electrons are within 1½ per cent for initial angles with the axis that are less than 10 deg, and a pretty good focus is obtained. The radial and angular displacement associated with this motion is shown in Fig. 14.2. The radial displacement is sinusoidal in form. The angular displacement is uniformly increasing with distance.

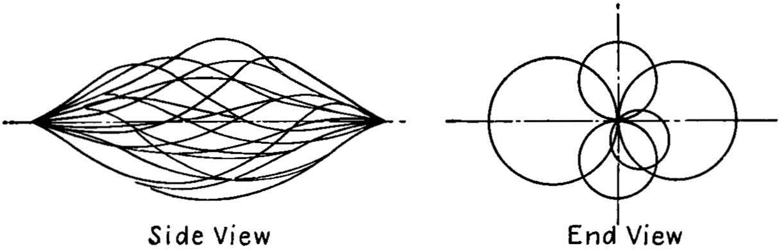


FIG. 14.1.—Helical electron paths in a uniform magnetic field.

The action of the long field parallel to the axis is more or less typical of the action of all magnetic lenses. All magnetic lenses depend upon a component of magnetic field parallel to the axis. The electron paths start from a point on the axis and return to it at a later point. In doing so the electrons move in a plane through the axis, which rotates continuously as the electron passes through the magnetic field.

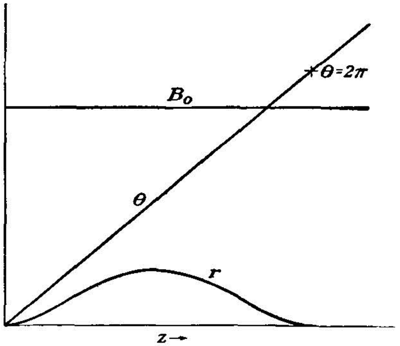


FIG. 14.2.—Radial and axial displacel ment of an electron in a uniform axia- field.

The action of a short magnetic lens may be understood by considering a fictitious case in which the magnetic field is short but uniform and parallel to the axis, as shown in Fig. 14.3. In this case an electron leaving the axis will move in a straight line at a constant velocity until it enters the magnetic field. When this happens, from Eq. (6.70), the radial component of velocity will react with the axial component of magnetic flux to produce an angular component of force. This imparts a twist to the electron path, and at the same time the angular (θ) component of velocity developed will react with the axial component of field to produce a radial component of force directed toward the axis, that serves to focus the electron and bring it back to the axis. The radial and angular displacements along the axis for this case are shown in Fig. 14.3.

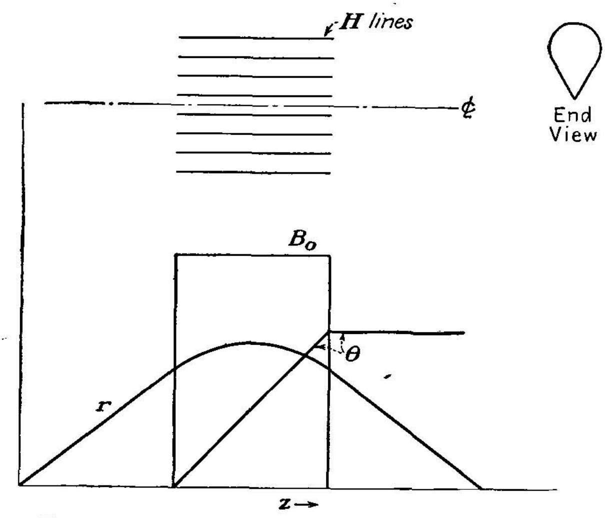


FIG. 14.3.—Action of a short fictitious magnetic lens.

In an actual short magnetic lens the magnetic-flux lines cannot terminate abruptly as in the fictitious example above. A typical field, such as might be produced by a circular current coil, is shown in Fig. 14.4. Fundamentally, the action here is the same as that described in the previous paragraph except that the action is more uniform. In particular, the angular displacement has no sharp corners but is continuous in slope. Note that the action in each of the three cases cited is apparently a function of the axial component of magnetic flux. It will be shown in subsequent sections that this is indeed the case and that the entire action of a magnetic lens can be described in terms of its axial variation of its axial component of magnetic field. Note further that if the magnetic-field strength in the last two cases cited is not of just the right value the electron may not hit the axis even though it is deflected back toward it. This is not the case with electrostatic lenses and means that magnetic lenses have an extra type of aberration to which the electrostatic lens is not subject.

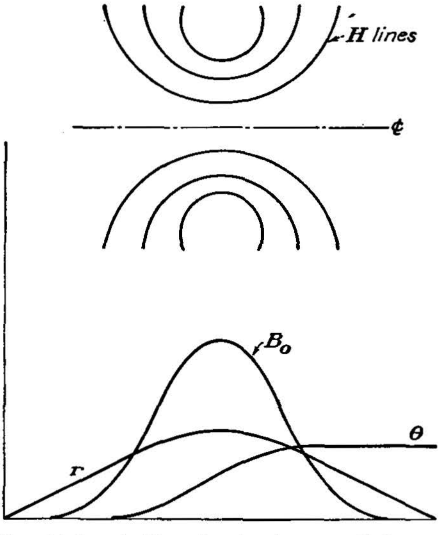


FIG. 14.4.—Action of a short magnetic lens.

14.2. Magnetic Fields with Rotational Symmetry. In the case of electrostatic fields with rotational symmetry it was found that the potential at any point in the field could be expressed in terms of the axial potential and its derivatives by means of a series expansion. This series proved very useful in studying the behavior of paraxial electrons. A similar situation applies to magnetic fields. The z component of a magnetic field obeys Laplace's law provided that the region under consideration does not include any current flow. If attention is restricted to the vicinity of the axis of a field produced by something like a circular coil, there is no current flow except that represented by the electron beam and this is so weak in terms of the magnetic field it produces that it can be neglected. The axial component of magnetic flux is of most importance; Laplace's equation for it is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\,\frac{\partial B_z}{\partial r}\right) + \frac{\partial^2 B_z}{\partial z^2} = 0 \tag{14.1}$$

This can be solved for B_z in the form of a power series in r by the technique used to obtain an expansion for electrostatic potential. The result is

$$B_z(r,z) = B_0 - \frac{r^2}{2^2} B_0'' + \frac{r^4}{2^2 \cdot 4^2} B_0^{(4)} + \cdots \tag{14.2}$$

where $B_0 = B_z(0,z)$, the axial value of the axial component, and the primes indicate derivatives with respect to axial distance z. This series can also be written as the summation

$$B_z(r,z) = \sum_{n=1}^{\infty} (-1)^{n+1} \left(\frac{r}{2}\right)^{2n-2} \frac{B_0^{(2n-2)}}{[(n-1)!]^2} \tag{14.3}$$

Magnetic fields of rotational symmetry will not have an angular component of flux but will have radial and axial components of flux. The radial and axial components of flux are related by the fact that the net outward flux over any small volume not containing current is zero. Mathematically, this is stated by saying that the *divergence* of magnetic flux is zero, and this condition is expressed by

$$\nabla \cdot \boldsymbol{B} = 0 \tag{14.4}$$

or

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_r) + \frac{\partial B_z}{\partial z} = 0 \tag{14.5}$$

Equation (14.4) is simply a shorthand vector notation for the relation of Eq. (14.5). When the series of Eq. (14.2) is substituted into Eq. (14.5) and this solved for B_r, there results

$$B_r = -\frac{r}{2} B_0' + \frac{r^3}{4 \cdot 2^2} B_0''' - \frac{r^5}{6 \cdot 2^2 \cdot 4^2} B_0^{(5)} + \cdots \tag{14.6}$$

in which the constant of integration has been set equal to zero because B_r is an odd function of r. This series may also be written as the summation

$$B_r = \sum_{n=1}^{\infty} \frac{(-1)^n r^{2n-1} B_0^{(2n-1)}}{2n[(n-1)!]^2 2^{2n-2}} \tag{14.7}$$

14.3. Electron Motion in a Magnetic Field Expressed in Cylindrical Coordinates. The equations of motion of an electron in a magnetic field as expressed in cylindrical coordinates have previously been given in Eq. (6.70) but will be repeated here for convenience of reference. In general, the force on a charged particle moving in a magnetic field is given by

$$\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B} \tag{14.8}$$

where $\boldsymbol{F}$, $\boldsymbol{v}$, and $\boldsymbol{B}$ are vector quantities representing force, velocity, and magnetic-flux density, respectively, and q is the charge of the particle. The symbol $\times$ indicates the so-called "vector product," which is a shorthand notation to indicate that the product lies at right angles to the plane of the vectors being multiplied and has the direction a right-handed screw would advance if the slot in its head were turned from alignment with the first to alignment with the second. Furthermore, the resultant vector has a magnitude equal to the product of the magnitude of the vectors being multiplied and the sine of the angle between them. Upon applying Newton's second law and the fact that the electron charge is $-e$, Eq. (14.8) becomes

$$m\boldsymbol{a} = e\boldsymbol{B} \times \boldsymbol{v} \tag{14.9}$$

since $\boldsymbol{A} \times \boldsymbol{B} = -\boldsymbol{B} \times \boldsymbol{A}$. In Eq. (14.9) $\boldsymbol{a}$ is acceleration and m is mass. Expanded in component form, this becomes three differential equations as follows:

$$\frac{m}{e}[\ddot{r} - r\dot{\theta}^2] = B_\theta \dot{z} - B_z r\dot{\theta} \tag{14.10}$$

$$\frac{m}{e}\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = B_z\dot{r} - B_r\dot{z} \tag{14.11}$$

$$\frac{m}{e}\ddot{z} = B_r r\dot{\theta} - B_\theta \dot{r} \tag{14.12}$$

in which the dots above the component variables indicate derivatives with respect to time.

14.4. Differential Equations of Motion of the Paraxial Electron. The set of component equations can be greatly simplified to yield the case of an electron in a magnetic field of rotational symmetry, moving close to the axis and making a small angle with it. In the first place, the rotational symmetry of field means that the angular (θ) component of magnetic flux is zero. The first terms of the series expansions of Eqs. (14.2) and (14.6) can then be substituted for the other components of magnetic flux; and when terms of order r^2 and higher are neglected, tremendous simplification results. Equation (14.11) can be integrated once with the above substitutions to give

$$\dot{\theta} = \frac{e}{m}\frac{B_0}{2} \tag{14.13}$$

the constant of integration being zero since the angular velocity is zero when the magnetic field is zero. This rather remarkable equation states that the angular velocity is proportional only to the axial component of magnetic field. Applying these substitutions to Eq. (14.10),

$$\ddot{r} = -r\left(\frac{e}{m}\frac{B_0}{2}\right)^2 \tag{14.14}$$

Equation (14.12) reduces to

$$\ddot{z} = 0 \tag{14.15}$$

which is approximate only to first order, of course, but is reasonable since there is no electric field contributing to the motion.

Time may be eliminated from Eq. (14.14) by using the approximate relations

$$v = \dot{z} \tag{14.16}$$

and

$$\ddot{r} = \frac{2Ve}{m}\frac{d^2r}{dz^2} \tag{14.17}$$

Substitution of these values into Eq. (14.14) yields

$$\frac{d^2r}{dz^2} + \frac{e}{8mV} B_0{}^2 r = 0 \tag{14.18}$$

Evaluating the constant of the second term numerically,

$$\frac{d^2r}{dz^2} + 2.20 \times 10^{10} \frac{B_0{}^2 r}{V} = 0 \tag{14.19}$$

for rationalized mks units. This expression is similar to the reduced form [Eq. (13.46)] of the paraxial differential equation for the electrostatic case. By the procedures indicated above the electron motion has been separated into radial and angular components. In most focusing problems the radial component may be treated alone without regard for the angle. It need be remembered only that the plane of the electron rotates progressively as the electron moves through the lens.

14.5. Focusing Properties of Magnetic Lenses. *General.* By exactly the same reasoning and process as that used in Sec. 13.4 the focal length of a thin magnetic lens can be deduced from Eq. (14.18). The result of this process is

$$-\frac{1}{f_1} = \frac{1}{f_2} = \frac{e}{8mV}\int_{z_1}^{z_2} B_0{}^2\,dz \tag{14.20}$$

Evaluating the constant

$$-\frac{1}{f_1} = \frac{1}{f_2} = \frac{2.20 \times 10^{10}}{V}\int_{z_1}^{z_2} B_0{}^2\,dz \qquad \text{meters}^{-1} \tag{14.21}$$

where z_1 is a point to the left of appreciable field variation and z_2 a corresponding point to the right and B_0 is the axial component of magnetic-flux density in webers per square meter (10^4 gausses). Two important conclusions are immediately available from the above equations for focal length. The first is that the focal length in the two directions is the same. The second is that the lens is always convergent since the quantity in the integrand is always positive.

The corresponding rotation of image is given directly from Eq. (14.13), making use of Eq. (14.16). It is

$$\theta = \frac{1}{2\sqrt{V}}\sqrt{\frac{e}{2m}}\int_{z_1}^{z_2} B_0\,dz \tag{14.22}$$

Evaluating the constant,

$$\theta = \frac{1.480 \times 10^5}{\sqrt{V}}\int_{z_1}^{z_2} B_0\,dz \qquad \text{radians} \tag{14.23}$$

The rotation has a clockwise direction in a magnetic field that has a component in the positive z direction.

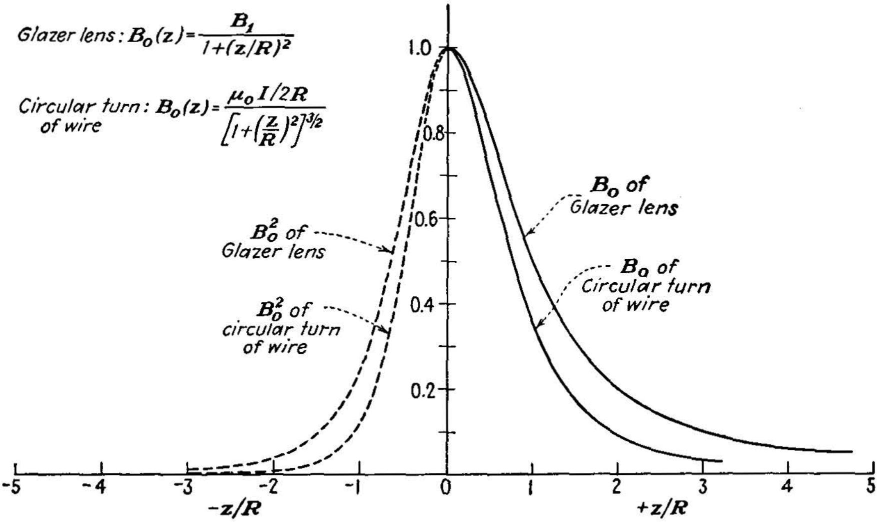


FIG. 14.5.—Axial magnetic-field distribution of a circular turn of wire and of the Glazer lens.

Magnetic Lens of a Circular Turn of Wire. A simple source of a magnetic field suitable for a magnetic lens is that of a circular turn of wire about the electron beam. Such a turn produces the necessary axial component of magnetic field having the desired rotational symmetry. The shape of the field is shown in Fig. 14.5. This is seen to approximate, roughly, a short uniform field parallel to the axis. The axial component of magnetic-flux density associated with a circular turn of wire is given by

$$B_0 = \frac{\mu_0 I R^2}{2(R^2 + z^2)^{3/2}} \tag{14.24}$$

where μ_0 is 1.257×10^{-6} henry per meter, the constant of proportionality between magnetic intensity and flux in rationalized mks units, and R is the radius of the turn of wire. Upon evaluating the coefficient, this becomes

$$B_0 = \frac{0.6285 \times 10^{-6} IR^2}{(R^2 + z^2)^{3/2}} \qquad \text{webers per meter}^2 \tag{14.25}$$

By using this expression for magnetic-flux density along the axis it is possible to evaluate the integral of Eq. (14.21), with the result

$$f = \frac{97.9VR}{I^2} \qquad \text{meters} \tag{14.26}$$

It is not possible to get a sufficiently strong lens with a single turn of wire, and therefore a coil of many turns is ordinarily used. When this is the case, the focal length becomes

$$f = \frac{97.9VR}{(NI)^2} \times \text{coil form factor} \qquad \text{meters} \tag{14.27}$$

where N is the number of turns and the coil form factor will generally assume a value between 1.00 and 1.25. The coil lens is weaker per ampere turn than the single-turn lens because the magnetic field is not so well concentrated. Where extremely strong lenses are desired, the field is further concentrated by means of iron pieces surrounding the coil.

The Glazer Lens. It is possible to calculate exactly the characteristics of a magnetic lens having an axial-flux-density function of the form $B_0 = \dfrac{B_1}{1 + \left(\dfrac{z}{R}\right)^2}$. This field form[1,2] is approximately that obtained from a large coil or from a coil with pole pieces. A brief study of this lens is valuable because it is possible to determine its optical characteristics exactly and compare them with those obtained from the approximate formulas given before. The assumed field form is plotted in Fig. 14.5. It is seen to be similar to that of a single turn of wire. The field of a circular turn of wire drops to 10 per cent of its peak value in 1.96 radii, whereas the field of the Glazer lens drops to 10 per cent in 3 radii.

---

[1] Glazer, W., Strenge Berechnung magnetischer Linsen der Feldform $H = \dfrac{H_0}{1+\left(\dfrac{z}{a}\right)^2}$, *Zeit. für Phys*, vol. 117, 285–315, 1941.

[2] Marton, L., and R. G. E. Hutter, Optical Constants of a Magnetic Type Electron Microscope, *Proc. I.R.E.*, vol. 32, pp. 546–552, September, 1944.

Upon applying the *approximate* expression for focal length of Eq. (14.20) to the assumed field form the focal length is found to be

$$f = \frac{16mV}{RB_1{}^2 e\pi} \qquad \text{meters} \tag{14.28}$$

The paraxial differential equation for the assumed field form is

$$\frac{d^2r}{dz^2} + \frac{eB_1{}^2}{8mV} \frac{r}{\left[1 + \left(\frac{z}{R}\right)^2\right]^2} = 0 \tag{14.29}$$

By means of judicious substitutions this equation can be converted to a form that is directly integrable.[1]

This yields a general-ray solution in the form

$$r(z) = R\sqrt{1 + \left(\frac{z}{R}\right)^2} \left\{C_1 \sin\left[\sqrt{1+p^2} \operatorname{arc\,cot}\left(\frac{z}{R}\right)\right] + C_2 \cos\left[\sqrt{1+p^2} \operatorname{arc\,cot}\left(\frac{z}{R}\right)\right]\right\} \tag{14.30}$$

where C_1 and C_2 are arbitrary constants and $p^2 = \frac{eB_1{}^2R^2}{8mV}$ is a lens-strength parameter. By proper choice of the constants C_1 and C_2 the general ray can be made to pass through any two points or meet any two conditions in general.

Since the general-ray equation is known, the focal points, focal lengths, and location of the principal planes of the Glazer lens can be found. The first principal ray is found by letting $\frac{dr}{dz}$ be zero and r finite at $z = +\infty$. The second principal ray is found by letting the slope be zero and the displacement finite at $z = -\infty$. The principal planes are located at the intersection of the initial and final straight-line portions of the principal rays. The focal points are found at the points at which the principal rays cross the axis.

The focal length of the Glazer lens is

$$f_1 = -f_2 = \frac{R}{\sin n\left(\frac{\pi}{\sqrt{1+p^2}}\right)} \tag{14.31}$$

where n assumes integral values. The significance of the focal length being multiple-valued is that for very strong fields a principal ray enter-

---

[1] Let $y = \frac{r}{R}$, $x = \frac{z}{R}$, and then make the further substitution $y = \frac{v(\phi)}{\sin\phi}$ and $x = \cot\phi$. This yields the differential equation $v''(\phi) = -(1 + k^2)v(\phi)$, which is readily solved.

ing the lens parallel to the axis will advance and oscillate transversely, crossing the axis several times. For normal applications the value of n is taken as 1. The above value of focal length has the same low-field value as given by the approximate formula of Eq. (14.28), that is, $\frac{2R}{\pi p^2}$. For larger values of field there may be a considerable divergence from the approximate value. The divergence does not, however, occur until the lens is strong enough for the electron to cross the axis within the region of appreciable field. A comparison of the focal lengths as determined by the exact and approximate formulas is given in Fig. 14.6.

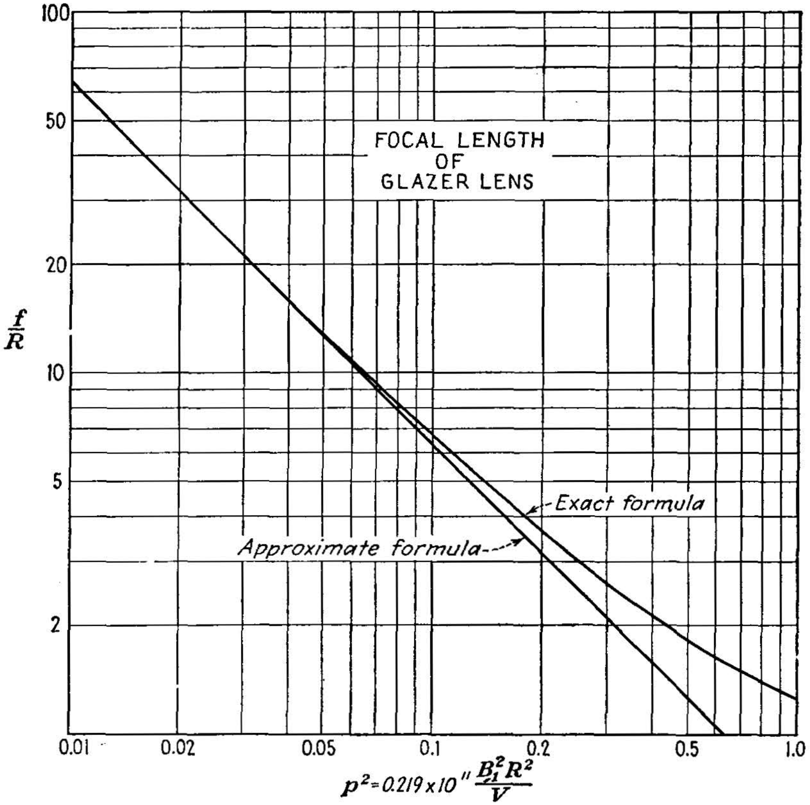


FIG. 14.6.—Focal lengths of the Glazer lens by exact and approximate formulas.

The location of the focal points is given by

$$F_1 = -F_2 = R \cot n \left(\frac{\pi}{\sqrt{1 + p^2}} \right) \tag{14.32}$$

This has the same weak-field asymptotic value as the focal length, which means that for weak fields the principal planes are located at the lens center. As the lens field is increased, the focal length becomes greater than the distance from the lens center to the focal points. This means that the first principal planes move away from the focal points. The first principal plane is to the right of the lens center and the second to the left. A plot of the focal length and focal-point position is shown in Fig. 14.7.

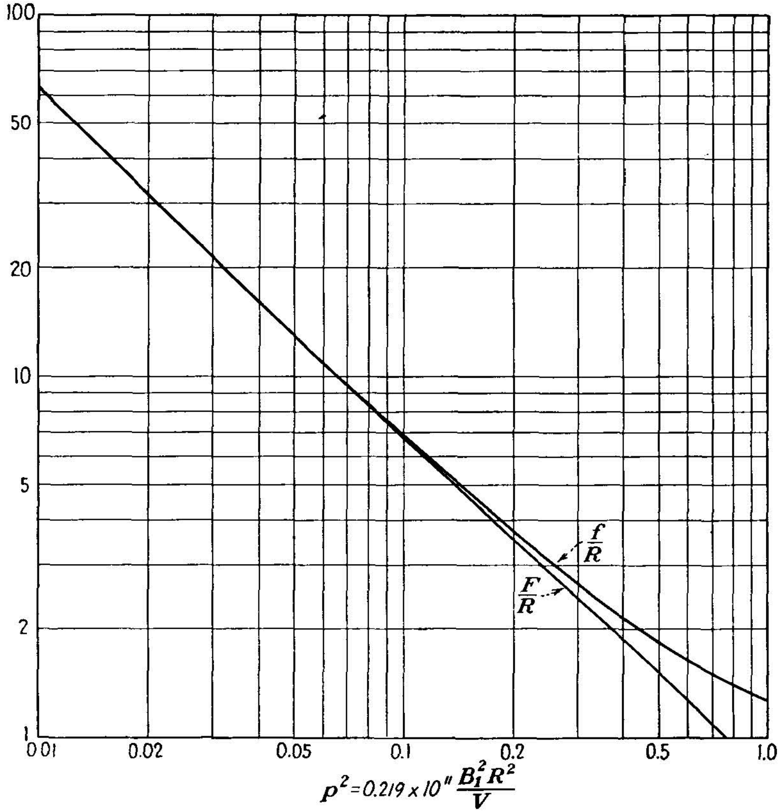


FIG. 14.7.—Focal length and principal-plane location of the Glazer lens.

14.6. Practical Magnetic Lenses. A coil of fine wire, square or rectangular in cross section, about the beam axis is a practical lens. Its strength is not very great, however, and its field is not very well confined. Both these features may be improved by partly shielding the coil with an iron shield but still maintaining a gap along which magnetic lines will pass parallel to the axis. In Fig. 14.8 are shown some practical lenses and their approximate fields. For extremely strong lenses such as are needed in electron microscopes the gap is made very small and is brought as close to the axis as possible by extended pole pieces. Such a lens is shown in Fig. 14.8*e*. It is not possible to calculate the performance of such a lens because of nonuniform saturation of the pole pieces. In part *d* of the figure there is shown a double lens composed of two sections containing coils passing currents in opposite directions. This makes the net image rotation through the lens zero for equal currents in the halves and tends to reduce the distortion associated with the image rotation.

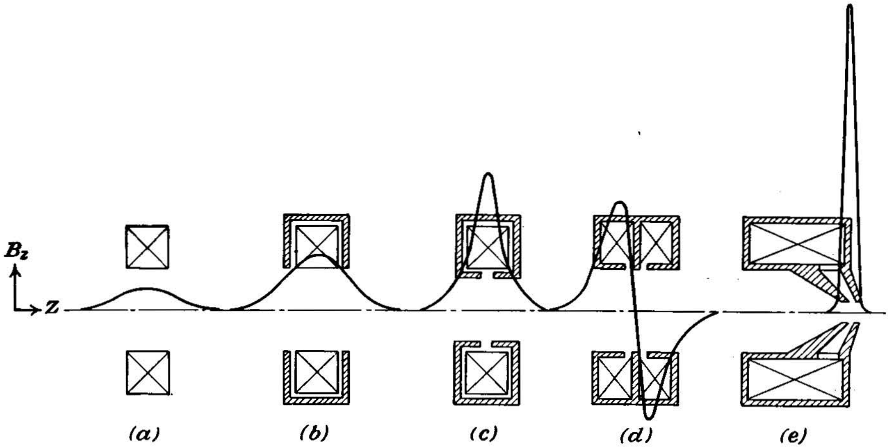


FIG. 14.8.—Practical magnetic focusing coils.

14.7. Magnetic-lens Defects. Magnetic lenses are subject to all the aberrations encountered in electrostatic lenses, plus a type of distortion associated with the image rotation. This type of distortion is known as "spiral distortion" and is illustrated in Fig. 14.9. It results from the fact that the rotation of different parts of the image is a function of the radial position. Its effect may be reduced by limiting the beam by

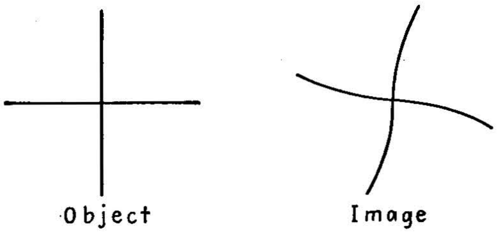


FIG. 14.9.—Spiral distortion in magnetic-lens images.

very small apertures, or it may be largely eliminated by using pairs of lenses giving rotation in opposite direction.

In addition to spiral distortion there may be distortion from current ripple in the magnetic coils or from stray fields. The effect of current ripple is to cause a point focus to become a blurred spot. Stray alternating fields will cause a point focus to become a short line.

14.8. The General Equations of Motion in Combined Electric and Magnetic Fields. In the previous work in this chapter there have been described the effects of a nonuniform magnetic field upon an electron presumed to be moving in a region of constant electrostatic potential. For completeness there will be outlined in this section the basic relations that apply to electrons moving in combined electric and magnetic fields of rotational symmetry. This involves considerable analysis the end point of which is the differential equation of motion of a paraxial electron in terms of the axial potential and the axial component of magnetic field. Although the yield for a great deal of work is quite small, the methods involved are fundamental and instructive enough to make the inclusion of this section worth while.

The force on an electron in a combined electric and magnetic field is given by

$$\boldsymbol{F} = m\boldsymbol{a} = e[\nabla V + \boldsymbol{B} \times \boldsymbol{v}] \tag{14.33}$$

where ∇V is the gradient of potential and the components of $\boldsymbol{B} \times \boldsymbol{v}$ have been given in Eqs. (14.10) to (14.12). Equation (14.33) is a compact representation of three coordinate equations and needs to be expanded for any specific application.

In the work with electrostatic fields it was found that the electric intensity and the corresponding forces on electrons were all derivable from the electric potential. Similarly, it is convenient to consider that the magnetic-flux vector $\boldsymbol{B}$ is derivable from a vector potential $\boldsymbol{A}$. The relations for the electrostatic case are similar but not exactly analogous to those for the magnetic case. Electrostatic potential fields are analogous to the irrotational flow of an incompressible fluid. Magnetic fields are analogous to the sourceless rotational flow of an incompressible flow.

The basic relations for electrostatic potentials are quickly listed. First the line integral of electric intensity around any closed path is *always* zero.

$$\int \boldsymbol{E} \cdot d\boldsymbol{l} = 0 \tag{14.34}$$

where the dot indicates the so-called scalar product, which is equal to the product of the magnitude of the vectors by the cosine of the angle between them. An equivalent statement of this is that the curl of the electric intensity, *i.e.*, the microscopic circulation, is always zero.

$$\nabla \times E = 0 \tag{14.35}$$

Whenever the curl of a vector is zero, then that vector is the gradient of some scalar function. Specifically,

$$E = -\nabla V \tag{14.36}$$

or intensity is the negative gradient of potential. The electrostatic potential results from a summation of the effect of various electric charges by the relation

$$V = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho\, dv}{r} \tag{14.37}$$

The corresponding relations for the magnetic field are also quickly given. The net outward flux through any closed surface is *always* zero.

$$\int B \cdot ds = 0 \tag{14.38}$$

An equivalent statement of this is that the divergence of magnetic flux is zero.

$$\nabla \cdot B = 0 \tag{14.39}$$

When the divergence of a vector is zero, then that vector is the curl of some other vector.

$$B = \nabla \times A \tag{14.40}$$

The vector *A* is called the "magnetic vector potential." Just as the electrostatic potential results from a summation of the effects of individual charges, so does the magnetic vector potential result from the summation of the effect of various currents.

$$A = \frac{\mu_0}{4\pi} \int \frac{J\, dv}{r} \tag{14.41}$$

in which *J* is vector current density. The vector *A* is seen to have the same direction as the currents that create it. The divergence of *A* is taken as zero in the static case.

When Eq. (14.40) is expanded and written in component form using cylindrical coordinates, it becomes, first in determinant form,

$$B = \frac{1}{r} \begin{vmatrix} i_r & ri_\theta & i_z \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial z} \\ A_r & rA_\theta & A_z \end{vmatrix} \tag{14.42}$$

where i_r, i_θ, and i_z are unit vectors in the r, θ, and z directions, respectively. When this determinant is expanded, the component equations become

$$B_r = \frac{1}{r}\frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z} \tag{14.43}$$

$$B_\theta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \tag{14.44}$$

$$B_z = \frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) - \frac{1}{r}\frac{\partial A_r}{\partial \theta} \tag{14.45}$$

These are the general equations relating B and A for cylindrical coordinates.

For the particular problems of electron optics most of the magnetic-lens fields are like those of a circular coil. In such there is only a θ component of current and hence only a θ component of A. Further, the θ component of either current or A does not vary with angle. Hence we may write

$$A_r = A_z = 0 \tag{14.46}$$

$$\boldsymbol{A} = A_\theta \boldsymbol{i}_\theta \tag{14.47}$$

$$\frac{\partial A_\theta}{\partial \theta} = 0 \tag{14.48}$$

With these restrictions the component relations between B and A become

$$B_r = -\frac{\partial A_\theta}{\partial z} \tag{14.49}$$

$$B_\theta = 0 \tag{14.50}$$

$$B_z = \frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) \tag{14.51}$$

From the above set of equations and the fact that $\nabla \times \boldsymbol{B} = 0$ it is possible to obtain a differential equation for A_θ alone. In subsequent work the θ subscript for A will sometimes be dropped for simplification, though it will be remembered that the vector $\boldsymbol{A}$ has a θ component only. Setting the curl of $\boldsymbol{B}$ equal to zero in terms of the θ component of $\boldsymbol{A}$,

$$\nabla \times \boldsymbol{B} = \frac{1}{r}\begin{vmatrix} \boldsymbol{i}_r & r\boldsymbol{i}_\theta & \boldsymbol{i}_z \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial z} \\ -\dfrac{\partial A}{\partial z} & 0 & \dfrac{1}{r}\dfrac{\partial (rA)}{\partial r} \end{vmatrix} \tag{14.52}$$

which expands into

$$r \text{ component of } \nabla \times \boldsymbol{B} = 0 - 0 \tag{14.53}$$

$$\theta \text{ component of } \nabla \times \boldsymbol{B} = -\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial (rA)}{\partial r}\right) - \frac{\partial^2 A}{\partial z^2} \tag{14.54}$$

$$z \text{ component of } \nabla \times \boldsymbol{B} = 0 - 0 \tag{14.55}$$

Upon setting the θ component of curl $\boldsymbol{B}$ above equal to zero there is obtained the differential equation for the vector potential A,

$$\frac{\partial^2 A}{\partial z^2} + \frac{\partial^2 A}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{A}{r}\right) = 0 \tag{14.56}$$

which, of course, applies only to magnetic fields having a rotational symmetry and produced by currents flowing exclusively in the θ direction. This equation is similar to but not identical with Laplace's equation but serves the same function in defining A as Laplace's equation does in defining V. This equation may be solved by series exactly as was done for potential and magnetic flux. In this case the series is restricted to odd powers of r because the vector potential like the current that generates it is an odd function of r. The resultant series expansion for A is

$$A_\theta(r,z) = \frac{rB_0}{2} - \frac{r^3 B_0''}{2^2 \cdot 4} + \frac{r^5}{2^2 \cdot 4^2 \cdot 6} B_0^{(4)} + \cdots + \frac{(-1)^{n+1} B_0^{(2n-2)}}{n[(n-1)!]^2}\left(\frac{r}{2}\right)^{2n-1} \tag{14.57}$$

where $B_0 = B(0,z)$ is the value of the axial component of magnetic flux.

By restricting Eq. (14.33) to fields of rotational symmetry and utilizing Eqs. (14.49) to (14.51) there result the component equations of motion

$$\frac{m}{e}(\ddot{r} - r\dot{\theta}^2) = -\dot{\theta}\frac{\partial}{\partial r}(rA_\theta) + \frac{\partial V}{\partial r} \tag{14.58}$$

$$\frac{m}{e}\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = \dot{z}\frac{\partial A_\theta}{\partial z} + \frac{\dot{r}}{r}\frac{\partial(rA_\theta)}{\partial r} \tag{14.59}$$

$$= \frac{1}{r}\frac{d}{dt}(rA_\theta) \tag{14.60}$$

$$\frac{m}{e}\ddot{z} = -r\dot{\theta}\frac{\partial A_\theta}{\partial z} + \frac{\partial V}{\partial z} \tag{14.61}$$

These are the basic equations from which now some simplified relations will be obtained. These equations are so far exact.

Equation (14.60) integrates to

$$mr^2\dot{\theta} = erA_\theta = \frac{er^2 B_0}{2} \tag{14.62}$$

for the paraxial case. The constant of integration is zero since $A_\theta = 0$ for $r = 0$ as may be seen by reference to its series expansion. Substituting the value of $\dot{\theta}$ from Eq. (14.62) into Eq. (14.58) yields

$$m\ddot{r} = e\frac{\partial}{\partial r}\left(V - \frac{1}{2}\frac{e}{m}A_\theta{}^2\right) \tag{14.63}$$

A similar operation upon Eq. (14.61) yields

$$m\ddot{z} = e\frac{\partial}{\partial z}\left(V - \frac{1}{2}\frac{e}{m}A_\theta{}^2\right) \tag{14.64}$$

The energy equation is obtained from these last two equations by multiplying the first by $\dot{r}$ and the second by $\dot{z}$, adding, and integrating. The result is

$$\frac{m}{2}[\dot{r}^2 + (r\dot{\theta})^2 + \dot{z}^2] = eV \tag{14.65}$$

Use has been made in obtaining this of Eq. (14.62). Note that the kinetic energy is independent of A and hence of the magnetic field. This is consistent with the idea previously propounded that a magnetic field can change only the direction of an electron and cannot change its energy because the force is always directed at right angles to the electron's velocity.

From Eq. (14.62) the approximate rotation of an electron is given by

$$\theta = \sqrt{\frac{e}{8m}}\int\frac{B_0}{\sqrt{V}}\,dz \tag{14.66}$$

Inserting the series expansions for A and V into Eqs. (14.63) and (14.64) gives

$$m\ddot{r} = e\left\{-\frac{r}{2}\left(V_0'' + \frac{1}{2}\frac{e}{m}B_0{}^2\right) + \frac{r^3}{2^2\cdot 4}\left[V_0{}^{(4)} + \frac{2e}{m}B_0B_0''\right] + \cdots\right\} \tag{14.67}$$

and

$$m\ddot{z} = e\left[V_0' - \frac{r^2}{2^2}\left(V_0''' + \frac{e}{m}B_0B_0'\right) + \cdots\right] \tag{14.68}$$

The paraxial components of these last two equations are found by retaining only first-order terms,

$$\ddot{r} = -\frac{re}{2m}\left(V_0'' + \frac{1}{2}\frac{e}{m}B_0{}^2\right) \tag{14.69}$$

$$\ddot{z} = \frac{e}{m}V_0' \tag{14.70}$$

Note that the radial component of acceleration due to the magnetic field is always convergent. The paraxial differential equation may now be obtained by eliminating time from the last two equations by the use of

$$\frac{dr}{dz} = \frac{\dot{r}}{\dot{z}} \tag{14.71}$$

$$\frac{d^2r}{dz^2} = \frac{\ddot{r}\dot{z} - \dot{r}\ddot{z}}{\dot{z}^3} \tag{14.72}$$

and

$$\dot{z}^2 = \frac{2eV_0}{m} \tag{14.73}$$

With these substitutions the paraxial differential equation is

$$\frac{d^2r}{dz^2} + \frac{V_0'}{2V_0}\frac{dr}{dz} + \frac{r}{4V_0}\left(V_0'' + \frac{eB_0{}^2}{2m}\right) = 0 \tag{14.74}$$

This can also be written in the form

$$\sqrt{V_0}\frac{d}{dz}\left(\sqrt{V_0}\frac{dr}{dz}\right) = -\frac{r}{4}\left(V_0'' + \frac{eB_0{}^2}{2m}\right) \tag{14.75}$$

These equations are seen to be of proper form because the paraxial differential equation of either a varying electric or magnetic field alone is derivable from them.

If in addition to the effect of the electric and magnetic fields there be considered the defocusing effect of the mutual radial repulsion of the electrons, then a factor of the form $-\frac{r\rho}{2\varepsilon_0}$ must be added within the parentheses of the last term of the above two equations, where ρ is the space-charge density within the beam and ε_0 is the dielectric constant of free space.

Note that the paraxial differential equations of Eqs. (14.74) and (14.75) are second-order linear differential equations. This means that even with combined electric magnetic fields a general ray can be expressed in terms of two independent principal rays.

CHAPTER 15

CATHODE-RAY TUBES

15.1. The General Form of Cathode-ray Tubes. The external physical form of cathode-ray tubes is well known. They generally have a glass envelope shaped like an Erlenmeyer flask. The electrical leads to the tube come out through a base at the mouth of the flask. The inside of the flask is coated with aquadag. The bottom of the flask is coated inside with a fluorescent material.

The internal parts of the cathode-ray tube include an electron gun, devices for horizontal and vertical deflection of the beam, and a fluorescent screen. The electron gun is a combination of electrodes for producing and focusing a beam of electrons. It consists of a cathode, a control electrode, and two or more electrodes forming an objective lens. One commonly used arrangement of these parts is shown in Fig. 15.1.

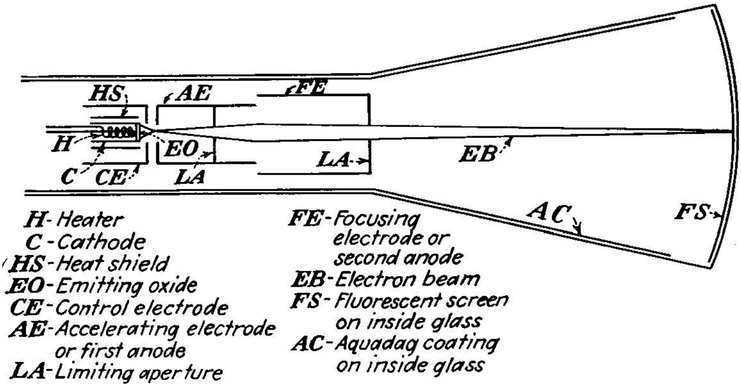


FIG. 15.1.—Schematic cathode-ray-tube structure.

The general description of the parts of the electron gun and their function is as follows: The cathode consists of a small capped cylinder of sheet nickel. The cap is coated with emitting material. The cathode is indirectly heated by an insulated filament wire inside the cylinder. The cathode is generally surrounded by a close-fitting but nontouching cylinder, which acts as a heat shield and increases the thermal efficiency of the cathode. The heat shield is supported at the nonemitting end of the cathode and projects slightly beyond the cap at the other end. This projection at the emitting end serves to keep the emitted electrons from spreading. Figure 15.2 shows this cathode construction. The control electrode takes the form of a cylindrical can completely surrounding the cathode and having a circular aperture in front of the emitting cap. It performs the same function as the control grid in a triode. The main focusing lens is in the form of one of the lenses described in the chapter on Electrostatic Electron Optics. The first part of this lens always has limiting apertures to keep the electrons from spreading too much by reducing the angle of the beam. With the arrangement shown in Fig. 15.1 there is a crossover point of the electrons between the control electrode and the first electrode of the objective lens. The spot seen on the screen of the tube is an image of the crossover portion of the beam, which is the cross section of minimum diameter. In the two-cylinder objective lens shown in Fig. 15.1 the small cylinder is called the "first anode" or "accelerating electrode." The large cylinder is called the "second anode" or "focusing electrode."

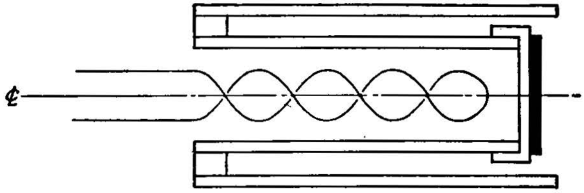


FIG. 15.2.—Typical cathode structure.

The relative potentials on the electrodes of the electron gun are quite important. For a typical tube with the electrode arrangement of Fig. 15.1 the electrode potentials are as follows:

Electrode	Potential relative to cathode, volts	Potential relative to ground, volts
Filament	0	−800
Cathode	0	−800
Heat shield	0	−800
Control electrode	−10 to +10	−790 to −810
First anode	+200	−600
Second anode	+800	0
Aquadag coating	+800	0

The physical construction of the electron gun requires a high degree of precision in the alignment of the electrodes. The electrodes are usually supported from glass or ceramic insulating rods, in turn supported on a stem similar to that used in vacuum tubes. Mica spacers or metal springs serve to center the gun within the neck of the envelope. Extreme care must be used in aligning the electrodes axially. Such alignment is usually achieved by means of a mandrel, which is removed after the electrodes have been spotted or crimped into place.

The beam-deflecting devices most commonly used are electrostatic deflecting plates or magnetic deflecting coils The deflecting plates are always placed inside the tube and are usually supported from the end of the electron gun. Some special-purpose tubes have the deflecting electrodes supported directly from the neck of the envelope, with leads brought out directly through the glass. Magnetic deflection, when used, is achieved by coils external to the tube. The coils are arranged so that they produce a component of magnetic field perpendicular to the axis of the tube.

The fluorescent screen at the end of the tube serves to reveal the position of the electron beam and to translate electrical impulses into a visual picture. The screen consists of a thin layer of fluorescent material on the inside of the tube, which lights up when struck by electrons. The fluorescent coating is generally a fairly good insulator so that it is necessary for the electrical circuit consisting of the power supply and the beam to be completed by means other than electrical conduction. The means in this case is secondary emission. As beam electrons strike the fluorescent screen, they liberate secondary electrons, which look for a more positive electrode to be drawn toward. This electrode is found in the aquadag coating, which is at beam, or ground, potential. The fluorescent screen will assume a negative potential because of an accumulation of beam electrons that are slow to leak off. This means that there exists a potential difference between the fluorescent screen and the aquadag coating that is in the right direction to attract the secondary electrons liberated by the beam impact.

15.2. Electron-gun Design. The fields and electron paths in the vicinity of the cathode of an electron gun are extremely complex. This makes the exact design of electron guns necessarily at least partly empirical. Although it is not possible to give equations resulting in exact design relations, it is possible to indicate the nature and magnitude of the effects encountered.

For low-current guns such as are used in ordinary cathode-ray tubes the electrostatic field in the vicinity of the cathode has the general shape shown in Fig. 15.3. The fields will be similar to those encountered in the vicinity of simple apertures, but modified by departures in the shape of the electrodes from that ideal configuration. Between the control electrode and the first anode the field will be approximately linear. In the vicinity of the cathode the field will be strongly curved in such a way as to cause all electrons emitted from the cathode to be drawn strongly toward the axis. Only a limited portion of the center of the cathode emitting area will present a positive gradient of potential to the emitted electrons. At cutoff the gradient of potential will be zero at the cathode center and negative in other parts. As the control grid is made more positive, a region of positive gradient will grow from the center until at sufficiently positive control-electrode potentials the entire surface of the cathode may emit.

Cutoff Relations in the Electron Gun. The control electrode has an action somewhat similar to that of the control grid in a vacuum tube except that in addition to controlling the gradient at the center of the cathode it controls the size of the emitting area. For this reason it is difficult to write a current-voltage relation, but it is possible to estimate the cutoff relation. Exact relations for the configuration of Fig. 15.3 are almost impossible to write, but the field configuration is approximated by the idealized electrode configuration of Fig. 15.4. For this configuration the aperture-field formula of Eq. (13.37) will apply very closely. Here the axial potential is given by

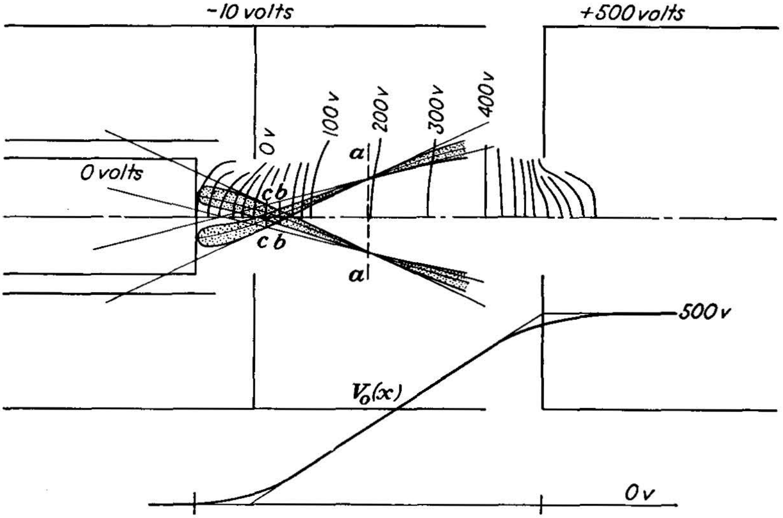


FIG. 15.3.—Field in the vicinity of the cathode of an electron gun.

$$V_0(z) = \frac{-V_2 d_{23} + (V_3 - V_2)d_{12}}{2d_{12}d_{23}}\left[|z| - \frac{2R}{\pi}\left(\frac{z}{R}\arctan\frac{R}{z} - 1\right)\right] + \frac{(V_3 - V_2)d_{12} + V_2 d_{23}}{2d_{12}d_{23}}z + V_2 \quad (15.1)$$

when $V_1 = 0$.

The gradient of potential at the cathode is found by taking the derivative of this expression with respect to z and then setting $z = -d_{12}$. The resulting expression is

$$\frac{dV_0(-d_{12})}{dz} = \frac{V_2d_{23} - (V_3 - V_2)d_{12}}{2d_{12}d_{23}}\left[1 - \frac{2R}{\pi}\left(\frac{1}{R}\arctan\frac{R}{d_{12}} - \frac{d_{12}}{R^2 + d_{12}^2}\right)\right] + \frac{(V_3 - V_2)d_{12} + V_2d_{23}}{2d_{12}d_{23}} \tag{15.2}$$

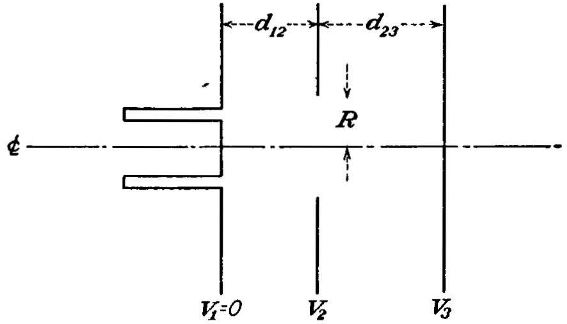


FIG. 15.4.—Idealized cathode-electrode configuration.

The cathode current will be cut off when the gradient at the cathode is zero. The equivalent amplification factor of the control structure is found by setting $\dfrac{dV_0}{dz}$ at $z = -d_{12}$ equal to zero and then taking the negative ratio of V_3 to V_2.

$$\mu = -\frac{V_3}{V_2} \tag{15.3}$$

This is an amplification factor that determines the current cutoff. The amplification factor has the specific value

$$\mu = \frac{\pi d_{23}}{d_{12}}\left[\frac{1}{\arctan\dfrac{R}{d_{12}} - \dfrac{1}{\dfrac{R}{d_{12}} + \dfrac{d_{12}}{R}}} - \frac{1}{\pi}\left(\frac{d_{12}}{d_{23}} + 1\right)\right] \tag{15.4}$$

A nomographic chart of equivalent amplification factor as a function of control-electrode-aperture radius and grid–first-anode distance, each expressed in units of cathode-grid distance, is given in Fig. 15.5. These values, while not exactly the same as those for the electrode structure of Fig. 15.3, will serve to indicate the order of magnitude and the nature of the variation of the equivalent amplification factor with the critical dimensions. Measured values of μ for the structure of Fig. 15.3 will be considerably higher than those obtained from Eq. (15.4) because of the shielding effect of the control-electrode-cylinder extension. Measured data on some specific electrode structures are available in the literature.[1]

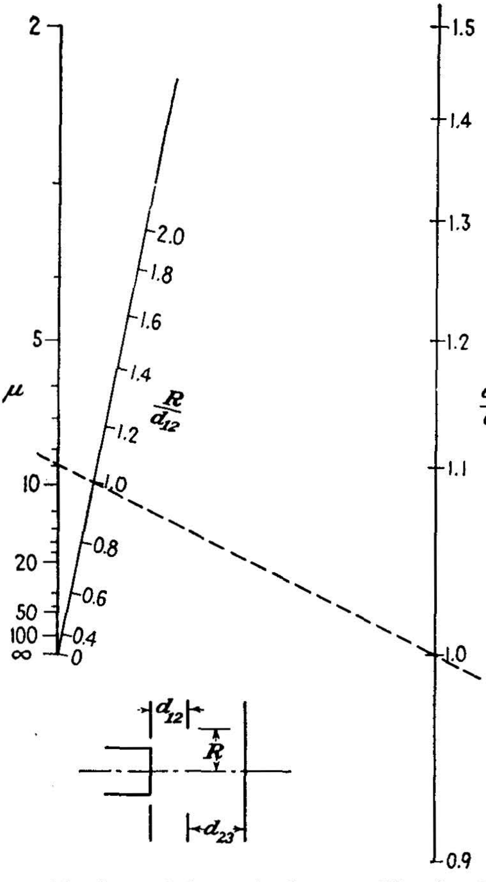


FIG. 15.5.—Nomographic chart of the equivalent amplification factor of an electron gun as given by Eq. (15.4).

Electron Paths in the Electron Gun. As may be seen from the equipotential plot of Fig. 15.3, the field in front of the gun cathode is strongly convergent. It is not easy to apply the methods described in the chapter on Electrostatic Electron Optics to this portion of the gun because the focusing field is so strong relative to the low-velocity electrons that a focus is obtained within the region of field variation. The type

---

[1] MALOFF, I. G., and D. W. EPSTEIN, "Electron Optics in Television," pp. 167–169, McGraw-Hill, New York, 1938.

of lens encountered here is sometimes referred to as an "immersion lens" because the object is immersed in the lens.

Some typical electron paths in the vicinity of the cathode of an electron gun are shown in Fig. 15.3. Rays leaving the cathode are propelled forward and attracted toward the axis. As a result of this action, the rays cross the axis at a point not very far out in the field. After crossing the axis the rays are curved the other way and are again bent toward the axis, but the action in this portion of the field is so weak that the rays invariably retain their divergent characteristic. Shown in the figure are three rays. These rays differ by virtue of the direction of the velocity of emission of the electrons. The three rays show the effect of emission velocity directed toward the axis, normal to the cathode, and away from the axis. All three rays are seen to come to a focus on the plane *a-a*. This represents an image of the cathode. The minimum diameter of the cross section of the beam occurs at the plane *b-b*. This plane of minimum cross section is called the "crossover" of the beam. It is seen to be much smaller in diameter than either the cathode or its image. The best spot is obtained by focusing this crossover rather than the cathode or its image on the fluorescent screen. Actually, the crossover cannot serve as object, but rather its virtual image at *c-c*, as found by projecting back straight lines from the region of uniform field, serves as object. This virtual image of the crossover is slightly larger than the crossover itself but is still smaller than the cathode or its image.

Since the beam crossover is used as the object whose image forms the working spot of the beam, its location and size are of considerable importance. These values are rather hard to determine exactly, but some good approximations can be given. The location of the crossover can be estimated by making use of the fact that the field in the vicinity of the cathode is approximately spherical. Hence, if the radius of curvature of the zero-potential contour can be found, it is to be expected that the crossover will occur at this radial distance from the cathode. It was shown in the chapter on Electrostatic Electron Optics that the radius of curvature of any equipotential surface in a field of rotational symmetry is given by

$$R_0 = \frac{2V_0'}{V_0''} \tag{15.5}$$

where V_0' and V_0'' are the first and second derivatives of the axial potential, respectively. The curvature radius can be obtained in a straightforward manner from Eq. (15.1), and at the cathode has the value

$$\frac{R_0}{R} = \pi\left[\frac{1}{\left(\frac{V_3}{V_2} - 1\right)\frac{d_{12}}{d_{23}} - 1} + \frac{1}{\pi}\left(\arctan\frac{R}{d_{12}} - \frac{1}{\frac{R}{d_{12}} + \frac{d_{12}}{R}}\right)\right]\left[1 + \left(\frac{d_{12}}{R}\right)^2\right]^2 \tag{15.6}$$

for $V_1 = 0$. As an example, if $R = d_{12} = d_{23}$ and $V_2 = 0$, then the value of μ from Eq. (15.4) is 9.02 and the value of R_0 from Eq. (15.6) is 1.14R. The very low value of μ results from the fact that the control-electrode-aperture diameter is twice the cathode–control-electrode distance in this example. From the value of the zero-potential radius of curvature the crossover is expected to occur nearly in the plane of the control electrode. If the cathode is flat, Eq. (15.6) will predict a smaller radius of curvature than actually exists because of the influence of the flat cathode upon the field. Cathodes may, however, readily be curved to fit the normal aperture fields.

The size of the crossover diameter may also be estimated by assuming that the field in the vicinity of the cathode is spherical. The finite size of the crossover results from electrons being emitted at all angles from each point on the cathode, and with appreciable velocity. The larger the emission velocity, the larger the crossover diameter. The electron behavior encountered is like that shown for the idealized spherical electrodes of Fig. 15.6. For this situation the radius of the crossover is given by

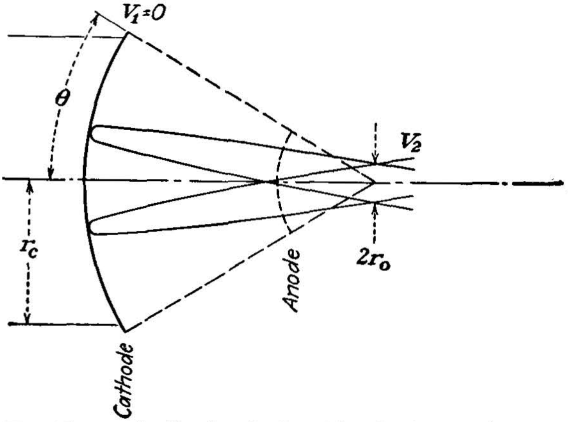


FIG. 15.6.—Idealized cathode with spherical field. This gives the notation for use in Eq. (15.7).

$$r_0 = \frac{+2r_c}{\sqrt{\dfrac{V_2}{V_e}}\sin 2\theta} \tag{15.7}$$

where r_c is the radius of the cathode, r_0 is the radius of the crossover, V_2 is the potential of the crossover, V_e is the voltage equivalent of the velocity of emission, and θ is the half angle of the cathode as viewed from the crossover.[1]

---

[1] RUSKA, E., Zur Fokusierbarkeit von Kathodenstrahlbündeln grosser Ausgangsquerschnitte, *Zeit. für Phys.*, vol. 83 (Nos. 9, 10), pp. 684–698, 1933.

This equation results from an analysis of the electron path in a spherical field. It is properly valid only for small values of θ, say less than 20 deg. Actually, electrons will be coming off the cathode with all possible velocities so that an average value of V_e must be used. If a value of V_e, that is, the voltage equivalent of the velocity not exceeded by 80 per cent of the electrons is used, then the crossover radius of Eq. (15.7) will contain at least 80 per cent of the beam current and probably more, for not all electrons emitted with greater velocities will have tangential components greater than that corresponding to V_e. The virtual image of the crossover will generally be larger than the actual crossover. An enlargement by a factor of two is, however, not often exceeded. The position of the virtual image of the crossover will be on the cathode side of the actual crossover and may even lie behind the actual cathode. It may generally be expected to lie within a distance equal to the cathode–control-electrode distance of the actual cathode, which is close enough for design of the subsequent lens system.[1]

Current-voltage relations for the electron gun are not readily specified analytically. If a low-mu gun structure is used and the control electrode is operated at zero potential, the cathode will come very close to being temperature-limited. Some specific measured data on gun current-voltage relations are available in the literature.[2]

The concept of the screen spot as an image of the beam crossover in front of the cathode is largely one of convenience. There is evidence that the screen spot is actually an image of the cathode. The size of the crossover may be obtained from optical considerations of the field in front of the cathode. At low beam voltages, however, the thermal velocities of emission of the electrons from the cathode are large enough compared with the potential of the crossover so that they are an appreciable factor in determining the spot size. At large beam voltages the thermal velocities may be expected to be low compared with the potential of the beam crossover so that they do not add appreciably to the size of the cathode image. An examination of the operation of tubes with beam potentials greater than 1,000 volts, from the viewpoint of straightforward cathode imaging, yields some useful information on the properties of beams.[3]

A rough optical approximation to the field action in front of the

---

[1] Zworykin, V. K., and G. A. Morton, "Television," pp. 368–383, Wiley, New York, 1940.

[2] Maloff and Epstein, *op. cit.*, pp. 171–176.

[3] Liebmann, G., Image Formation in Cathode Ray Tubes and the Relation of Fluorescent Spot Size and Final Anode Voltage, *Proc. I.R.E.*, vol. 33, pp. 381–389, June, 1945.

cathode may be had by considering the equivalent lens to be made up of two regions of constant index of refraction with a spherical refracting surface between them, as shown in Fig. 15.7. For this equivalent lens Lagrange's law will hold,

$$\left(\frac{n_2}{n_1}\right) M_l M_a = 1 \tag{15.8}$$

where n_1 and n_2 are equivalent indices of refraction, $M_l = \frac{y_2}{y_1}$ is the lateral magnification, and $M_a = \frac{\alpha_2}{\alpha_1}$ is the angular magnification. The lens equation in this case is

$$\frac{n_2}{z_i} + \frac{n_1}{z_0} = \frac{n_2 - n_1}{R_s} \tag{15.9}$$

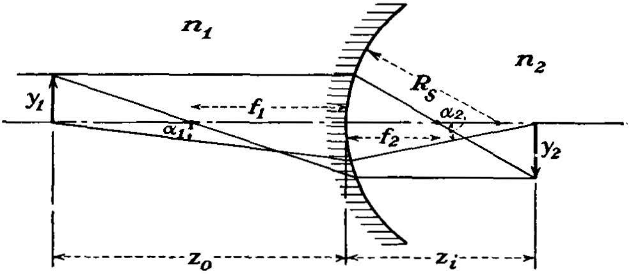


FIG. 15.7.—Spherical-surface refraction equivalent of cathode-lens action.

where z_0 and z_i are object and image distance, respectively, and R_s is the radius of curvature of the spherical refracting surface.[1] Solving Eq. (15.9) for z_i,

$$z_i = \frac{n_2 R_s z_0}{z_0(n_2 - n_1) - n_1 R_s} \tag{15.10}$$

From Eq. (15.8) the linear magnification is

$$M_l = \frac{y_2}{y_1} = \frac{n_1 z_i}{n_2 z_0} \tag{15.11}$$

Substituting the value for z_i from Eq. (15.10) into Eq. (15.11),

$$M_l = \frac{1}{\left(\frac{n_2}{n_1} - 1\right)\frac{z_0}{R_s} - 1} \tag{15.12}$$

---

[1] A derivation of this expression is available in almost any book on geometrical optics.

If it is now considered that the object is located very close to the refracting surface so that z_0 is small compared with R_s, then approximately

$$M_l = -1 \tag{15.13}$$

which says that the image is the same size as the object and is roughly independent of the indices of refraction. Further,

$$\frac{n_2 M_a}{n_1} = -1 \tag{15.14}$$

from Eq. (15.8), which says that the angle which a ray makes with the axis at the image is inversely proportional to the index of refraction n_2 if n_1 is held constant. If the assumptions made here are approximated in a cathode-ray tube, then it is to be expected that the size of the cathode image is independent of the voltage of the first accelerating electrode. Then, since, as was shown in Figs. 13.37 to 13.45 of the chapter on Electrostatic Electron Optics, the magnification of the usual electrostatic objective lens is approximately eight-tenths of the ratio of image to object distance independent of voltage ratio, it is to be expected that the spot size is also independent of the beam voltage. It also follows from Eq. (15.14) that the product of the beam voltage and beam area in the fore part of the objective lens is a constant,

$$\pi r^2 V_0 = K \tag{15.15}$$

where r is the radius of the beam in an arbitrary plane. Measurements on actual tubes show that both these expectations are realized very closely for beam voltages above 1,000 volts.[1] The above performance applies only if the limiting apertures intercept a negligible amount current.

Focusing System. The production of a beam crossover of small diameter and high current density is the principal problem in electron-gun design. The rest of the design problem is relatively simple. The beam crossover need only be followed by one of the types of objective lenses described in the chapter on Electrostatic Electron Optics. The cylinder lenses are found to be most suitable, and there is not much to choose between them. In fact, almost any kind of lens will do, for it is always possible to find a voltage ratio that will focus the beam crossover on the screen. When cylinder lenses are used, it is necessary to put limiting apertures within the first cylinder to limit the initial divergent action of the beam. This is illustrated in the schematic drawing of Fig. 15.1. A limiting aperture is often put at the end of the second cylinder as well.

---

[1] LIEBMANN, *op. cit.*

The data of Figs. 13.37 to 13.45 can be used directly to design the focusing system. The object distance is simply taken as the distance from the beam crossover to the reference point in the lens. The image distance is the distance from the reference point in the lens to the fluorescent screen. Focusing, with the electrode arrangement of Fig. 15.1, is obtained by adjusting the first-anode voltage, all other voltages being kept fixed. Intensity of the beam is controlled by adjusting the control-electrode voltage. With this arrangement the two adjustments indicated will have a principal effect upon focus and beam intensity, respectively, but it will be noticed that the adjustment of the beam intensity affects the focus somewhat, and vice versa. While adjustment of the control electrode has the principal effect of changing the beam current, it also changes the location of the beam crossover and so affects the focus. Adjustment of the first-anode voltage has the principal effect of adjusting the focus, but the field of the first anode reaches back to the cathode and

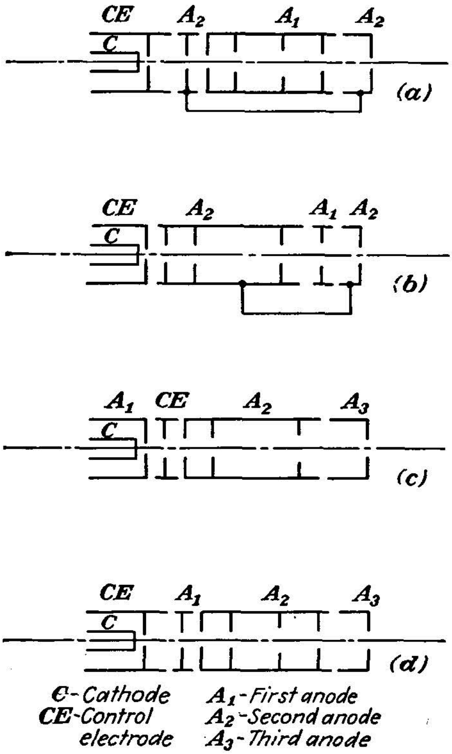


FIG. 15.8.—Typical electron-gun structures using electrostatic focusing.

changes the intensity somewhat. This interaction of controls can be improved by making use of a different electrode arrangement.

Alternative Electrode Structures. In Fig. 15.8 are shown some alternative electron-gun structures that are extensively used. In the arrangement of electrodes shown for gun *a* the cathode–control-electrode structure is about the same as that just discussed. The focusing action in this case, however, is divided into two parts so that there are really two objective lenses. Thus, the accelerating anode is split in two, with the focusing anode located between the two parts. The principal advantage of this electrode arrangement over that shown in Fig. 15.1 is that the interaction between the intensity and focusing controls is greatly reduced. With this arrangement the electrodes adjacent to the control electrode are kept at a fixed potential. This means that any changes in the focusing field of the objective lenses are shielded from the control-grid region by the first part of the accelerating electrode. A better capture of secondary electrons liberated at the limiting apertures may also be effected. The action of the other guns is evident from their structure.

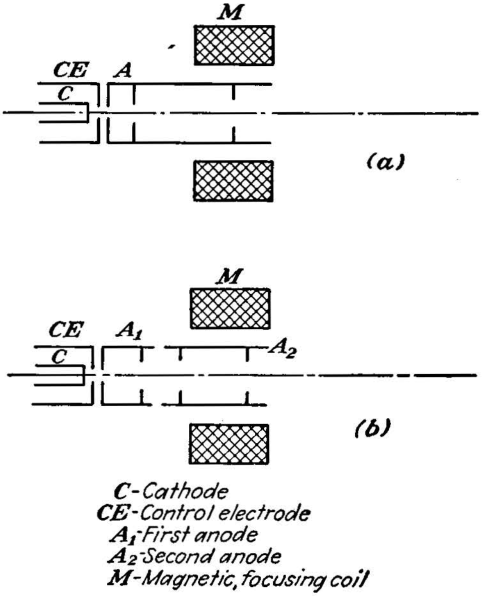


FIG. 15.9.—Electron guns with magnetic focusing.

Where magnetic focusing is used, the simple arrangements of Fig. 15.9 are adequate. The arrangement of gun *a* consists of only a cathode, control electrode, accelerating electrode, and magnetic focusing coil. The magnetic coil can usually be put outside the tube. This arrangement, while simpler of construction, requires that part of the power supplied to the tube must be regulated to give constant current for the magnetic focusing coils. The arrangement of gun b actually involves a combination electrostatic and magnetic focusing action.

15.3. Deflection Devices. Electrostatic Deflecting Plates.—Electrostatic deflection plates have already been discussed in the chapter on Laws of Electron Motion. The deflection obtained from electrostatic deflecting plates is given by Eq. (6.23), which states that the deflection is equal to half the beam length multiplied by the ratio of the deflecting voltage to the beam voltage and by the ratio of the axial deflecting-plate length to the deflecting-plate spacing.

$$y_s = \frac{lbV_d}{2aV_0} \tag{6.23}$$

where y_s is the spot deflection at the fluorescent screen in any units of length, l is the beam length from plates to screen in the same units, b is the deflecting-plate length in the same units, a is the deflecting-plate spacing in the same units, V_d is the deflecting potential, and V_0 is the beam potential. Of principal significance is the fact that the spot deflection is proportional to the deflecting voltage and inversely proportional to the beam voltage.

Magnetic Deflection. Magnetic deflection of a beam may be achieved by applying a magnetic field perpendicular to the beam for a short distance of its length. The electrons moving through this magnetic field will move in a short section of an arc of a circle if the field is constant, emerging at an angle with their original direction. The radius of curvature of an electron moving at right angles to a constant field was given by Eq. (6.62) as

$$R = 3.37 \times 10^{-6} \frac{\sqrt{V}}{B} \qquad \text{meters} \tag{6.62}$$

where V is in volts equivalent to the velocity and B is webers per square meter (10^4 gausses). Consider the deflecting arrangement of Fig. 15.10. The magnetic field is shown by the dots in the rectangle astride the beam. If the field is constant within this rectangle, the beam will move in the arc of a circle of radius given by Eq. (6.62). Upon emerging from the magnetic field the electrons will move in straight lines at an angle θ with the original path given by

$$\tan \theta \cong \frac{b}{R} = \frac{y_s}{l} \tag{15.16}$$

Hence the deflection is given by

$$y_s = \frac{lb}{R} = \frac{lbB}{3.37 \times 10^{-6}\sqrt{V}} \qquad \text{meters} \tag{15.17}$$

for B in webers per square meter. To obtain deflection in centimeters, express B in gausses, and drop the factor 10^{-6} in the denominator. Magnetic deflecting coils are invariably placed outside of the tube neck and take the form of a saddle-shaped coil.

Relative Merits of Electrostatic and Magnetic Deflection. Both electrostatic and magnetic deflection are capable of giving linear deflection over the entire tube face. The differences in their operation lie only in their sensitivity and frequency characteristics. There is an advantage in using magnetic deflection at high beam voltages, for a relatively smaller increase in deflecting field is necessary. This results from the fact that electrostatic deflection is inversely proportional to beam voltage, whereas magnetic deflection is inversely proportional to the square root of the beam voltage. Hence, if beam voltage were raised from 1,000 to 4,000 volts, four times the voltage would be necessary to give the same electrostatic deflection, whereas only twice the magnetic coil current would be necessary to give the same magnetic deflection. For this reason, magnetic deflection is commonly used in high-voltage television viewing tubes. A disadvantage of magnetic deflection is that a negative-ion spot forms in the middle of the screen, due to negative ions emitted from the cathode,

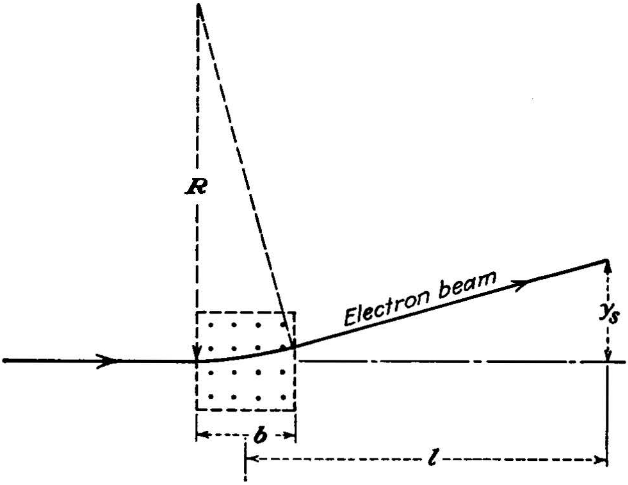


FIG. 15.10.—Magnetic deflection of an electron beam by a region of uniform magnetic field, as given by Eq. (15.17).

which, because of their great mass, are scarcely deflected by the magnetic fields.[1] With electrostatic deflection a negative-ion spot does not form because the negative ions are deflected the same as are the electrons. Electrostatic deflection has the advantage as far as frequency characteristics go. With ordinary construction, electrostatic deflection can resolve frequencies as high as several hundred kilocycles. The practical upper limit of magnetic deflection is of the order of 10 kc. Magnetic deflecting coils are most suitably fed from a high-impedance source. Since the coil represents a fairly high inductance, the voltage appearing across it for the same current increases linearly with frequency. This means that excessive voltages are reached at relatively low frequencies.

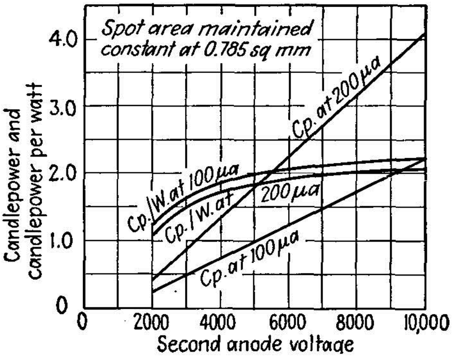


FIG. 15.11.—Brightness and luminescent efficiency of willemite as a function of beam voltage. (*Maloff and Epstein.*)

An advantage of magnetic deflection which electrostatic deflection does not possess is that it is more suitable for radial deflection and polar representation. Magnetic deflecting coils can be made to rotate about the tube and so give polar representations where the frequency of rotation required is not too high.

Visual versus Deflection Sensitivity. The light output from a spot on a fluorescent screen under beam excitation is found to be approximately linear with beam voltage in accordance with *Lenard's equation,*

$$CP = AI(V - V_0) \tag{15.18}$$

where CP is the candle-power output, A is a constant of the material of the order of 2 candle power per watt, I is the beam current, V is

---

[1] BACHMAN, C. H., and C. W. CARNAHAN, Negative-ion Components of the Cathode Ray Beam, *Proc. I.R.E.*, vol. 26, pp. 529–539, May, 1938.

the beam voltage, and V_0 is the voltage at which fluorescence starts, somewhere between 500 and 1,000 volts. Curves of candle power and candle power per watt as a function of beam voltage for willemite are shown in Fig. 15.11. From this and Eq. (6.23) it is seen that if the attempt is made to increase the brightness of the trace by using a higher beam voltage a corresponding decrease in deflection is suffered. With the gun arrangement of Fig. 15.1 this sets a practical limit to the deflection sensitivity of the order of 0.1 mm per d-c volt with a beam voltage of 3,000 volts. The deflection sensitivity can be increased by reducing the beam voltage, but this correspondingly reduces the spot brightness.

Postdeflection Acceleration. The dilemma of having to sacrifice deflection sensitivity to achieve visual sensitivity, or vice versa, can be circumvented by making use of the principle of postdeflection acceleration.[1,2] A schematic drawing of a tube making use of this principle is shown in Fig. 15.12. The principle that is used increases the deflection sensitivity by deflecting the electron beam at relatively low voltage and then subsequently accelerating it before the electrons hit the screen. With this arrangement the beam is deflected at relatively low velocity, giving a good deflection sensitivity, and then is subsequently accelerated, giving a good visual sensitivity. Part of the increase in deflection

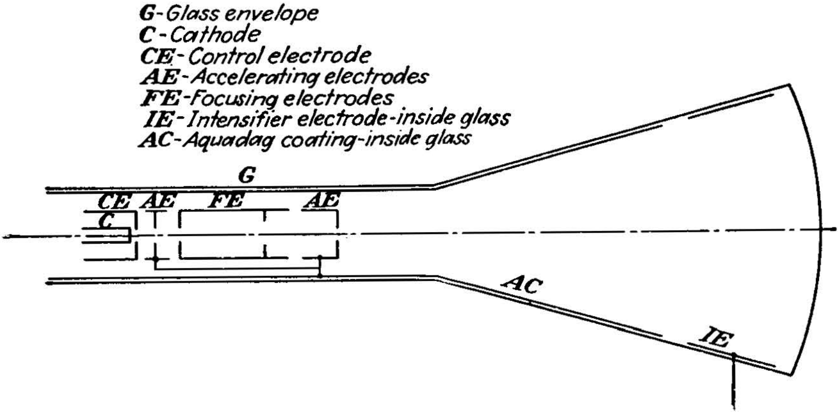


FIG. 15.12.—Structure of the postdeflection-acceleration tube.

---

[1] DE GEIER, J., A Cathode Ray Tube with Post Acceleration, *Philips Tech. Rev.* vol. 5, pp. 245–252, September, 1940.

[2] PIERCE, J. R., After Acceleration and Deflection, *Proc. I.R.E.*, vol. 29, pp. 28–31, January, 1941.

sensitivity gained by this arrangement is lost because the final accelerating field is somewhat convergent, thus reducing the deflection.

The postdeflection-acceleration arrangement makes use of an ordinary electron gun supplemented by a so-called "intensifier electrode," which takes the form of a ring of conducting material inside the tube near the fluorescent screen and operated at about twice the voltage of the last previous electrode.

Some typical voltages as used in this arrangement are as follows:

Electrode	Potential relative to cathode, volts	Potential relative to ground, volts
Cathode	0	−1,500
Control electrode	−10 to +10	−1,490 to −1,510
Accelerating electrode	1,500	0
Focusing electrode	375	−1,125
Intensifier electrode	3,000	+1,500

With this arrangement of electrodes the deflecting plates, which are situated between the second part of the accelerating electrode and the aquadag coating, are operated at zero direct voltage, as are also the adjacent electrodes. With the above operating conditions a deflection sensitivity of 0.3 mm per volt may be realized. In general, an improvement of 3 to 5 times in deflection sensitivity may be obtained by this arrangement of electrodes.

The amount of the beam intensification may be extended considerably even beyond that indicated above. By putting in a number of intensifier electrodes with potentials progressively greater, spot brightness may be increased by a factor of 10, and yet the deflection sensitivity may be increased slightly over that which would obtain if the final intensifier potential were applied to the last gun electrode and the intensifier electrode were removed. Tubes with final intensifier-electrode potentials as high as 15,000 volts are considered commercially feasible.[1]

15.4. Fluorescent Materials. The characteristics of the fluorescent material used for a cathode-ray screen are critical factors in the successful operation of the tube. The various characteristics such as spot brightness, spectral characteristics, trace persistence, secondary emission, and voltage characteristics are all controllable by the composition and pro-

---

[1] CHRISTALDI, P. S., Cathode Ray Tubes and Their Applications, *Proc. I.R.E.*, vol. 33, pp. 373–381, June, 1945.

cessing of the material. Hundreds of fluorescent materials have been studied, and by now the data on such materials are very numerous.[1–5]

Definitions. Strictly speaking, the term "fluorescence" as applied to cathode-ray-tube screen operation is a misnomer, but it is so widely used that it will also be applied here. Properly, one should distinguish between the three terms "luminescence," "fluorescence," and "phosphorescence." These may be distinguished, briefly, as follows:

Luminescence. This refers to visible and near-visible radiation in excess of black-body radiation due to some form of excitation. The term applies to the radiation both during and after excitation. It can be classified according to the means of excitation into many classes, such as cathode luminescence (the luminescence produced by the impact of electrons), photoluminescence (the luminescence caused by exposure to radiation), electroluminescence (the luminescence given off by ionized gases), and bioluminescence (the luminescence of living organisms). About 10 kinds of luminescence can be enumerated.

Fluorescence. Fluorescence is luminescence during excitation. In the case of cathode luminescence this refers to the light emitted during the period of electron bombardment.

Phosphorescence. Phosphorescence is the luminescence occurring after excitation. In a cathode-ray tube this is the radiation given off after the beam excitation has ceased.

Phosphor. Materials that manifest cathode luminescence are known by the general name of phosphors.[6]

Since phosphorescence as well as fluorescence is involved in cathode-ray-tube operation, it would be more suitable to refer to screen action as "cathode luminescence" than as "fluorescence."

General Make-up of Phosphors. A great number of materials will exhibit luminescences when bombarded with electrons. Practically all nonmetallic inorganic crystals will exhibit this effect, as will also glasses and some organic materials. Most of these will, however, react so

---

[1] LEVERENZ, H. W., and F. SEITZ, Luminescent Materials, *Jour. Appl. Phys.*, vol. 10, pp. 479–493, July, 1939.

[2] ZWORYKIN and MORTON, *op. cit.*, Chap. II.

[3] LEVERENZ, H. W., Cathode Luminescence as Applied in Television, *RCA Rev.*, vol. 5, pp 131–175, October, 1940.

[4] STAUFFER, L. H., Characteristics of Fluorescent Materials, *Electronics*, vol. 14, pp. 32–34, October, 1941.

[5] KUSHEL, I., Phosphors and Their Behavior in Television, *Electronic Ind.*, vol. 4, pp. 100–105, 132, 134, December, 1945

[6] PERKINS, T. B., Cathode Ray Terminology, *Proc. I.R.E.*, vol. 23, pp. 1334–1343, November, 1935.

weakly as to be useless. To be suitable for practical purposes a material must produce a high brightness, be stable under electron bombardment, have a suitable color, and have a persistence that is not too great.

The basic ingredients of a practical luminescent material are a base material, a flux, and an activator. The base material is generally a crystalline, colorless semiconductor. Good base materials are the oxides and sulphides of zinc, cadmium, magnesium, and silicon. Oxides of copper, iron, and nickel are not good bases. The flux is some material such as sodium chloride that is used to catalyze crystallization of the base and is subsequently removed. The activator is one of a group of metals including silver, copper, manganese, and chromium. The presence of 10 to 100 parts per million of such metals may increase the light output of the base material by a factor of 10 to 100. Various other metals such as lead, iron, nickel, and cobalt will inhibit radiation to such an extent that the presence of one part per million of these metals will ruin the luminescence. The activator serves to furnish a material with additional energy levels for the excited electrons to jump between. The theory of luminescence is qualitatively understood, but so many anomalies exist that there is no direct procedure that can be applied to synthesizing a suitable phosphor.[1]

Phosphors are prepared by mixing the base material and flux, heating to crystallize, drying, and regrinding for application. Screens may be deposited from settling out of a liquid suspension or by spraying the material suspended in a volatile organic liquid such as acetone to which has been added a small amount of binder. In the settling process a mild electrolyte such as ammonium carbonate is used to prevent the particles from settling nonuniformly.

Luminous Properties of Fluorescent Materials. One of the best and most widely used fluorescent materials is zinc orthosilicate, $ZnO + SiO_2$:Mn, with a manganese activator. In its natural form this is known as "willemite." The natural material is subject to great variations in performance due to impurities, and therefore only synthetic materials are now used. Synthetic willemite gives the bright-green trace so well known to users of test oscilloscopes.

The light output of synthetic willemite follows quite closely Lenard's law as previously given. Curves of light output in candle power and luminous efficiency in candle power per watt as a function of beam voltage were previously given in Fig. 15.11. Actually, Lenard's law does not hold exactly for the fluorescent material but does so only apparently in Fig. 15.11 because the output is plotted against beam

---

[1] LEVERENZ, H. W., Phosphors Versus the Periodic System of Elements, *Proc. I.R.E.*, vol. 32, pp. 256–263, May, 1944.

potential, which is higher than the screen potential. When the output is plotted against screen potential, then it is found that the light output is given by

$$CP = AI(V_s - V_0)^n \tag{15.19}$$

where A is a constant, I is the beam current, V_s is the fluorescent-screen potential and V_0 is the screen potential at which luminescence starts, and n is an exponent that is nearly 2 for synthetic willemite and in general has a value between 2 and 2.8.[1,2]

The spectral characteristics of willemite are compared with the sensitivity of the human eye in Fig. 15.13. This figure shows that most of the radiant energy from this material is concentrated in the green region of the spectrum. Phosphors are available giving almost any desired color response. A few of the most useful phosphors are listed in Table VII. A white luminescence may be obtained by mixing a yellow-color-producing phosphor such as zinc cadmium sulphide with a green-blue-color-producing phosphor such as zinc sulphide. Further specific characteristics of commercial phosphors are given in Appendix IV.

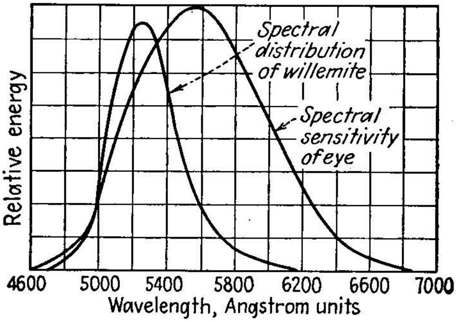


FIG. 15.13.—Spectral characteristics of willemite.

The persistence characteristics of the luminescence are quite important. In test oscilloscopes and television kinescopes a relatively short persistence time is desired. In some transient studies and most radar applications a long persistence is desired. Most of the phosphors have short-persistence characteristics, while a few of the yellow-green sulphides have long-persistence characteristics. Synthetic willemite will build up to 50 per cent of its maximum radiation in about 2.5 milliseconds.

---

[1] NELSON, H., Method of Measuring Luminescent Screen Potential, *Jour. Appl. Phys.*, vol. 9, pp. 592–599, September, 1938.

[2] FONDA, G. R., Phosphorescence of Zinc Silicate Phosphors, *Jour. Appl. Phys.*, vol. 10, pp. 408–420, June, 1939.

TABLE VII
CHARACTERISTICS OF THE PRINCIPAL PHOSPHORS[1]

No.	Phosphor	Chemical composition	Color	Spectral maximum, angstrom units	Approx. candle power per watt
1	Zinc oxide	ZnO	Violet	Ultraviolet	<0.1
2	Zinc sulphide	ZnS:Ag	Blue-violet	4,700–4,500	5*
3	Calcium tungstate	$CaWO_4$	Blue	4,300	<1
4	Zinc silicate	$ZnO + SiO_2$	Blue	4,200	<1
5	Zinc sulphide	ZnS	Light blue	4,700	1–5*
6	Zinc aluminate	$(ZnO + Al_2O_3)$:Mn	Green-blue	5,130	~1
7	Zinc silicate (willemite)	$(ZnO + SiO_2)$:Mn	Blue-green	5,230	3*
8	Zinc sulphide	ZnS:Cu	Green	4,700–5,250	>4*
9	Zinc germanate	$(ZnO + GeO_2)$:Mn	Yellow-green	5,370	1.5
10	Beta zinc silicate	$(ZnO + SiO_2)$:Mn	Green-yellow	5,600–5,700	3*
11	Zinc beryllium silicate	$(ZnO + BeO + SiO_2)$:Mn	Green to orange	5,230–6,500	1–2*
12	Zinc cadmium sulphide	$(ZnS + CdS)$:Ag	Blue to red	4,700–>7,000	5*
13	Calcium silicate	$(CaO + SiO_2)$:Mn	Green to orange	5,500–6,500	<1
14	Cadmium silicate	$(CdO + SiO_2)$:Mn	Orange yellow	5,850	~1
15	Magnesium silicate	$(MgO + SiO_2)$:Mn	Orange-red	6,400–6,700	<1
16	Zinc aluminate	$(ZnO + Al_2O_3)$:Cr	Red	>7,000	<1
17	Zinc beryllium zirconium silicate	$[ZnO + BeO + (Ti - Zr - Th - O_2) + SiO_2]$:Mn	White	4,200 + 5,500–6,000	~1*
18	Magnesium tungstate	$MgO + WO_3$	Very light blue	4,800	<1
19	Zinc borate	$(ZnO + B_2O_3)$:Mn	Yellow-orange	5,400–6,000	~1*
20	Cadmium borate	$(CdO + B_2O_3)$:Mn	Green-orange	5,300–6,300	<1
21	Cadmium tungstate	$CdO + WO_3$	Light blue	4,900	<1*

---

[1] Reprinted by permission from "Television" by V. K. Zworykin and G. A. Morton, published by John Wiley and Sons, Inc.

* Used in television.

The radiation will decay to 50 per cent of its maximum in 3 to 5 milliseconds. The decay is approximately logarithmic except for the first half millisecond, during which time it is more rapid than logarithmic.[1]

Electrical Characteristics of Phosphors. The potential that a fluorescent screen will assume will depend upon the beam potential, the secondary-emission characteristics of the screen, and the current-voltage transmission characteristics of the screen to the more positive adjacent electrodes. The secondary-emission characteristics of the screen have the form of the general characteristics described in the chapter on Electronic Emission. A typical secondary-emission characteristic showing the ratio of secondary to primary current is shown in Fig. 15.14. The screen can function properly only over the range of voltages for which the ratio of secondary to primary currents is greater than unity. Below the voltage at which the ratio is first unity the screen will block and repel beam electrons. The screen will "stick" at the potential at which the ratio again drops to unity, and it will not be possible to raise the screen above this potential. The screen-voltage–beam-voltage characteristics can be estimated by combining the effect of the secondary-current characteristic with the current-voltage transmission characteristic of the screen in conjunction with its adjacent electrode. The current

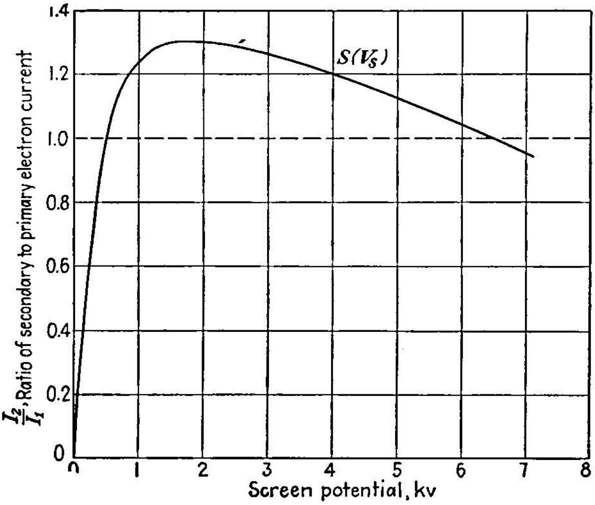


FIG. 15.14.—Ratio of secondary- to primary-electron current as a function of screen voltage of a fluorescent screen.

---

[1] NELSON, R. B., R. P. JOHNSON, and W. B. NOTTINGHAM, Luminescence during Intermittent Electron Bombardment, *Jour. Appl. Phys.*, vol. 10, pp. 335–342, May, 1939.

taken from the screen secondaries by the adjacent electrodes will depend upon the relative potential of the screen and the adjacent collector electrode. This function will have the general form shown in Fig. 15.15. When the collector is more than 20 volts more positive than the screen, it will collect virtually all the secondary electrons liberated by it. When the collector and the screen are at the same potential, the collector will still collect about half the secondary electrons. When the collector is more than 20 volts more negative than the screen, it will take virtually none of the secondary electrons. The relation between Figs. 4.18 and 15.15 will be apparent. The difference is due to the difference in physical form of the electrodes corresponding to each curve. Let the collector current be indicated by

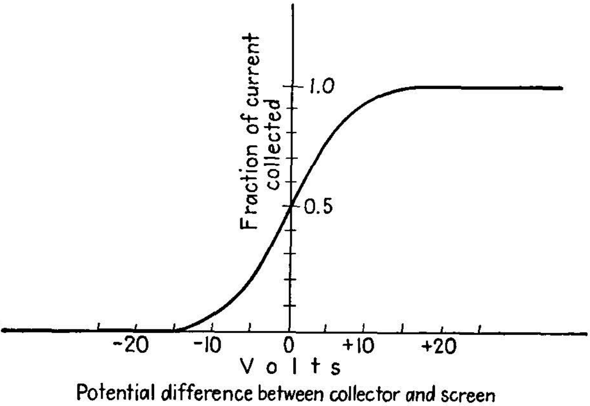


FIG. 15.15.—Collector current of a cathode-ray tube as a function of the difference of collector and fluorescent-screen potential.

$$I_c = I_2 T(V_c - V_s) \tag{15.20}$$

where I_c is the collector current, I_2 is the secondary current liberated by primary-electron impact, V_c is the collector potential, V_s is the screen potential, and $T(V_c - V_s)$ is the current-transmission function shown in Fig. 15.15. Let the secondary-ratio function of Fig. 15.15 be given by

$$\frac{I_2}{I_1} = S(V_s) \tag{15.21}$$

where I_1 is the beam current striking the screen and $S(V_s)$ is the secondary-ratio function shown in Fig. 15.14. For current equilibrium the collected current must equal the beam current. Equating the collected and beam current as given in the above two equations,

$$T(V_c - V_s) = \frac{1}{S(V_s)} \tag{15.22}$$

This neglects conduction-current components, which are, however, ordinarily quite small. It is possible to find the screen-voltage–beam-voltage (the latter being the same as the collector voltage) function

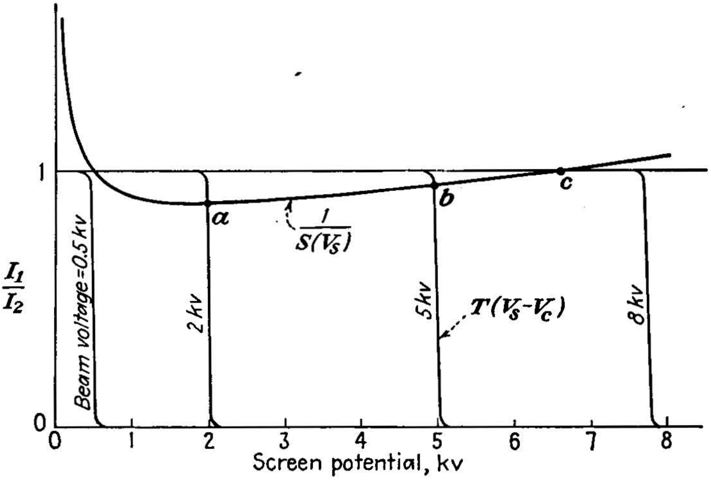


FIG. 15.16.—Graphical construction of the screen-potential–beam-potential characteristic.

graphically from Eq. (15.22). The method of construction is shown in Fig. 15.16. In this are plotted the reciprocal of the current-ratio function of Fig. 15.14 and the collector-current–voltage-difference function of Fig. 15.15 on a scale of screen voltage.

From Eq. (15.22) the screen potential is given by the intersection of the two curves for any reference beam potential. The entire curve desired is constructed point by point by shifting the collector-current–voltage-difference curve to correspond to different beam voltages and taking the corresponding intersections.[1,2] Some shifted transmission curves are shown. The resulting screen-voltage–beam-voltage characteristic is shown in Fig. 15.17. Points *a*, *b*, and *c* are taken from the similarly designated intersections in Fig. 15.16. It is seen that the screen

---

[1] NOTTINGHAM, W. B., Electrical and Luminescent Properties of Willemite under Electrical Bombardment, *Jour. Appl. Phys.*, vol. 8, pp. 762–778, November, 1937.

[2] NOTTINGHAM, W. B., Electrical and Luminescent Properties of Phosphor under Electron Bombardment, *Jour. Appl. Phys.*, vol. 10, pp. 72–83, January, 1939.

potential never exceeds the beam potential. The curve confirms the conclusion that the screen will not accept electrons below the potential at which the secondary current ratio is unity, nor can the screen be raised to a higher potential than that at which the ratio again drops to unity. Sticking potentials for screens ordinarily lie between 5,000 and 8,000 volts, though they can be raised to as high as 15,000 volts.[1] The critical blocking voltage will ordinarily lie in the vicinity of 200 volts.

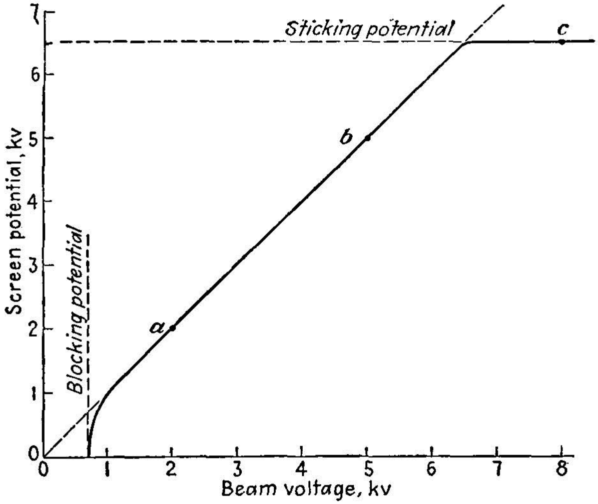


FIG. 15.17.—Screen-potential–beam-potential characteristic of a fluorescent screen.

15.5. Limitations of Spot Size. *Effect of Thermal Velocity of Emission.* It has already been mentioned that the size of the beam crossover in front of the cathode which is subsequently imaged into the spot is determined by the thermal velocities of emission of the electrons. The approximate size of the crossover for any limiting velocity of emission is given by Eq. (15.7). Actually, electrons are coming off the cathode with all velocities, as given by a Maxwellian distribution, so that there is no sharp edge to the beam; rather, it is found to have a cross section approximating the Gauss error curve

$$J(r) = A\epsilon^{-Br^2} \tag{15.23}$$

where $J(r)$ is the current density at any radius r and A and B are con-

---

[1] Beam potentials may be raised to as high as 30 kv by the use of metallized screens. See EPSTEIN, D. W., and L. PENSAK, Improved Cathode Ray Tubes with Metal-Backed Luminescent Screens, *RCA Rev.*, vol. 7, pp. 5–10, March, 1946.

stants related to the total current and rate of decay of current with radius, respectively.[1]

Associated with this effect, it is found from a consideration of the optics of thermally emitted electrons that the maximum current density with perfect focusing at a crossover of cathode image is given by

$$\frac{J}{J_0} = \frac{1}{M^2}\left[1 - (1 - M^2 \sin^2 \phi) \exp \frac{Ve}{kT}\left(\frac{M^2 \sin^2 \phi}{1 - M^2 \sin^2 \phi}\right)\right] \tag{15.24}$$

where J is current density at the crossover or cathode image, J_0 is cathode current density, M is the ratio of crossover or cathode-image diameter to cathode diameter, ϕ is the half angle of the cone including all electron paths reaching the point in question, T is cathode temperature in degrees Kelvin, k is Boltzmann's constant, and V is the potential at the point in question.[2–4]

Limiting values of Eq. (15.24) are of interest. For M large,

$$\frac{J}{J_0} = \frac{1}{M^2} \tag{15.25}$$

For M small,

$$\frac{J_m}{J_0} = \left(1 + \frac{eV}{kT}\right) \sin^2 \phi \tag{15.26}$$

where the symbol J_m is substituted for J because this is the largest possible value of current density that can be achieved under any conditions.

A curve of $\frac{J}{J_m}$ for various values of M and ϕ is shown in Fig. 15.18 for the case of $\frac{eV}{kT} = 10{,}000$ (this corresponds to a voltage of about 800 volts since $\frac{e}{k}$ has a value of 11,600 and T is about 1000°K for an oxide emitter.) ϕ in this case is understood to be the value determined by a stop or limiting aperture at, before, or after the crossover. As ϕ is decreased, more and more electrons with high thermal-emission velocities are thrown away so that a greater fraction of the cathode current is

---

[1] Law, R. R., High Current Electron Gun for Projection Kinescopes, *Proc. I.R.E.*, vol. 25, pp. 954–976, August, 1937.

[2] Langmuir, D. B., Theoretical Limitations of Cathode Ray Tubes, *Proc. I.R.E.*, vol. 25, pp. 954–976, August, 1937.

[3] Pierce, J. R., Limiting Current Densities in Electron Beams, *Jour. Appl. Phys.*, vol. 10, pp. 715–724, October, 1939.

[4] Pierce, J. R., A Figure of Merit for Electron Concentrating Systems, *Proc. I.R.E.*, vol. 33, pp. 476–478, July, 1945.

wasted and yet the more nearly the maximum possible current density is realized.

Since the extent to which the limiting current density can be approached depends upon the fraction of current used, it is convenient to draw a curve relating these two quantities. Let the ratio of the actual to the maximum current density be called the *intensity efficiency.*

$$\text{Intensity efficiency} = \frac{J}{J_m} \qquad (15.27)$$

The value of this expression is readily obtained from the quotient of Eqs. (15.24) and (15.26). Let the fraction of cathode current used be called the *current efficiency.*

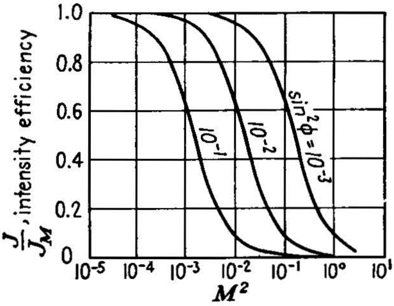


FIG. 15.18.—Intensity efficiency of an electron gun as a function of cathode magnification for various aperture sizes.

$$\text{Current efficiency} = \frac{JM^2}{J_0} \qquad (15.28)$$

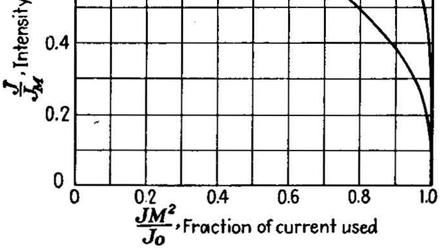


FIG. 15.19.—Intensity efficiency of an electron gun as a function of current efficiency.

For a given value of $\frac{V}{T}$ both of the above expressions are functions of $M \sin \phi$ alone. It further turns out that the relation between the two efficiencies varies numerically only a few per cent for voltages above 10 volts. A curve showing the relation between the intensity efficiency and the fraction of the current used is given in Fig. 15.19. Also shown in the figure is the curve for the line-focus case. These curves show that in order to approach the limiting maximum value of current density it is necessary to waste most of the current with limiting apertures. The above equations do not include the effects of electron collisions or lens aberrations but are limitations imposed by thermal velocities alone. A figure of merit for electron guns is the ratio of the area of the aperture that, in an equivalent ideal system, would pass as much cathode current as does the actual gun to the area of the actual aperture.[1] An equivalent figure of merit for television kinescopes in terms of the deflection angle and the number of scanning lines can be worked out by application of the above formulas.[2]

Space-charge Limitation of Spot Size. Another serious limiting factor in the production of small beam spots is the space-charge mutual repulsion between electrons in the beam, which prevents the electrons from coming together into a point focus. In a convergent beam, as the beam tends to come to a smaller diameter, the electrons get closer together, the space-charge density increases within the beam, and hence the mutual-repulsion forces become greater. This means that the radial components of velocity which the electrons have become less and less as the beam becomes more and more constricted, until finally they become zero at some finite beam diameter and then the beam begins to spread again.[3–5]

This action may be pictured by considering the behavior of the electrons in a cross section of the beam as seen by an observer moving along with the electrons. To such an observer, there is no axial motion, and only radial effects can be observed. The action is actually independent of the axial velocity. To make the problem soluble the following conditions will be assumed:

1. Electrons are uniformly distributed throughout the cross section of the beam.
2. Every electron has a radial component of velocity that is proportional to its radial distance from the axis.

These conditions are close enough to the actual conditions to make the answers based upon these assumptions useful. The first condition will hold if only a small fraction of the cathode current or if a high-current-density cathode, to be described later, is used. The second condition is the assumption made in treating paraxial electrons and is the condition for uniform convergence of the beam when small angles are involved. The general picture encountered in a convergent beam

---

[1] *Ibid.*

[2] LAW, R. R., Factors Governing Performance of Electron Guns in Television Cathode-ray Tubes, *Proc. I.R.E.*, vol. 30, pp. 103–105, February, 1942.

[3] WATSON, E. E., The Dispersion of the Electron Beam, *Phil. Mag.*, Ser. 7, vol. 3, pp. 849–853, April, 1927.

[4] BORRIES, B. V., and J. DOSSE, Zerstreuung von Elektronenstrahlen durch eigene Raumladung, *Arch. Elektrotech*, vol. 32, pp. 221–232, 1938.

[5] THOMPSON, B. J., and L. B. HEADRICK, Space Charge Limitations on the Focus of Electron Beams, *Proc. I.R.E.*. vol. 28, pp. 318–324, July, 1940.

is shown in Fig. 15.20 (the radial scale in this figure is greatly exaggerated, and the axial scale is foreshortened). The beam is seen to decrease in diameter to a minimum cross section and then expand again. At any cross section as in *a* the radial velocity at any point in the cross section is given by

$$v_r = kr \tag{15.29}$$

The outward force on any electron in the cross section is given by

$$F = eE = \frac{e\lambda}{2\pi\varepsilon_0 r} = \frac{e\rho r}{2\varepsilon_0}^{*} \tag{15.30}$$

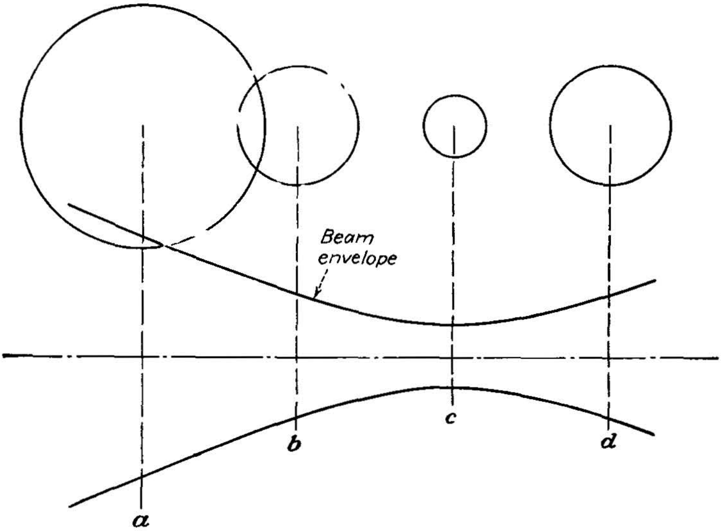


FIG. 15.20.—Effect of space-charge repulsion on a convergent beam.

where λ is the charge per unit length of the beam. This follows from the expression for electric intensity about a linear distribution of charge and from the fact that

$$\lambda = \pi r^2 \rho \tag{15.31}$$

where ρ is the volumetric space-charge density within the beam. It is seen that the outward force on any electron is proportional to the radial distance also. As a result of this relation, the percentage change in radial velocity of any electron will be constant throughout the beam, and hence

---

* In addition to the outwardly directed electrostatic force there is also an inwardly directed magnetic force. This magnetic force is only $\frac{v^2}{c^2}$ times as big as the electrostatic force, where v is the electron velocity and c is the velocity of light, and thus it is negligible for beam voltages under 10,000 volts. See BORRIES and DOSSE, *op. cit.*

the beam will constrict uniformly, maintaining the conditions that the electrons are distributed uniformly throughout the beam and that the radial velocity of any electron is proportional to its radial distance from the axis. At b the cross section of the beam is less than at a. At c the cross section of the beam assumes its minimum diameter. At this point the radial velocity of the electrons is zero. Beyond this point the beam expands to the larger diameter shown at d. In the region to the left of c the radial velocities of the electrons are directed inward, while to the right the radial velocities are directed outward. The behavior of the beam will be the same whether the electrons are moving to the left or to the right in Fig. 15.20. The *shape* of the beam envelope will not change with beam voltage or beam current, though the radial and axial scales will change. For purposes of analysis it is convenient to start at the cross section of minimum diameter and to study the beam's subsequent spread. This study yields a universal beam-spread curve, which can then be applied to any problem.

From Eq. (15.30) the radial acceleration of any outer electron of an initially parallel beam is given by

$$m\frac{d^2r}{dt^2} = F = \frac{\pi {r_0}^2\rho_0}{2\pi\varepsilon_0 r}e \tag{15.32}$$

in which the numerator is the charge per unit length in terms of the initial values of radius and space-charge density. If this equation is simplified by the substitutions

$$R = \frac{r}{r_0} \tag{15.33}$$

$$R' = \frac{dR}{dz}, \qquad R'' = \frac{d^2R}{dz^2} \tag{15.34}$$

$$v = \frac{dz}{dt} \tag{15.35}$$

then there results

$$\frac{2\varepsilon_0 mv^2}{e}RR'' = \rho_0 \tag{15.36}$$

This may be integrated by putting it into the form

$$\frac{\varepsilon_0 mv^2}{e\rho_0}2R'R'' = \frac{R'}{R} \tag{15.37}$$

with the result

$$K(R')^2 = \ln R \tag{15.38}$$

where $K = \dfrac{\varepsilon_0 mv^2}{e\rho_0}$ and the constant of integration is zero since $R' = 0$, $R = 1$, for $z = 0$. Extracting the square root,

$$R' = K^{-1/2} \sqrt{\ln R} \tag{15.39}$$

which for purposes of integration is best put into the form

$$dz = \frac{K^{1/2}\, dR}{\sqrt{\ln R}} \tag{15.40}$$

This has the solution

$$z = K^{1/2} \int_1^R \frac{dR}{\sqrt{\ln R}} \tag{15.41}$$

This equation gives the envelope of the beam as an integral function of R for any position z. Not much can be done to simplify this expression, for it does not integrate into simple standard functions. The shape of the envelope of the beam is best presented in curve form. First observe that the constant can be converted into a simpler form, making use of the fact that

$$V = \frac{mv^2}{2e} \tag{15.42}$$

and

$$I = \pi r_0{}^2 \rho_0 v \tag{15.43}$$

so that

$$\frac{z}{r_0} = 32.3 \frac{V_{kv}{}^{3/4}}{I_{ma}{}^{1/2}} \int_1^R \frac{dR}{\sqrt{\ln R}} \tag{15.44}$$

where V_{kv} is the potential in kilovolts and I_{ma} is the current in milliamperes. Although the integral cannot be expressed in terms of simple functions, one further change of variable is useful. If the substitution $R = \epsilon^{t^2}$ is made, then

$$\int_1^R \frac{dR}{\sqrt{\ln R}} = 2 \int_0^{\sqrt{\ln R}} \epsilon^{t^2}\, dt \tag{15.45}$$

Values of the right-hand integral above are tabulated on page 106 of Jahnke and Emde's Tables of Functions.[1] The plot of the values of Eq. (15.44) yields the *universal* beam-envelope curve of Fig. 15.21.[2]

The universal beam-spread curve of Fig. 15.21 gives the shape of a beam of initially parallel electrons. The curve applies for electrons moving either to the right or to the left and is symmetrical about the value $z = 0$. To apply the curve to any problem it is necessary only to enter

---

[1] Teubner, Leipzig, 1933.

[2] The spread of a beam subjected to an axial gradient of potential can be analyzed by a similar method. See Moss, H., A Space Charge Problem, *Wireless Eng.*, vol. 22, pp. 316–321, July, 1945.

the curve properly and then take numerical values from the curve. The curve will apply to low-current cathode-ray beams as well as to high-current power-tube beams. The curve shows that the spread of a beam is increased as the current is increased and the voltage is decreased. Considering the action of a convergent beam, the minimum spot diameter is decreased as the current is decreased and the voltage is increased.

An alternative representation of Fig. 15.21 is given in the nomographic chart of Fig. 15.22. This nomographic chart gives the spread of a beam with initially parallel electrons directly from the beam current, beam voltage, and beam length, without calculation. The diagonal line from lower left to upper right is a construction line. To use this graph draw a line from the left to the right scale through points corresponding to the voltage and current involved. Through the intersection of this line with the diagonal construction line draw a line through the proper point on the beam-length scale at the bottom, and extend it until it intersects the beam-spread scale at the top. The value from this scale will be the beam spread directly.

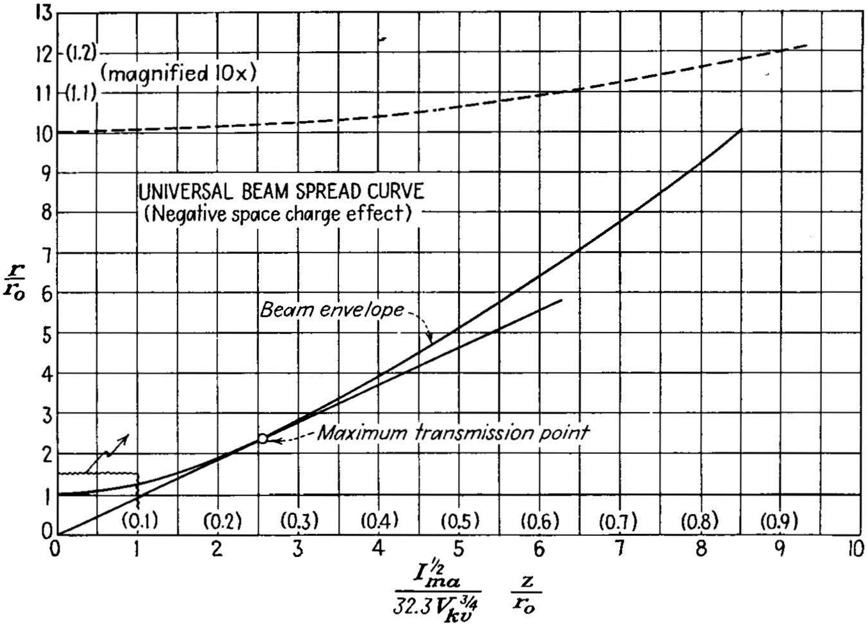


FIG. 15.21.—Universal beam-spread curve. This is a graphical representation of Eq. (15.44).

The functional relations of Eq. (15.44) have been verified experimentally.[1] It is found that although the current and voltage dependence predicted is obeyed correctly the actual values of spread are about six-tenths of the theoretical value (at pressures of 5 × 10^{-7} mm of mercury). This reduction is due to a partial neutralization of negative space charge by the presence of positive ions, which are created by collision of beam electrons with gas molecules. Even at the highest vacuums obtainable, there are theoretically enough positive ions created to

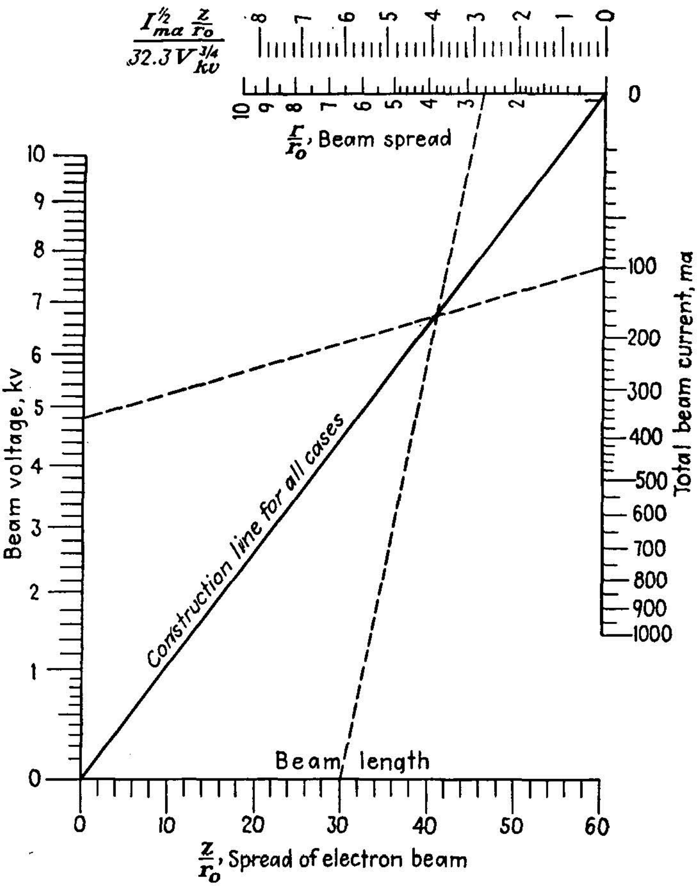


FIG. 15.22.—Nomographic chart of beam spread as a function of beam voltage, beam current, and beam length, as given by Eq. (15.44).

---

[1] See "The Production and Control of Electron Beams," by K. R. Spangenberg, L. M. Field, and R. Helm, published by Federal Telephone and Radio Corporation, New York, 1942.

neutralize the beam completely. This does not occur, however, for the positive ions drift down the beam from positive to negative electrodes and run out of the beam at the cathode end almost as fast as they are created.[1]

A number of important problems can be solved by the use of the universal beam-spread curves of Fig. 15.21. In the design of high-power klystron tubes there arises the practical problem of putting the maximum current down a cylinder of given dimensions with a given voltage. From the above discussion of space-charge spread it would be expected that the beam should be initially convergent, come to a minimum diameter somewhere in the cylinder, and then spread again until it just fills the end of the cylinder. From symmetry it may be predicted that the beam will have its minimum diameter at the middle of the cylinder. It therefore only remains to specify the initial angle of convergence and find the current that can be transmitted. If the beam enters the cylinder at a point on the curve of Fig. 15.21 having coordinates x and y, then for the minimum beam diameter to occur at the center of the cylinder of length l and diameter d

$$\frac{l}{d} = \frac{z}{r} \tag{15.46}$$

where

$$z = xr_0 32.3 \frac{(V_{kv})^{3/4}}{(I_{ma})^{1/2}} \tag{15.47}$$

and

$$r = yr_0 \tag{15.48}$$

Therefore

$$l \frac{(I_{ma})^{1/2}}{(V_{kv})^{3/4}} = 32.3 \frac{xd}{y} \tag{15.49}$$

Therefore, to transmit maximum current at minimum voltage, that is, to have the beam impedance a minimum, the beam must enter the cylinder at a point on the curve of Fig. 15.21 that has the maximum ratio of x to y or the minimum ratio of y to x. This is the point on the curve where a line through the origin is tangent to the curve. The coordinates of this point are

$$x = \frac{1}{32.3} \frac{(I_{ma})^{1/2}}{(V_{kv})^{3/4}} \frac{z}{r_0} = 2.60 \tag{15.50}$$

and

$$y = \frac{r}{r_0} = 2.35 \tag{15.51}$$

---

[1] *Ibid.*

Thus the initial angle is

$$\theta = \arctan\left(\frac{d}{l}\right) \tag{15.52}$$

since substitution of Eqs. (15.50) and (15.51) into Eq. (15.49) gives $\frac{z}{r} = \frac{l}{d}$. The maximum current is transmitted at a given voltage when the beam is so directed on entering the cylinder that in the absence of electrostatic repulsion between the electrons it would converge to a point at the center of the cylinder. Under these conditions the minimum radius is $\frac{1}{2.35} = 0.425$ times the cylinder radius. These relations are illustrated in Fig. 15.23. The value of θ is not extremely critical, for the optimum is very broad. The minimum beam radius may be varied from 0.25 to 0.6 of the cylinder radius, with a loss of only 10 per cent of the maximum current.

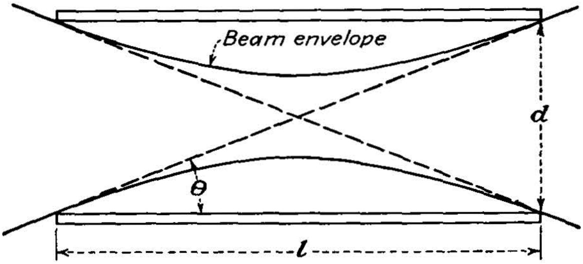


FIG. 15.23.—Diagram showing transmission of maximum current through a cylinder.

When the beam is directed into the cylinder as shown in Fig. 15.23, the maximum value of current that can be transmitted is

$$I_{\max} = 1{,}230(V_{kv})^{3/2}\left(\frac{d}{l}\right)^2 \quad \text{ma}^* \tag{15.53}$$

---

* Even with a strong axial magnetic field to prevent beam-spreading there is a maximum current that can be transmitted along a beam. As current is increased, the potential at the beam center drops below the value at the edge by the amount $V = 0.478\, I_{ma}(V_{kv})^{-1/2}$ volts. This potential difference finally becomes so large that the beam is blocked by space-charge action at a value of $I_{\max} = 1.025\,(V_{kv})^{3/2}$ amperes for a beam completely filling a conducting tube, independent of the tube dimensions. If the beam does not completely fill the tube, then the blocking action will occur at a lower beam current. Greater current can be transmitted if the negative electron space charge is neutralized by positive ions, although even here there is a limit to the current that can be transmitted. See HAEFF, A. V. Space Charge Effects in Elec-

The beam impedance corresponding to this condition is

$$Z = 813 \left(\frac{l}{d}\right)^2 (V_{kv})^{-1/2} \qquad \text{ohms} \tag{15.54}$$

and this is the minimum that can be achieved under the applied restrictions.

As an example of the operation of the above equations, let it be desired to find the maximum current that can be transmitted through a cylinder ¼ in. in diameter and 1 in. in length at 1,000 volts. Equation (15.53) gives $I_{\max} = 77$ ma. The corresponding beam impedance is 13,000 ohms. Actual currents may be slightly higher in practice because of a partial neutralization of the negative space charge by positive ions in the beam.[1]

Higher values of current cannot be passed through a cylinder if it is permitted to waste current. Thus, consider the case of a beam of initially parallel electrons, and let the current be increased. As the current is increased the beam will spread. The current transmitted down a cylinder will increase at first as the effect of increasing the current predominates and then decrease as the effect of beam spread predominates. Maximum current will be transmitted when the cylinder area is 18 per cent of the area of the beam if it has been permitted to spread. Under these conditions the transmitted current is

$$I_{\max} = 305 \left(\frac{d}{l}\right)^2 (V_{kv})^{3/2} \qquad \text{ma} \tag{15.55}$$

which is about one-fourth of the value for a properly convergent beam with no current wasted.

Effect of Secondary Emission. Beam spots will be enlarged slightly by the effect of stray secondary electrons liberated at the limiting apertures. This may or may not be serious depending upon the particular electrode configuration used. In general, secondaries from limiting apertures located near the cathode will give most trouble because these

---

tron Beams, *Proc. I.R.E.*, vol. 27, pp. 586–602, September, 1939; SMITH, L. P., and P. L. HARTMAN, Formation and Maintenance of Electron Beams, *Jour. Appl. Phys.*, vol. 11, pp. 220–229, March, 1940; PETRIE, D. P. R., The Effect of Space Charge on Potential and Electron Paths of Electron Beams, *Elec. Commun.*, vol. 20 (No. 2), pp. 100–111, 1941; PIERCE, J. R., Limiting Stable Current in the Presence of Ions, *Jour. Appl. Phys.*, vol. 15, pp. 721–726, October, 1944.

[1] FIELD, L. M., K. R. SPANGENBERG, and R. HELM, Control of Electron-Beam Dispersion at High Vacuum by Ions, *Elec. Commun.* vol. 24 (No. 1), pp. 108–121, 1947.

will be accelerated by almost the full potential of the system. The stray electrons show as a fuzzy edge to the beam and fairly widespread stray light. It is possible, however, to design electron guns so that stray secondary electrons hardly affect the spot size.

Halation. "Halation" is a term well known in photography. It refers to the "halolike" rings that sometimes appear around bright points of light. The effect is due to light rays being reflected back and forth between the surfaces of a film or, in the case of the cathode-ray tube, back and forth between the faces of glass. When the electron beam strikes the willemite surface on the inside of the end of the tube, a bright spot is formed that radiates in all directions. Those rays which are emitted perpendicular to the glass and moderately close to the perpendicular will pass through the glass and can be seen outside. Rays that are emitted from the spot at a large angle with the perpendicular to the glass will strike the air-glass surface at a low angle and be reflected back into the glass, where they will be reflected back and forth, with a gradual loss of energy due to scattering effects. For the usual glass (index of refraction of about 1.5) only about half the light emitted from the spot on the screen will pass through the glass without multiple reflection. The effect on an outside observer is that there is a bright spot surrounded by a ring of lower intensity. Studies of the effect on the various parameters show that halation is reduced if the fluorescent screen is in moderately poor optical contact with the glass, if it is moderately absorbing, and if the glass is moderately thick.[1]

15.6. High-efficiency Cathodes. When it is desired to obtain high current from a cathode, then the design of the gun becomes complicated by considerations of space charge and the efficiency of the structure, *i.e.*, the fraction of the cathode current that is utilized in the beam, becomes of importance. If the attempt is made to operate the type of gun already described at very high currents, difficulty is immediately encountered in that the space-charge repulsion of the electrons causes the beam to spread so much that a large portion of the cathode current is lost to the various gun electrodes.

The general problem of determining electron paths under conditions of space-charge repulsion is very difficult to solve.[2] As yet no solutions for space-charge flow in cases where the electron paths are curved are known. This means that the design of high-current high-efficiency

---

[1] LAW, R. R., Contrast in Kinescopes, *Proc. I.R.E.*, vol. 27, pp. 511–524, August, 1939.

[2] SPANGENBERG, K. R., Use of the Action Function to Obtain the General Differential Equations of Space Charge Flow in More than One Dimension, *Jour. Franklin Inst.*, vol. 232, pp. 365–371, October, 1941.

cathodes is very difficult. The most successful high-efficiency cathodes are those designed upon a principle enunciated by J. R. Pierce. The laws of space-charge flow are known for a few simple geometries such as plane, cylindrical, and spherical. In each of these cases, the electrons move in straight lines, and the behavior of the electrons can be described in terms of a single parameter representing distance. Pierce has suggested that the conditions of uniform space-charge flow can be achieved in a cathode if a segment of such flow is utilized and the cathode and accelerating electrodes be shaped so as to maintain along the edge of the beam the same potential variation which would exist if there were a uniform extensive space-charge flow. Cathodes designed on this principle are often referred to as "Pierce cathodes."[1]

Parallel Flow of a Rectangular Beam. The laws of space-charge flow of electrons between parallel planes are known (see Sec. 8.2). The potential variation along the direction of electron flow is as the four-thirds power of the distance from the cathode. Hence it would be expected that, if a beam in the form of a rectangular strip were cut out of such a flow and if electrode shapes were such that they would create a potential variation as $x^{4/3}$ along the edge of the beam, the beam would be subjected to the same conditions which exist in the extensive space-charge flow and hence would maintain its property of parallel flow. Specifically, the cathode electrodes must create a potential field with the following properties,

$$V(x,0) = Ax^{4/3} \tag{15.56}$$

where A is merely a numerical constant and

$$\frac{dV}{dy}(x,0) = 0 \tag{15.57}$$

where the edge of the beam is along the line $y = 0$.

The above conditions may be achieved by the electrode configuration of Fig. 15.24. The conditions expressed by the above two equations are achieved along the bisector of a 135-deg inside corner (three-fourths of 180 deg). This follows from the application of the transformation

$$W = Z^{4/3} \tag{15.58}$$

to the lines of constant u and v in the W plane. Hence, if an inside 135-deg corner be split in two and each half be applied to one side of the rectangular strip beam, the conditions for plane-parallel space-charge flow are maintained.

---

[1] PIERCE, J. R., Rectilinear Electron Flow in Beams, *Jour. Appl. Phys.*, vol. 11, pp. 548–554, August, 1940.

The conditions outside of the beam are shown in Fig. 15.25. The gun structure in this case is a unipotential one. A good beam is formed, but it will diverge after passing through the anode because of the lens action of the slot aperture. This type of gun has the advantage that it draws current uniformly from the cathode. The laws of plane-parallel space-charge flow apply directly so that it is easy to design. The anode can be shaped like any of the equipotentials in the plot of Fig. 15.25.

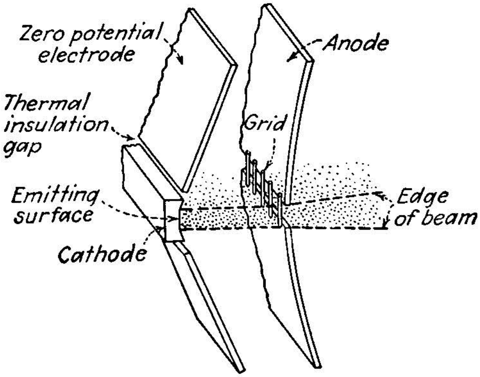


FIG. 15.24.—Pierce cathode structure for rectangular-beam parallel flow.

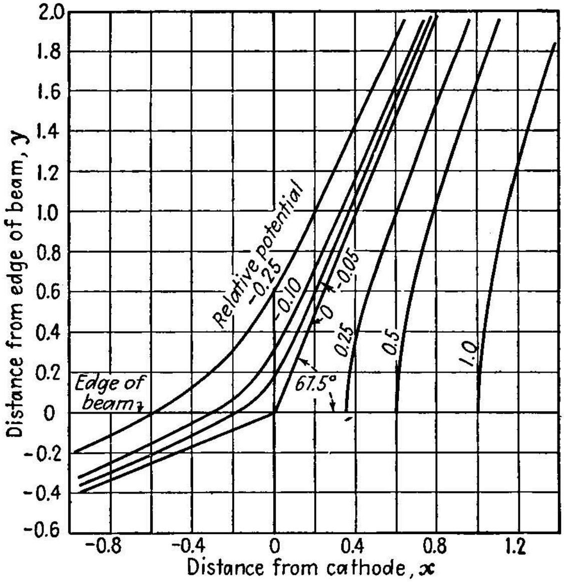


FIG. 15.25.—Potential field required to produce a parallel-flow rectangular beam.

Some difficulty would be encountered in properly restraining the strip beam at its ends. This case serves primarily as an example of the application of the Pierce principle. It also is about the only case for which the electrode shapes can be determined exactly. The cathode electrode is given by

$$y = x \tan 67.5^\circ \tag{15.59}$$

The anode is given by

$$r^{4/3} \cos\left(\frac{4\theta}{3}\right) = \text{const} \tag{15.60}$$

In a practical case it is necessary to have a break in the zero-potential electrode where it joins the emitting portion of the surface in order to improve the thermal efficiency of the cathode. A small gap as shown in Fig. 15.24 will not disturb the flow conditions much.

Parallel-flow Cylindrical Beam. If it is desired to build a unipotential cathode gun producing a parallel-flow cylindrical beam, it is necessary that the cathode and accelerating electrodes produce the following field conditions along the edge of the beam,

$$V(r_0,z) = Az^{4/3} \tag{15.61}$$

and

$$\frac{dV}{dr}(r_0,z) = 0 \tag{15.62}$$

where r_0 is the radius of the beam. This problem has thus far defied analytical solution. Approximate electrode shapes may, however, be found with an electrolytic tank set up to represent this problem. A wedge-shaped piece of electrolyte is used by tilting a tray, placing an insulating strip of material in the tank to represent the edge of the electron beam, and then bending sheet electrodes until shapes are found such that the potential along the insulating strip follows a four-thirds-power law with distance. The insulating strip simulates the electron beam because it imposes the condition of Eq. (15.62). Since no current can flow into the insulator, there will be no component of gradient normal to the strip.

The resulting fields and electrodes have the shape shown in Fig. 15.26. Close to the beam the zero-potential electrode will be a section of a cone with a half angle of 67.5 deg. This is expected from the results of the case studied in the previous subsection. Close to the edge of the beam the conditions are almost identical with the plane-rectangular-strip case, and hence a 67.5-deg angle with the zero-potential electrode is indicated. At great distances from the beam the zero-potential electrode will be a section of a cone with a half angle of 71 deg. A cone of this angle

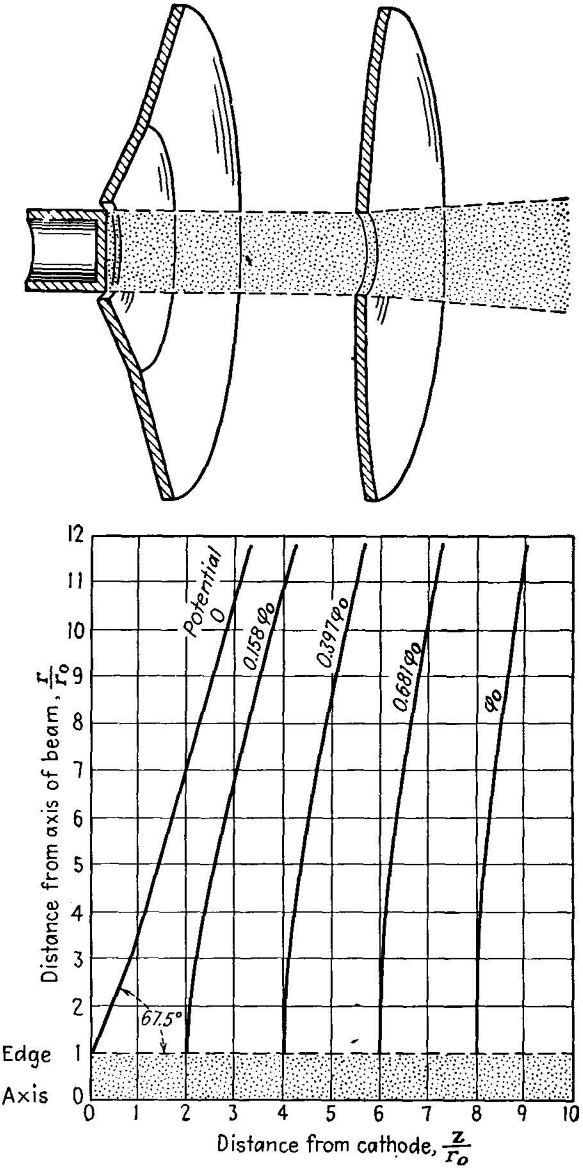


FIG. 15.26.—Unipotential cathode structure required to produce a parallel-flow cylindrical beam, and the associated field.

will give a four-thirds-power variation of potential along its axis. The anode will be a surface of revolution curved in the direction of increasing potential. Such a set of electrodes will produce a parallel cylindrical beam. If the region beyond the anode is field-free, however, the beam will diverge owing to the lens action of the aperture in the anode. For this reason a better type of gun is sought.

Convergent Radial Flow of a Conical Beam. The divergent effect of the aperture in the anode noted in the two cases studied above is unavoidable. Even if the aperture is covered with a grid, the individual holes in the grid will each have an action similar to that of the large aperture and with the same focal length. The only difference will be that the grid will produce more scattering of the electrons and hence will produce a divergent beam with a less sharply defined edge. Because of the inescapable divergent action of the aperture it is desirable to produce a beam which is initially quite strongly convergent so that the divergent action of the aperture in the anode will leave the beam still convergent.

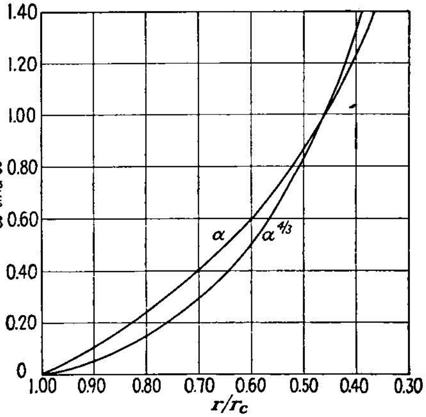


FIG. 15.27.—Curves of α and $\alpha^{3/4}$ as a function of r_c/r. See Appendix VII for values of α^2.

A convergent beam may be had by utilizing a circular conical section of the radial flow between concentric spheres with the cathode outside (see Sec. 8.4). The radial current flow in a cone of semiangle θ cut out of a sphere of radius r_c is

$$I = \frac{0.928(V_{kv})^{3/2}\sin^2\left(\frac{\theta}{2}\right)}{\alpha^2} \quad \text{amperes} \tag{15.63}$$

where V_{kv} is the beam voltage in kilovolts and α is the function of $\frac{r}{r_c}$ given by Eq. (8.32). The factor $\sin^2\left(\frac{\theta}{2}\right)$ takes account of the fact that the current flow takes place not over the entire sphere but merely over a cone of semiangle θ cut out of the sphere. The voltage as a function of radius is given by

$$V_{kv} = \frac{1.051 I^{2/3} \alpha^{4/3}}{\left[\sin\left(\frac{\theta}{2}\right)\right]^{4/3}} \tag{15.64}$$

in which the significant relation is the dependence upon the four-thirds power of α. Curves of α and $\alpha^{4/3}$ as a function of the ratio of cathode to

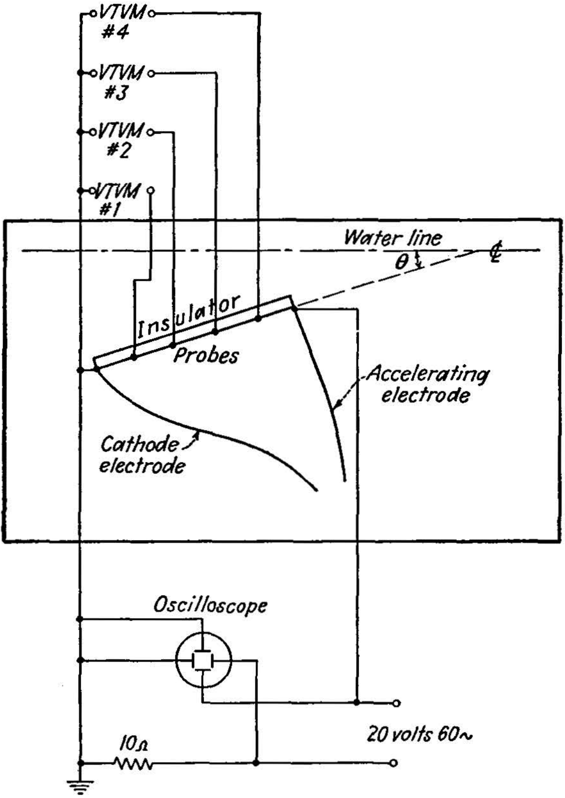


FIG. 15.28.—Electrolytic tank arrangement for the determination of unipotential gun-electrode shapes.

anode radius are given in Fig. 15.27. Numerical values are given in Appendix VII. Except for a proportionality constant, the curve of $\alpha^{4/3}$ is a universal curve of voltage variation as a function of radius for spherical flow. Equation (15.64) is therefore one of the conditions that applies along the edge of the conical beam. The other condition is

$$\frac{dV}{d\theta} = 0 \tag{15.65}$$

where θ is the polar angle of a spherical coordinate system. Equation (15.65) will apply inside the beam and at its edge, but not outside.

Electrode shapes for a conical section of a spherical flow can be determined by means of an electrolytic tank, as described before. A specific arrangement suitable for this purpose is shown in Fig. 15.28. A tilted tank is used to obtain a wedge-shaped portion of electrolyte. The vacuum-tube voltmeters are conveniently made with adjustable sensitivities. These sensitivities should be adjusted so that when the desired potential distribution is achieved each voltmeter gives some convenient deflection, such as half scale. This procedure makes the determination of the electrode shapes relatively simple and rapid. The electrodes are conveniently made of thin copper sheet so that they can be bent into any shape. The cathode electrode should make an angle of 67.5 deg with the beam edge simulated by the insulating strip. The oscilloscope shown serves to check the power factor of the electrolyte and the presence of contact potentials. It is connected to plot the Lissajous pattern of current against voltage. The pattern should show a straight line or at worst a long, thin ellipse, corresponding to a small phase angle. The voltmeter probes should be spaced to give equal increments of voltage rather than of distance.

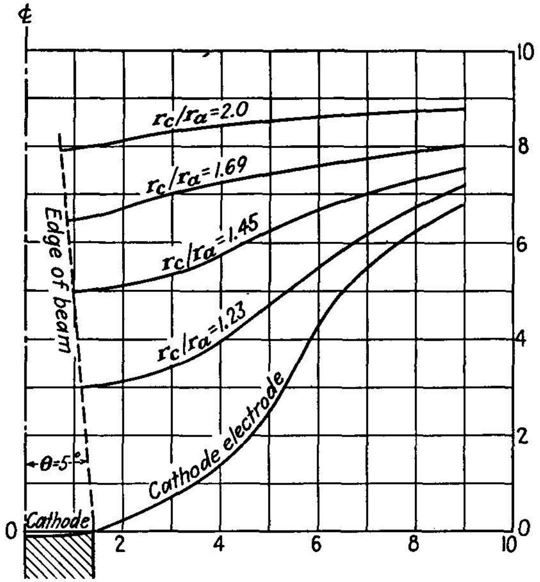


FIG. 15.29.—Unipotential gun-electrode shapes for the production of a 5-deg convergent beam.

Some resultant electrode shapes for different angles of beam con-

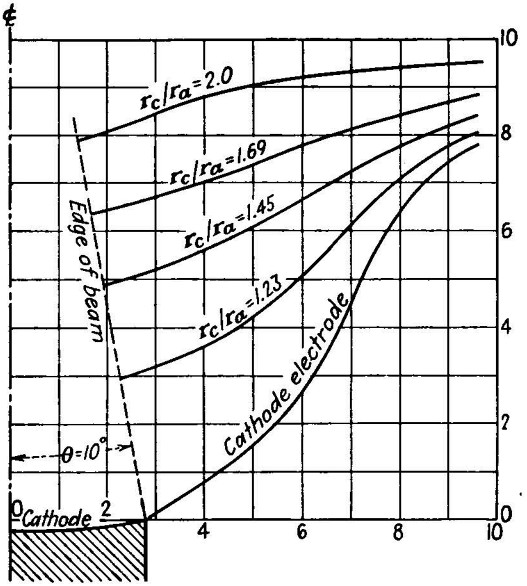


FIG. 15.30.—Unipotential gun-electrode shapes for the production of a 10-deg convergent beam.

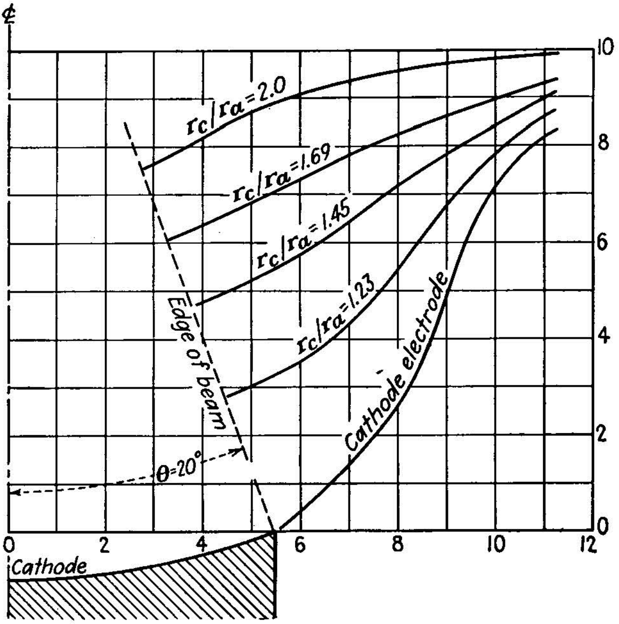


FIG. 15.31.—Unipotential gun-electrode shapes for the production of a 20-deg convergent beam.

vergence are shown in Figs. 15.29 to 15.32.[1] These curves are universal in that they will hold for any magnitude of applied voltage and that the anode electrode can be any of the equipotential curves shown.

The anode aperture in a unipotential Pierce cathode will always give rise to a divergent focusing action. This means that the beam on leaving the anode will always be less convergent than on entering it. The focal length of the anode aperture lens will be given very closely by

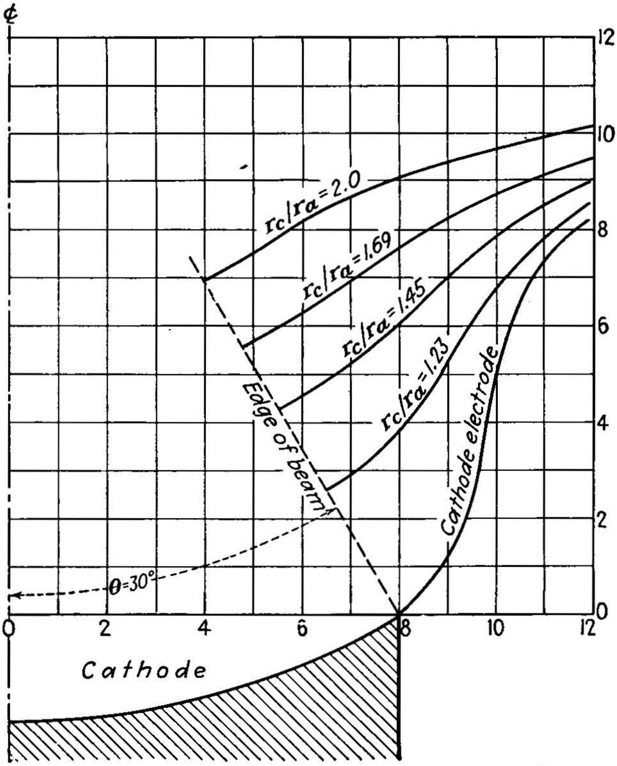


FIG. 15.32.—Unipotential gun-electrode shapes for the production of a 30-deg convergent beam.

$$f = \frac{4V}{E} \tag{15.66}$$

from the aperture-lens formula of Eq. (13.56) on the assumption that the

---

[1] These electrode shapes were determined by Robert Helm and were first published by SPANGENBERG, FIELD, and HELM, *op. cit.* See also HELM, R., K. R. SPANGENBERG, and L. M. FIELD, Cathode-Design Procedure for Electron-Beam Tubes, *Elec. Commun.*, vol. 24 (No. 1), pp. 101–107, 1947.

gradient of potential beyond the anode aperture is zero. Upon evaluating the focal length by means of Eq. (15.66), it is found that

$$\frac{f}{r_c} = \frac{-3\alpha}{\dfrac{d\alpha}{dR}} \tag{15.67}$$

where $R = \frac{r}{r_c}$. A curve of $\frac{f}{r_c}$ as a function of $\frac{r}{r_c}$ is given in Fig. 15.33.

The focal length of the aperture lens being known, it is possible to determine the exit angle γ of the beam for any entrance angle θ. The basic dimensions of the electron gun are shown in Fig. 15.34. For this structure the usual lens formula applies,

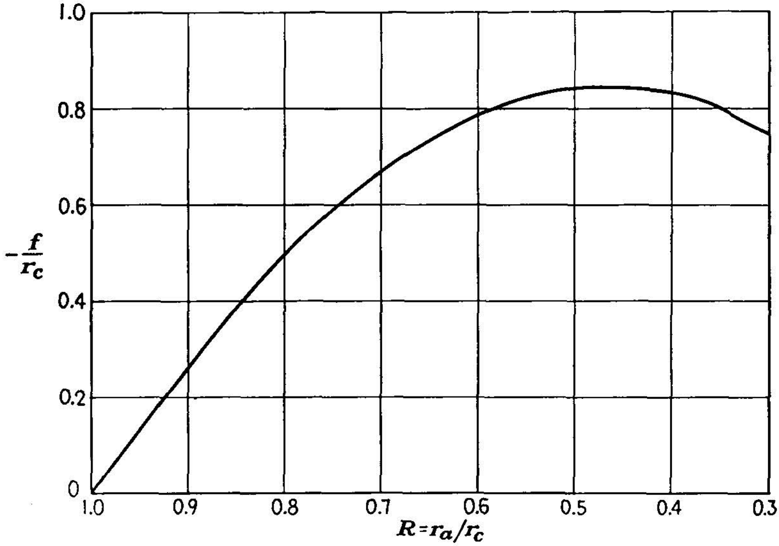


FIG. 15.33.—Focal length of a unipotential convergent-beam gun as a function of the ratio of anode to cathode radius.

$$\frac{1}{r_a} - \frac{1}{b} = -\frac{1}{f} \tag{15.68}$$

In this equation the lens has been assumed to be located at the intersection of the anode sphere with the axis rather than in the plane of the aperture. Equation (15.68) can be put into the form

$$\frac{b}{r_c} = \frac{1}{\dfrac{r_c}{f} + \dfrac{r_c}{r_a}} \tag{15.69}$$

This equation shows that the distance from the anode to the beam focal point (in the absence of space-charge spread) depends only upon the ratio of cathode to anode radius, since focal length, by Eq. (15.67), depends only upon this ratio. The distance from the anode to the beam focal point is independent of the entrance angle θ of the beam because the ratio of entrance to exit angle for any ray of the beam will be a constant. A curve of the distance from the anode to the beam focal point as a function of the ratio of cathode to anode radius is given in Fig. 15.35. The lens action is convergent only for ratios of cathode to anode radius greater than 1.455. Smaller ratios of cathode to anode radius approach the plane-electrode case, which is strongly divergent.

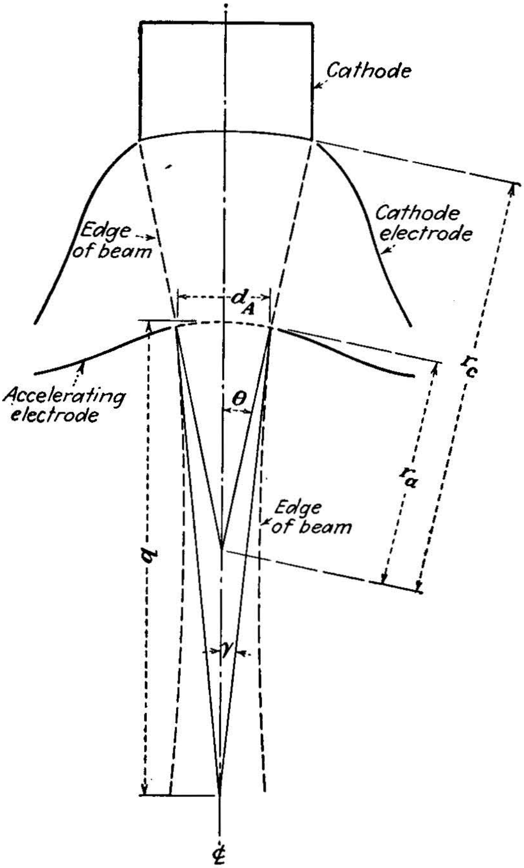


FIG. 15.34.—Diagram of a unipotential convergent-beam gun.

The relation between the entrance and exit angles in Fig. 15.34 is readily deduced from the geometry and has the form

$$\frac{\sin \theta}{\sin \gamma} = \frac{b}{r_a} = \frac{b}{r_c}\frac{r_c}{r_a} \tag{15.70}$$

in which the symbols have the significance given in Fig. 15.34. The above equation shows that the ratio of the sines of the entrance and exit angles depends only upon the ratio of cathode to anode radius, since $\frac{b}{r_c}$ is a function of this ratio, also. The relation between the entrance angle θ, the exit angle γ, and the ratio of cathode to anode radius is given in Fig. 15.36. This representation has the advantage over the many others possible in that the curves of constant θ, γ, and $\frac{r_c}{r_a}$ are straight lines. The region of divergent lens action lies below the θ axis and so does not appear on the curve sheet. This curve sheet has been converted into a universal design chart by superimposing curves of constant beam perveance on the other curves. Beam perveance in this case is defined as

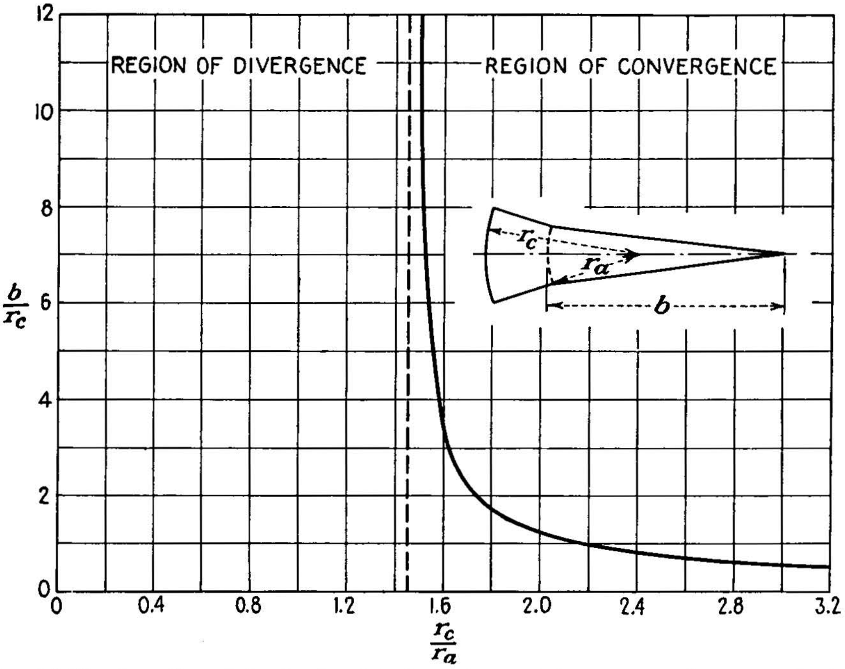


FIG. 15.35.—Location of the focal point of a unipotential convergent-beam gun in the absence of space-charge repulsion.

$$G = \frac{I_a}{(V_{kv})^{3/2}} \tag{15.71}$$

which, from Eq. (15.63), has the value

$$G = \frac{0.928 \sin^2 \left(\dfrac{\theta}{2}\right)}{\alpha^2} \tag{15.72}$$

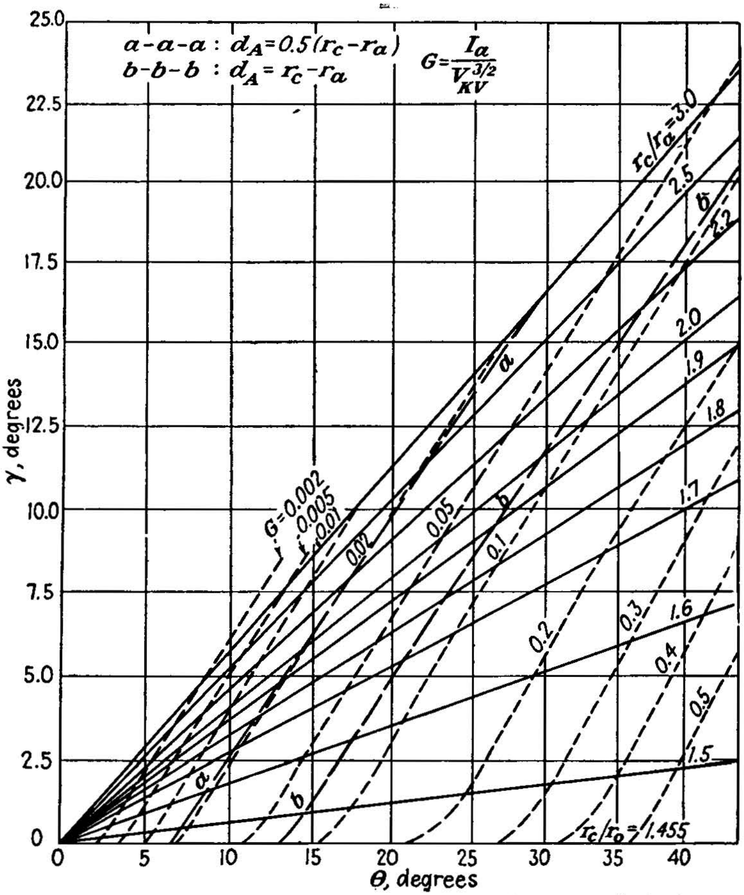


FIG. 15.36.—Unipotential convergent-beam gun chart for use in designing guns like that of Fig. 15.34.

A nomographic chart of the relation of Eq. (15.71) giving perveance for any beam voltage and current is given in Fig. 15.37.

By means of Figs. 15.37 and 15.36 it is a simple matter to select values of cathode to anode radius and entrance and exit angles for any

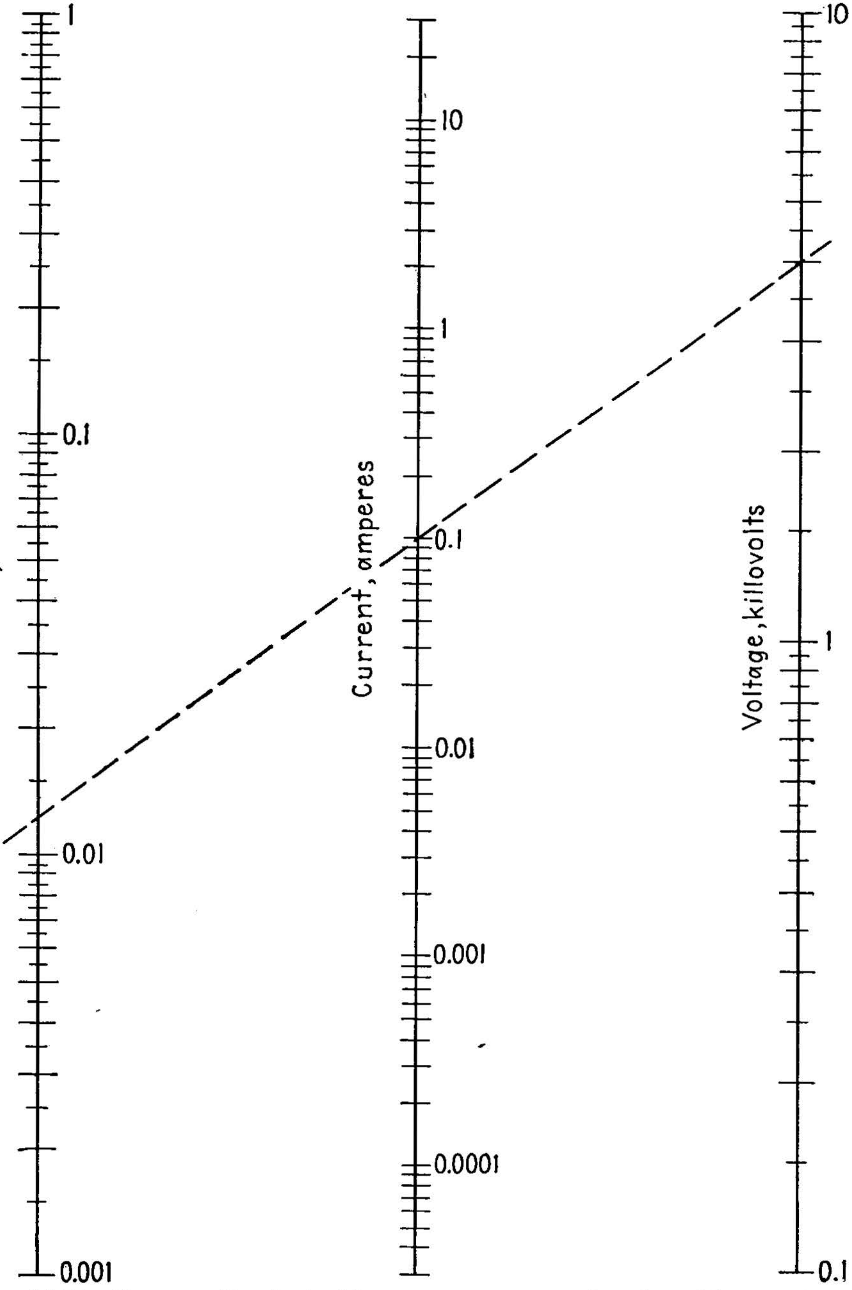


FIG. 15.37.—Nomographic chart of beam perveance as a function of beam voltage and current, for use in Fig. 15.36.

desired voltage and current. Figure 15.36 gives the relation between the four variables $G, \frac{r_c}{r_a}, \theta$, and γ. Any two may be taken as independent variables. When their values are prescribed, the values of the other two variables are determined. Where the anode aperture has no grid, it is necessary only to make sure that the aperture is not too big so that it will not disturb the field in the cathode-anode region. Also shown in Fig. 15.36 are contours of anode aperture diameter equal to 50 per cent and 100 per cent of the cathode-anode distance. The value of the gradient of potential at the center of the cathode is reduced about 5 per cent when the anode aperture diameter is 70 per cent of the cathode-anode spacing. Aperture diameters larger than this should not be used without attempting to compensate for the reduced cathode gradient by changing the electrode shapes. When a spherical grid is used to cover the anode aperture, no such limitations are encountered.[1]

Guns designed from the chart of Fig. 15.36 have performed as theoretically predicted. It is not unreasonable to expect that 90 or possibly 95 per cent of the cathode current will become useful beam current and that current densities as high as half the maximum theoretical value as limited by thermal-emission velocities will be attained.

For determination of the beam action after leaving the anode aperture, reference is made to the universal beam-spread curve of Fig. 15.21. If this curve is entered at the right point, the subsequent beam envelope will be like that of the universal curve to the left of the point, it being assumed that the beam leaving the anode aperture is convergent. For convenience in entering the universal curve, its slope at any point is given, along with the universal curve replotted in Fig. 15.38. This slope is given by

$$M = \tan \gamma \tag{15.73}$$

and when divided by $K = \frac{I^{1/2}}{(V_{kv})^{3/4}}$ gives the proper scale value for entering the slope curve. From the corresponding point on the envelope curve it can then be determined where the minimum diameter of the beam will occur and what the subsequent beam spread will be. Actual beam-spreading action is usually only about two-thirds of the values predicted here because of a partial neutralization of the negative space charge by positive ions in the beam.

Example: Suppose it is desired to design a cathode that will put a beam of 40 ma at a voltage of 1,000 volts through a cylinder 6 cm in length and 1 cm in

---

[1] For an alternative treatment of this subject see SAMUEL, A. L., Some Notes on the Design of Electron Guns, *Proc. I.R.E.*, vol. 33, pp. 233–240, April, 1945. This article also contains design data for the line-focus case.

diameter. Then the conditions of Fig. 15.23 apply, and tan γ will have the value of 0.167, which corresponds to a value of γ of 9.5 deg. Opposite this value of exit angle and the corresponding value of perveance on the design chart of Fig. 15.37, it is found that a cathode-beam angle of 22.5 deg and a ratio of cathode to anode radius of 2.13 are required. The required electrode shapes may be found from a slight interpolation of the shapes given in Figs. 15.31 and 15.32.

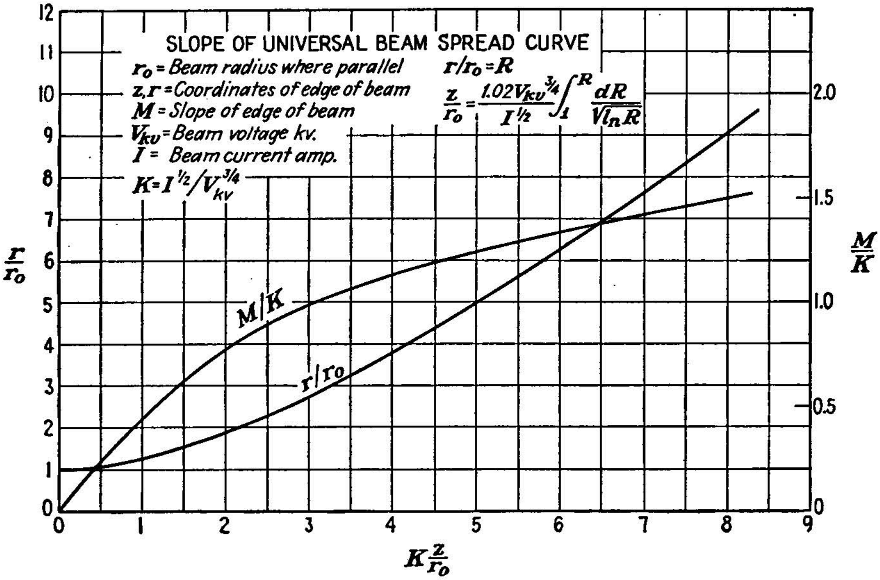


FIG. 15.38.—Slope of universal beam-spread envelope.

15.7. Ultra-high-frequency Deflection Effects. There is a limit to the frequency of wave forms, which can be observed on a cathode-ray tube with electrostatic deflecting plates. The deflection equation [Eq. (6.23)] is evaluated for a direct potential and is valid for alternating potentials only if the beam electrons' transit time through the deflecting plates is so small a fraction of the cycle that the deflecting plate voltage does not change appreciably while any single electron is influenced by it. In a representative tube having, say, a beam voltage of 1,000 volts and a deflecting-plate length of 2 cm, the transit angle will not become appreciable until the frequency is of the order of 50 mc, at which frequency ordinary sweep circuits have failed and the problem of getting the voltage on the deflecting plates is considerable. However, there are an increasing number of applications in which it is desired to observe very high and ultra-high-frequency phenomena so that it is worth while to make a brief study of transit-time effects to determine the limitations of ordinary tubes and serve as a guide to the design of special tubes.

Let the notation of Fig. 6.2 be used and let the *static-deflection case* be reviewed for comparison. The crosswise acceleration of an electron entering the field between the plates is

$$\frac{d^2y}{dt^2} = \frac{eV_d}{ma} \tag{15.74}$$

where y is transverse displacement, e and m are charge and mass of the electron, respectively, V_d is the deflecting potential, and a is the deflecting-plate spacing. A first integration of this equation gives

$$\frac{dy}{dt} = \frac{eV_dt}{ma} \tag{15.75}$$

in which the constant of integration is zero since $\frac{dy}{dt} = 0$ when $t = 0$. The transverse velocity at the end of deflecting plates of length b is

$$\frac{dy}{dt} = \frac{eV_db}{mav_0} \tag{15.76}$$

where v_0 is the velocity of the beam electrons. From this equation the deflection y_s of a spot on a screen a distance l from the end of the deflecting plates is

$$\frac{y_s}{l} = \frac{eV_db}{mav_0{}^2} \tag{15.77}$$

since $\frac{y_s}{l} = \frac{\frac{dy}{dt}}{v_0}$. The deflection sensitivity, or deflection per unit deflecting voltage, is

$$A_0 = \frac{y_s}{V_d} = \frac{leb}{mav_0{}^2} \tag{15.78}$$

The *dynamic-deflection case* can be handled in much the same manner.[1] For this case let the instantaneous voltage between deflection plates

---

[1] Of the rather extensive periodical literature on this subject the following articles are the most fundamental:

HOLLMANN, H. E., Die Braunsche Röhre bei sehr hohen Frequenzen, *Hochfrequenz. und Elektroakustik*, vol. 40, pp. 97–103, September, 1932.

LIBBY, L. L., Cathode Rays for the Ultra-high Frequencies, *Electronics*, vol. 9, pp. 15–17, September, 1936.

BOWIE, R. M., Cathode Ray Wave Form Distortion at Ultra-high Frequencies, *Electronics*, vol. 11, pp. 18–19, 29, February, 1938.

HOLLMANN, H. E., Ultra-high Frequency Oscillography, *Proc. I.R.E.*, vol. 28, pp. 213–219, May, 1940.

HARRIES, J. H. OWEN, Deflected Electron Beams, *Wireless Eng.*, vol. 21, pp. 267–277, June, 1944.

be given by $V_d \cos \omega t$. Then the acceleration of an electron at any instant t is

$$\frac{d^2y}{dt^2} = \frac{eV_d \cos \omega t}{ma} \tag{15.79}$$

where all the symbols have their previous significance. A first integration gives

$$\frac{dy}{dt} = \frac{eV_d\,(\sin \omega t - \sin \omega t_0)}{ma\omega} \tag{15.80}$$

where t_0 is the time the electron enters the alternating field and the particular value of the constant of integration given results from the condition that $\frac{dy}{dt} = 0$ when $t = t_0$. For simplification in interpretation let the time the electron spends in the alternating field be represented by

$$T = t - t_0 \tag{15.81}$$

$$\frac{dy}{dt} = \frac{eV_d\,[\sin \omega(T + t_0) - \sin \omega t_0]}{ma\omega} \tag{15.82}$$

Integration of Eq. (15.80) gives

$$y = \frac{eV_d}{ma\omega^2}\,[\cos \omega t_0 - \cos \omega t - \omega(t - t_0) \sin \omega t_0] \tag{15.83}$$

where the particular value of the constant of integration results from the condition that $y = 0$ when $t = t_0$. The above equation is better written in terms of the time the electron is exposed to the alternating field as

$$y = \frac{eV_d}{ma\omega^2}\,[\cos \omega t_0 - \cos \omega(T + t_0) - \omega T \sin \omega t_0] \tag{15.84}$$

This equation gives the path of the electron when used parametrically with the expression

$$x = v_0 T \tag{15.85}$$

for the time the electron is exposed to the alternating field. A set of curves showing the path of electrons *between the deflecting plates* over a period of two complete cycles is shown in Fig. 15.39. It is seen that the path for any starting time is a straight line at some angle with the axis with a superimposed transverse sinusoidal motion. The angle of the straight-line component of the path depends upon the starting angle, being zero when the electron enters at a peak of the instantaneous alternating voltage and maximum when it enters at an instant of zero alternating voltage. The amplitude of the alternating component of transverse sinusoidal motion is the same for any starting time and is proportional to the deflecting gradient of potential and inversely proportional to the square of frequency. As an aid to visualization of the electron behavior, it may be stated that this path is the same as that of a ball rolled along a plank which rocks with a sinusoidal motion, the ball having an initial velocity parallel to the long dimension of the plank.

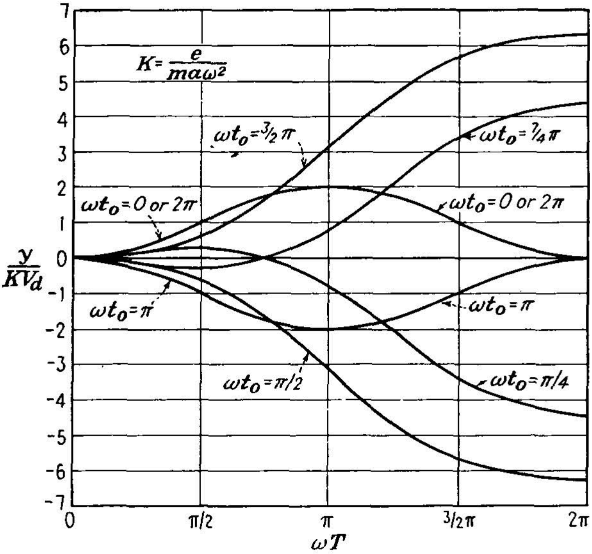


FIG. 15.39.—Path of an electron between deflecting plates when the transit time is large compared to the period of the deflecting voltage.

The electron will move in a straight-line path in the field-free region beyond the end of the deflecting plates with a slope determined by Eq. (15.82) and $\frac{dx}{dt} = v_0$. As before,

$$\frac{y_s}{l} = \frac{\frac{dy}{dt}}{v_0} \tag{15.86}$$

so that the dynamic-deflection sensitivity, or deflection per volt, is

$$A(\omega) = \frac{el}{ma\omega v_0} [\sin \omega(T + t_0) - \sin \omega t_0] \tag{15.87}$$

Upon invoking Eq. (15.78) and simplifying by trigonometric transformation, the ratio of the dynamic to the static deflection sensitivity is

$$\frac{A(\omega)}{A_0} = 2 \frac{v_0}{b\omega} \sin \frac{\omega T}{2} \cos \omega \left(t_0 + \frac{T}{2}\right) \tag{15.88}$$

which by virtue of the fact that $T = \frac{b}{v_0}$ is readily written in the simpler form

$$\frac{A(\omega)}{A_0} = \frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \cos \omega \left(t_0 + \frac{T}{2}\right) \tag{15.89}$$

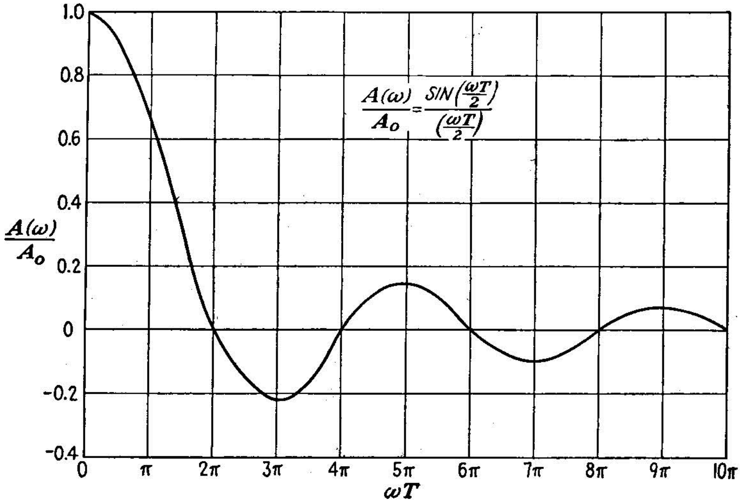


FIG. 15.40.—Ratio of dynamic to static sensitivity of deflecting plates as a function of transit angle.

From this equation several important properties of deflection at the ultrahigh frequencies are evident. In the first place the deflection is sinusoidal with time so that there is no distortion of waves of a *single frequency*.[1] In the second place the maximum value of the deflection varies as the ratio of the sine of half the transit angle through the deflecting plates to half the transit angle. A curve of deflection sensitivity as a function of transit angle ωT is given in Fig. 15.40. It is seen that the deflection is apparently zero whenever the transit angle through the deflecting plates is some integral multiple of 2π radians. This is consistent with the observation in Fig. 15.39 that the slope of the electron trajectory is zero every 2π radians. The above has assumed that the deflection in passing through the deflecting plates is small compared with the subsequent deflection over the relatively long distance to the screen, l. The exact actual deflection will, of course, be the sum of the values given by Eqs. (15.84) and (15.86). This means that the deflection will not be quite zero even when the transit angle through the deflecting plates is a

---

[1] BOWIE, *op. cit.*

multiple of 2π but will be a minimum given by the value from Eq. (15.84) when T has the value $2n\pi$, where n is a positive integer.

It is of interest to record the values of deflecting-plate transit angles for which the dynamic deflection sensitivity drops to some arbitrary fractions of the static sensitivity. The dynamic deflection sensitivity will be 0.9 of the static sensitivity when the transit angle through the deflecting plates is 0.794 radian, or 45.5 deg. It will be 0.707 of the static sensitivity when the transit angle is 2.78 radians, or 159 deg. It will be a minimum when the transit angle is 2π radians, or 360 deg.

When the same ultra-high-frequency voltage is applied to both horizontal and vertical deflecting plates, the resultant Lissajous figure will not be a straight line because of the phase shift occurring between the two sets of deflecting plates. For a pure sine wave the resultant figure will be an ellipse. When the applied voltage contains harmonics, the resultant figures will have odd shapes not encountered at low frequencies.[1] Harmonic analyses can be made from the resultant figures. They can also be made from the so-called "inversion spectrograms," which are obtained by applying the complex ultra-high-frequency wave to the vertical plates, allowing the beam voltage to vary sinusoidally over a suitable range, and applying a fixed magnetic field parallel to the vertical deflecting field. The transverse deflection of the beam will vary with the beam voltage because of the effect of the magnetic field, and the different harmonic components of the wave under observation will experience different vertical deflections at the different velocities, in accordance with Eq. (15.89).

15.8. Photography of Cathode-ray Traces. In the observation of wave-form phenomena it is frequently important to obtain a permanent record. This is readily done by simply taking a picture of the screen trace. The science of photographing cathode-ray traces has now reached such a state of development that, except for special applications, it has rendered other methods of recording wave forms virtually obsolete.[2–6]

---

[1] HOLLMANN, H. E., Ultra-high Frequency Oscillography, *Proc. I.R.E.*, vol. 28, pp. 213–219, May, 1940.

[2] FELDT, R., Photographing Patterns on Cathode Ray Tubes, *Electronics*, vol. 17, pp. 130–137, 262, 264, 266, February, 1944.

[3] GRAY, C., Notes on Cathode Ray Photography, *Radio Research Laboratory Seminar Rept.*, Jan. 23, 1945.

[4] Cathode Ray Tubes, *RCA Manual* TS2, pp. 86–93, 1935.

[5] "Photographic Papers for Recording Purposes," Eastman Kodak Company, Rochester, N. Y., 1942.

[6] "Photographic Materials Available for Use with Oscillograph, Cathode Ray Tubes, and Similar Recording Instruments," Eastman Kodak Company, Rochester, N.Y., 1941.

In taking pictures of cathode-ray-tube traces the experimenter has under control at least nine different factors all of which will contribute to the contrast of the resulting picture. These are

1. Cathode-ray-tube beam power.
2. Type of fluorescent screen.
3. Writing speed of the beam trace.
4. Exposure time.
5. Magnification of the camera lens.
6. Lens speed, or stop.
7. Film sensitivity.
8. Developer.
9. Development time.

TABLE VIII
PHOTOGRAPHIC PROPERTIES OF COMMON FLUORESCENT SERIES

	Type of screen*		
	P1 medium-persistence green	P2 long-persistence green	P5 short-persistence blue
Visual brightness, ft-lamberts	7.5	1.55	0.9
Relative brightness (Weston 603 meter, Viscor filter)	8.3	1.7	1.0
Relative film speed	0.63	0.25	1.0
Test film	Agfa SSS Ortho	Agfa SSS Ortho	Agfa Fluorapid Blue
Photographic efficiency	0.076	0.15	1.0
Ft-lamberts for equal photographic effect	13.2	6.7	1.0

* See Appendix IV for specific characteristics.

In the manipulation of the above factors the objectives sought are a dense negative trace with a high contrast. The effect of the separate factors listed above will now be briefly discussed.

Beam Power. It has already been mentioned that the brightness of a beam trace is approximately linear with beam voltage at a fixed current. It is also approximately linear with beam power. Hence the greater the beam power for a given spot size, the greater the brightness of the spot and the easier it is to get a satisfactory picture.

Screen Types. The three most commonly used screens today are the P1 medium-persistence green, P2 long-persistence green, and P5 short-persistence blue. The principal characteristics of these screens are listed in Table VIII.

In the above table, visual brightness gives the relative response of the eye. Relative brightness is a standard meter reading. For each screen the type of film that gives the densest trace for a standard developing procedure is used. Photographic efficiency is the ratio of relative film speed to relative brightness. Data in the last row are obtained from the reciprocal of photographic efficiency.

Writing Speed. The writing speed is simply the speed of the beam trace. Naturally, the greater the speed, the less the photographic effect.

Time, Stop, and Magnification. These factors are interdependent. Although it is possible to give specific coefficients that will determine exposure time for a given set of operating conditions, these conditions are subject to so much variation that it is almost necessary in all cases to obtain the correct exposure time by a trial set of pictures. When the best exposure time has been so determined for one set of operating conditions, times for other conditions are readily determined by simple formulas.

In the photography of *recurrent traces*, the exposure time can be made as long as desired, subject only to the limitation of fogging due to stray light. The exposure time necessary will be determined by the beam power, the lens stop, and the image magnification according to the formula

$$t = \frac{KF^2(M+1)^2}{w} \qquad (15.90)$$

where t is exposure time (conveniently, sec)
K is exposure constant (determined by experiment)
F is lens stop (ratio of lens focal length to aperture diameter)
M is image magnification (ratio of object to image size)
w is beam-power density (watts per cm^2 of fluorescent area as determined from beam power and trace area)

If the correct exposure time is experimentally determined for one set of operating conditions, it is a simple matter to evaluate the coefficient K and the above formula then gives the exposure time for any other set of operating conditions.

In the photography of *transient phenomena* where only a single trace of the pattern occurs the camera lens is left open and the beam-trace brightness is determined by the writing speed and the beam power. The corresponding photographic image density is determined by the lens stop and image magnification. The relation between writing speed and the other factors is given by

$$v_s = \frac{CW}{F^2(M+1)^2} \qquad (15.91)$$

where C is proportionality constant

v_s is spot velocity (conveniently, km per sec)

W is beam power (conveniently, watts)

F is lens stop

M is image magnification

When a suitable exposure is obtained by test, it is a simple matter to calculate the proportionality constant C. The above formula then gives the relation between the four parameters involved for any other set of operating conditions to obtain the same film-trace density. At ordinary potentials (2.5 kv), recordable writing speeds are of the order of 5 to 50 km per sec. With standard tubes and high accelerating potentials (10 kv) writing speeds as high as 1,000 km per sec have been recorded.

The maximum lens aperture that can be used is, of course, determined by the lens speed. Lenses with f ratings of $f/4.5$ are usually available.

TABLE IX

PHOTOGRAPHIC-FILM SENSITIVITIES

Film	Weston Speed Rating (Daylight)
Agfa SSS Pan	200
Agfa SSS Ortho	100
Eastman Ortho X	100
Eastman Superpan Press	100
Eastman Super XX	100
Defender Ortho X-F	50
Defender X-F Pan	50
Agfa Fluorapid Blue	
Eastman X-ray Blue	

Lenses with speeds as high as $f/1.5$ are available for Leica and Contax cameras.

Examination of Eqs. (15.90) and (15.91) shows that the film density may be increased, other factors being equal, by reducing the magnification to get a smaller image. The gain that can be effected in this way is not large, however, and the maximum gain possible over a magnification of 1 is a factor of 4 in writing speed or equivalent exposure time.

Film Sensitivity. A number of special and ordinary films are available for cathode-ray-trace photography. In Table IX is given a list of available films in the approximate order of their sensitivity.

The Agfa SSS Ortho gives the best result of all films for the P1 and P2 screens. The Agfa Fluorapid Blue gives the best results of all films for the P5 screen. It should be noted, however, that the Eastman Super XX requires only about twice as much exposure as the best film in all cases to give an equivalent image.

Film speeds can be increased 50 to 100 per cent by hypersensitizing the film with ammonia vapor. The well-known expedients of hypersensitizing with mercury vapor or preexposing to get above the fog level do not seem to do much good in the photography of cathode-ray traces, where the work is often at minimal levels of exposure.

Developers and Development. Standard developers and standard developing procedures with the usual precautions can be used. The commercial developers D19 (high contrast), D72, and Ansco 47 are satisfactory. Development can be carried beyond the recommended time to increase contrast up to the point where fogging becomes excessive or the gelatin softens too much.

CHAPTER 16

ULTRA-HIGH-FREQUENCY EFFECTS IN CONVENTIONAL TUBES

16.1. Introduction. It is well known that, as frequency is raised, tubes are progressively less effective as amplifiers and oscillators. Amplifiers require greater driving power, and the output drops off correspondingly. If the frequency is raised high enough, the gain of an amplifier will drop to unity or less. At the same time this is happening, the input impedance of the amplifier drops, as does also the maximum impedance that can be realized in the plate circuit. Oscillator output drops even more rapidly with frequency than does amplifier output. At the same time the limitations on output change. At low frequency the output for continuous operation is often limited by the plate dissipation. As the high-frequency limit of oscillation is reached, the grid dissipation commonly becomes the limiting factor while the plate hardly gets hot at all.

All the above effects come about because of a combination of electronic and circuital phenomena. Depending upon the design of the tube, electronic considerations may limit the output before the circuit limitations do as the frequency is raised, or vice versa.

16.2. Causes of Decreased Output at Ultra-high Frequencies. Numerous factors contributing to a reduction of output at ultra-high frequencies can be listed. The total number of contributing factors can be divided into roughly three groups. These are

1. Circuit-reactance limitations.
2. Circuit- and tube-loss limitations.
3. Electron-transit-time limitations.

At the ultra-high frequencies there exists a situation which is quite different from that which exists at low frequencies. At low frequencies the electrical circuits and the tube are quite distinct. As frequency increases, this ceases to be true and it is found that part of the resonant circuits exist inside of the tube. This comes about because electrode leads have a small but finite inductance. As frequency rises into the ultra-high classification, the reactance of this inductance becomes appreciable. This means that the voltage across the external terminals will not appear across the electrodes. In addition, while the interelectrode capacities may be small, at the ultra-high frequencies they may be a large fraction of the capacity required to give resonance in an external circuit. As such, they represent a limitation in terms of actual operation. The combination of the electrode-lead inductance and the interelectrode capacity may give rise to resonances in the ultra-high-frequency region. Even if resonances do not occur, the combination of the reactances within the tube may constitute a network that mismatches the equivalent tube generator and the load. All in all, there are a number of respects in which the circuit reactances combine to limit tube performance at the ultra-high frequencies. These detrimental effects can be combated in two ways, (1) by making the tube smaller, which reduces the inductances and capacities in direct proportion to the linear dimension, and (2) by making the tube structure such that the electrode leads can be incorporated into external concentric-line resonators.

The power losses associated with a tube and circuit all tend to increase with frequency. At ultra-high frequencies all currents flow in thin surface layers because of skin effect. The associated resistance and losses increase with the square root of frequency because the thickness of the layer in which the current flows decreases in this manner as frequency increases. Glass and other insulating supports have losses associated with the molecular movements produced by the electric fields. These "dielectric hysteresis losses," as they are called, will usually vary approximately as the first power of frequency. In addition, there will be appreciable radiation from an exposed piece of wire such as an electrode lead. The power radiated from a short length of wire carrying current increases as the square of the frequency. All the above factors contribute to a general reduction in tube efficiency as frequency is increased. Resistance losses may be made low by increasing the area of the surfaces carrying current. Dielectric losses may be reduced by proper positioning of glass with respect to points of low electric field. Radiation losses can be reduced by enclosing the tube and circuit or by using a concentric-line construction so that the tube and circuit fields are entirely confined.

Electron-transit-time effects can contribute to reduced tube output in many ways. If the transit times of the electron are appreciable fractions of the ultra-high-frequency cycle, then plate current will lag negative grid voltage and there will be a reduced output in an oscillator because plate current and voltage are out of phase. Associated with increased transit time there is a dispersal, or debunching, of electrons, which has the result that plate-current pulses are not so sharp as the pulses liberated from the cathode. In addition, there will be an energy interchange between the electric fields and the electrons in flight so that as frequency increases the grid-input impedance will have a resistance component which decreases with frequency even though no electrons strike the grid. Furthermore, all the tube constants such as the amplification factor will become complex instead of real numbers as a result of a shift in phase and what is generally a reduction in magnitude. There is not much that can be done about electron-transit-time effects except to raise the voltages and reduce the dimensions, both of which processes have definite limitations. In addition, certain tube types are less adversely affected by electron-transit-time effects than others. The tetrode, for instance, suffers less from the adverse effects of electron transit time than does the triode.

In the subsequent sections there will be given a brief analytical treatment of all the above effects. No complete analysis that embraces all aspects of ultra-high-frequency tube operation is available. Rather, the process of estimating the situation is that of looking through different windows, corresponding to different avenues of approach, and then trying to piece the complete picture together from the partial revelations obtained.

16.3. Onset of Tube-reactance Limitations. The most important reactance encountered in a vacuum tube is that associated with the lead inductance. It is possible to speak of the inductance of a piece of straight wire or of an unclosed circuit in general. It must be borne in mind, however, that the inductance of the unclosed circuit is considered as part of some closed circuit the total inductance of which is equal to the sum of the self-inductances of all its parts plus the sum of the mutual inductances of each one of the component parts relative to every other part. In cases where the mutual inductances between various parts of the same closed circuit are small the total inductance is simply the sum of the self-inductances of the component parts. Taken in this sense, the inductance of a straight piece of wire at very high frequencies is

$$L = 0.00508l\left(2.303 \log_{10} \frac{4l}{d} - 1 + \frac{d}{2l}\right) \quad \text{microhenrys} \quad (16.1)$$

where l is the length of the wire in inches and d is the wire diameter in inches. The last term in the parentheses is negligible if l is more than $100d$. A family of curves giving the dependence of inductance upon wire length and diameter is shown in Fig. 16.1. The inductance is seen to increase as the wire diameter is made smaller or as the wire length is increased. In tubes, therefore, leads should be as large as possible in diameter and as short as possible in length. As an example of how large lead reactances can be, consider the case of a lead that is 100 mils in diameter and 1 in. in length, as frequently occurs in small transmitting tubes. This lead has an inductance of approximately 0.015 microhenry, as may be seen from Fig. 16.1. At 500 mc this represents a reactance of 47 ohms, which is fairly high.

Cathode-inductance-feedback Limitations. Since the tube lead reactances are internal to the tube, there will be coupling between the input and output circuits due to grid and plate currents flowing through the common cathode lead inductance. This will have the effect of introducing feedback into the stage involving the tube and may cause the grid-input impedance to be affected adversely. If all the tube lead inductances and interelectrode capacities are considered, some rather complex relations are encountered.[1,2] In general, the effect of the internal tube

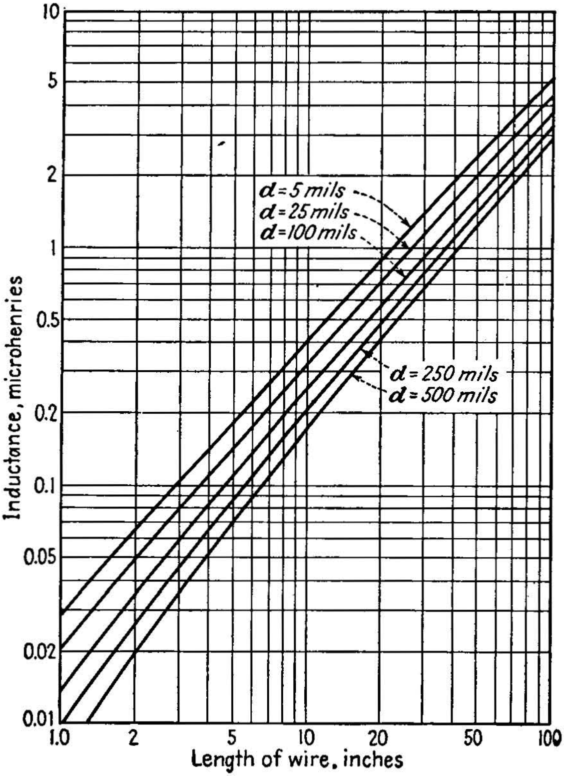


FIG. 16.1.—Inductance of a round straight wire.

---

[1] STRUTT, M. J. O., and A. VAN DER ZIEL, The Causes for the Increase of the Admittances of Modern High-frequency Amplifier Tubes on Short Waves, *Proc. I.R.E.*, vol. 26, pp. 1011–1032, August, 1938. Contains good bibliography.

[2] SARBACHER, R. I., and W. I. EDSON, "Hyper and Ultra-high Frequency Engineering," pp. 431–436, Wiley, New York, 1943.

reactances is to decrease the impedances presented at the tube input terminals. This may be seen by considering the onset of reactance effects in a triode at ultra-high frequencies.

Consider the triode circuit of Fig. 16.2, in which there are considered only the effect of the cathode lead inductance and the cathode-grid capacity.[1] Then the signal voltage V_s differs from the voltage that appears between the grid and the cathode by the voltage drop in the cathode lead inductance. Thus

$$V_s = V_g + j\omega L_c I_p \qquad (16.2)$$

But the plate current will be approximately proportional to the negative of the product of the grid input voltage and the mutual conductance of the tube since at ultra-high frequencies the plate-load resistance will usually be small.

$$I_p = G_m V_g \qquad (16.3)$$

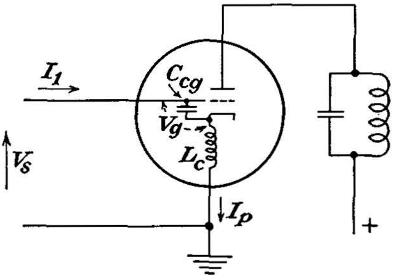


FIG. 16.2.—The equivalent circuit of a triode amplifier at ultra-high frequencies.

The input current to the tube will produce a voltage drop across the grid-cathode capacity that is equal to the tube input voltage

$$V_g = \frac{I_1}{j\omega C_{cg}} \qquad (16.4)$$

where I_1 is the input-circuit current. Making this substitution into Eq. (16.2) along with Eq. (16.3),

$$V_s = \frac{I_1(1 + j\omega L_c G_m)}{j\omega C_{cg}} \qquad (16.5)$$

in which the second term in the numerator is numerically small compared with unity. Accordingly, the input admittance of the tube is approximately

$$Y_{in} = \frac{I_1}{V_s} = j\omega C_{cg}(1 - j\omega L_c G_m) \qquad (16.6)$$

since $(1 + a)^{-1}$ is approximately equal to $1 - a$ when a is small compared with unity. The first term of the input admittance will be recognized as the normal capacitive susceptance of the tube. The second term is a real positive term representing a conductive component of input admittance and having the value

$$G_{in} = \omega^2 L_c C_{cg} G_m \qquad (16.7)$$

---

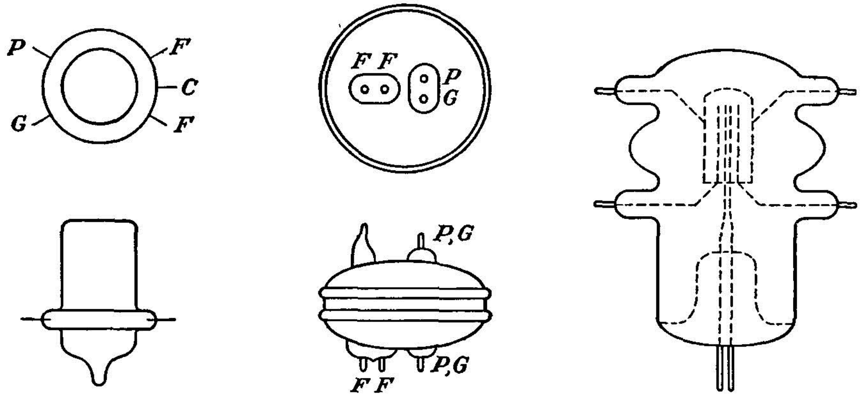
[1] FREEMAN, R. L., Input Conductance Neutralization, *Electronics*, vol. 17, pp. 24–25, October, 1939.

This input conductance corresponds to a resistance in parallel with the input capacity whose value decreases inversely as the square of the frequency. This resistance consumes power, which increases as the square of the frequency for a given driving voltage. There is no real loss of power involved here. The driving power consumed in this fashion is simply transmitted to the plate circuit. The equivalent input resistance encountered here can be fairly low. For a tube with a cathode lead inductance of 10^{-8} henry and a mutual conductance of 9,000 micromhos operating at a frequency of 30 mc the equivalent input resistance is of the order of 25,000 ohms. In addition to the input conductance due to cathode-inductance feedback there is a similar component of conductance due to electron-transit-time effects, as will be seen. The transit-time conductance varies in the same fashion with frequency, *i.e.*, as the square of frequency The equivalent resistances that are due to cathode-inductance feedback and electron transit time are in parallel, and any measurement will involve the effect of both. In triodes the equivalent resistance due to transit-time effects may be smaller than that due to cathode-inductance feedback. In multielectrode tubes the transit-time resistance will usually be much larger than the feedback resistance. The components of the input conductance can be separated by making measurements with and without a bit of external inductance inserted in series with the cathode lead.

Interelectrode-capacity Limitations. In addition to the lead inductance, the interelectrode capacitances play an important role in the operation of tubes in the ultra-high-frequency region. Interelectrode capacitances due to active parts of the tube structure are incapable of reduction beyond a certain point. However, in many tubes the interelectrode capacity results largely from capacity between the leads in parts of the tube where electrons do not flow. Thus the receiving-tube practice of bringing all the tube leads out through a single glass stem at the bottom of the tube is very bad from the standpoint of the interelectrode capacity.

Arrangements that bring out the leads separately as much as possible are preferred from the standpoint of low interelectrode capacity. Examples of such arrangements are to be found in the acorn tube, the doorknob tube, and certain low-power radiation-cooled tubes (see Fig. 16.3). In the acorn tube the leads are brought out radially in such a way that the capacity between them is greatly reduced. The leads of the doorknob tube likewise are brought out rather well spaced. In the radiation-cooled tubes the leads are brought out widely separated. In addition, in some forms there are double leads, which can be paralleled to reduce the inductance. When this is done, the interelectrode capacity is almost entirely that found in the active portion of the tube where the electron flow is concentrated. Further reduction here is possible only by scaling down the size of the tube, which in turn limits the power the tube can develop because the heat-dissipation capacities are reduced.

If the resonant circuits of the tubes are made of lumped reactance elements, then the lead inductance and interelectrode capacity determine the highest frequency at which the tube can be operated. This highest frequency is the frequency at which the interelectrode capacity resonates with the shortest external connection between the tube electrodes. For the tubes shown in Fig. 16.3 this frequency will be of the order of 2,000 mc for the acorn tube, 1,000 mc for the doorknob tube, and 500 mc for the radiation-cooled tube. These frequencies may be exceeded if a transmission-line type of resonant circuit is used, for then the connecting link between electrodes may effectively be pushed inside the tube.



955
Acorn tube
(a)

368
Door knob tube
(b)

Radiation cooled
Type-227A
(c)

FIG. 16.3.—Common ultra-high-frequency tube types.

The interelectrode capacity is an important factor in determining what plate-load resistance can be realized. This in turn determines the gain and power output that can be made available. The equivalent shunt resistance of a parallel resonant circuit can be written in a number of ways, among which there are

$$R_{sh} = \frac{1}{\omega_0{}^2 RC} \tag{16.8}$$

$$R_{sh} = \frac{Q}{\omega_0 C} \tag{16.9}$$

where R is the equivalent series resistance, C is the total capacity determining the resonance, and ω_0 is the angular resonant frequency. For operation at a given frequency it is seen that in order to increase the shunt resistance it is necessary to decrease the capacity. This can be done up to a point by reducing the capacity and increasing the inductance to maintain the same frequency of resonance. Eventually, this process is limited by the fact that the capacity external to the tube has been reduced to zero and the shunt resistance is determined by the tube interelectrode capacity. The larger the interelectrode capacity, the smaller the shunt resistance that can be realized. Accordingly, the power output tends to drop off as the load resistance or as the square of the frequency as frequency increases.

For amplifier operation the gain–band-width product is of considerable importance. This product is one that depends upon the ratio of the tube mutual conductance to the circuit capacity, various numerical coefficients applying for different circuits.[1] Consider the case of a tube with a simple single tuned circuit as a coupling and frequency-determining element between it and the next stage. The gain of such a stage is approximately equal to the product of the tube mutual conductance and the circuit shunt resistance.

$$A = G_m \frac{Q}{\omega C} \tag{16.10}$$

where A is the stage voltage gain. The corresponding band width depends upon the circuit Q and the operating frequency according to

$$\Delta f = \frac{f_0}{Q} \tag{16.11}$$

Accordingly, the gain–band-width product is

$$A \, \Delta f = \frac{1}{2\pi} \frac{G_m}{C} \tag{16.12}$$

The gain–band-width product can be increased by reducing the circuit capacity up to the point where that capacity is the interelectrode capacity of the tubes involved. Accordingly, it is again desirable to have tubes with well-separated leads to reduce the interelectrode capacity.

16.4. The Nature of Currents Induced by Electron Motion at Ultrahigh Frequencies. *The Plane Diode without Space Charge.* At low frequencies, the current flowing to any electrode in a vacuum tube is considered as resulting from the arrival of electrons at the electrode in accordance with the equation

$$i = nev \tag{16.13}$$

---

[1] WHEELER, H. A., Wide-band Amplifiers for Television, *Proc. I.R.E.*, vol. 27, pp. 429–438, July, 1939.

where n is the number of electrons per unit length of beam, $-e$ is electron charge, and v is the electron velocity. This concept is satisfactory as long as the time required for an electron to move from one electrode to another is so short that it can be considered as being virtually instantaneous. If, however, the time required is appreciable when measured in time units of the period of the alternating frequency involved, then this concept is no longer adequate. The question arises as to whether there is any current in the electrode circuit while the electron is en route. It turns out that there is such a current; and since the electron transit time may be an appreciable fraction of the period involved it needs to be considered. When the electron transit time is appreciable, it is no longer true that the electrode current is determined by the rate of arrival of electrons at the electrode. The current may be greater or less.

The correct concept of electrode current is that it is determined by rate of change of the charge on the electrode induced by the electron in flight. This induced current is the real current, and its magnitude is readily determined. Consider the situation shown in Figs. 16.4*a*, *b*, and *c*. Here there is shown an electron moving from the cathode to the plate of a plane-electrode diode. From the electron there emanate $-e$ lines of electric flux, which terminate on a like amount of positive charge on the cathode and plate. When the electron is close to the cathode as in Fig. 16.4*a*, then most of the lines from the electron terminate on the cathode, with the result that the positive charge so induced on the cathode is larger than the positive charge induced on the plate. When the electron is midway between cathode and the plate as in Fig. 16.4*b*, then half the lines terminate on the cathode and half terminate on the plate, with the result that the induced charges on cathode and plate are equal. When the electron is close to the plate as shown in Fig. 16.4*c*, then more lines terminate on the plate than on the cathode.

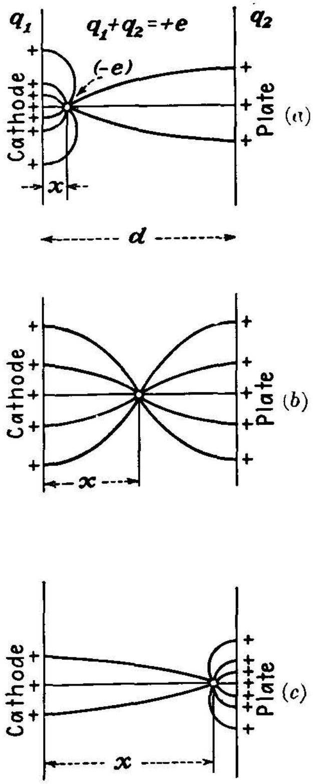


FIG. 16.4(*a*,*b*,*c*).—The electric field of a single electron in flight between parallel planes.

The exact magnitude of the induced charges described above may be calculated from the equality of the work done in transferring the charge from cathode to plate to the energy gained by the electron in its movement. Let q_1 be the charge induced on the cathode and q_2 the charge induced in the plate. As the electron moves across from cathode to plate, the battery effectively transfers a charge q_2 from cathode to plate. This means that the battery does work of the amount Vq_2. At the same time the electron has moved a distance x under the influence of a field of strength $\frac{-V}{d}$ so that work of the amount $\frac{Vex}{d}$ has been done on it. Accordingly,

$$Vq_2 = \frac{Vex}{d} \tag{16.14}$$

from which

$$q_2 = \frac{ex}{d} \tag{16.15}$$

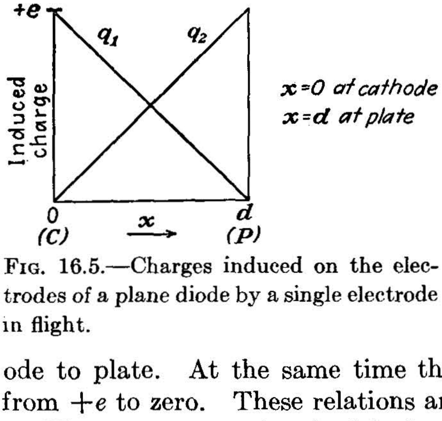


FIG. 16.5.—Charges induced on the electrodes of a plane diode by a single electrode in flight.

Since the total charge induced on both cathode and plate must equal $+e$, it must be true that

$$q_1 = e\left(1 - \frac{x}{d}\right) \tag{16.16}$$

This means that the induced *charge* on the plate grows linearly with electron position from a value of zero to $+e$ as the electron moves across the diode from cathode to plate. At the same time the induced cathode charge decreases from $+e$ to zero. These relations are shown in Fig. 16.5.

The current associated with the induced charges resulting from the motion of an electron is given simply by the time rate of charge. Thus

$$i = \frac{dq_2}{dt} = \frac{e}{d}\frac{dx}{dt} = \frac{ev}{d} \tag{16.17}$$

This is the current flowing to the plate. The above is one of the most important fundamental relations in the field of high-frequency-tube behavior. The magnitude of the circuit current associated with the electron flight is shown in Fig. 16.6. For the parallel-plane diode considered here, the field will be linear, and the velocity of the electron if emitted with zero velocity will increase linearly with time. This gives rise to a triangular-shaped pulse of current. Furthermore, the induced current depends only on the electron velocity and is independent of electron position to the extent that the position is independent of the velocity. It is seen that current starts to flow the moment the electron enters the interelectrode space and continues until it reaches the plate. It is not true that current flows only when the electron reaches the plate. The area under the current pulse is $+e$ from Eq. (16.17).

The total current that flows to any electrode is found by adding up the triangular pulses of current associated with each electron. This summation will generally result in a current curve that lags the emitted-electron current by an angle proportional to the product of the angular frequency and the transit time. Currents may even be induced in electrodes to which no electrons flow if the number or velocity of the electrons approaching the electrode is different from the number or velocity of the electrons receding from it. This is the case with the control grid in ordinary triodes and multielectrode tubes when operating Class A.[1,2]

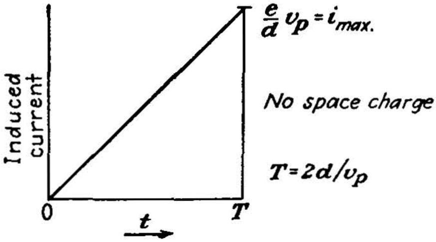


FIG. 16.6.—Induced current resulting from an electron in transit in a parallel-plane diode without space charge.

The General Case. The relations given above are a special case of a more general relation in that they are restricted to the plane-electrode diode in the absence of space charge. The general relation that applies for any field configuration is

$$i_n = e\frac{dV_n}{ds}v \tag{16.18}$$

where i is the induced current flowing to any electrode, e is the magnitude of the electron charge, and $\frac{dV_n}{ds}$ is the gradient of potential in the direction of the electron velocity that would exist at the electron's instantaneous position if the given electrode were raised to *unit* positive potential and

---

[1] NORTH, D. O., Analysis of the Effects of Space Charge on Grid Impedance, *Proc. I.R.E.*, vol. 24, pp. 108–136, January, 1936. One of the earliest papers to make use of the relation $i = \frac{ev}{d}$.

[2] THOMPSON, B. J., Review of Ultra-high Frequency Vacuum Tube Problems, *RCA Rev.*, vol. 2, pp. 146–155, October, 1938.

all other electrodes were grounded.[1-5] It will be noticed that the relation of Eq. (16.18) reduces to the relation of Eq. (16.17) for the plane-electrode case. From the general relation of Eq. (16.18) it is seen that the induced current is maximum when the electron is moving along a path for which the gradient of potential resulting, when the electrode in question is raised to unit positive potential and all other electrodes are grounded, is itself maximum. If the electron were to follow an equipotential line under the conditions stated above, the induced current would be zero.

Induced Currents in the Space-charge-limited Diode. The shape of the induced-current pulse associated with a single electron transit in a plane diode is slightly different when the diode is space-charge-limited from what it is when it is not. This comes about because the potential variation with distance is a four-thirds-power law in the presence of space charge, while it is linear in its absence. As a result, the electron velocity follows a two-thirds-power law of variation with distance for the space-charge-limited case, whereas it follows a one-half-power law in the absence of space charge. Accordingly, the velocity of an electron varies with the square of the time in the space-charge-limited case, whereas it varies linearly with time in the absence of space charge. In addition, the transit time in the presence of space charge has been shown in Sec. 8.10 to be 50 per cent greater than in its absence. As a result, the potential, velocity, and induced current in the space-charge-limited case will have the form shown in Fig. 16.7. For comparison, the corresponding relations that hold in the complete absence of space charge are shown dotted. The current pulse with space charge is sharper, which means that its fundamental component is smaller and is retarded more than in the space-charge-free case. The difference between the behavior with and without complete space-charge saturation is, however, small enough so that for most qualitative evaluations the triangular

---

[1] SHOCKLEY, W., Currents Induced by a Moving Charge, *Jour. Appl. Phys.*, vol. 9, pp. 635–636, October, 1938.

[2] RAMO, SIMON, Currents Induced by Electron Motion, *Proc. I.R.E.*, vol. 27, pp. 584–585, September, 1939.

[3] JEN, C. K., On the Induced Current and Energy Balance in Electronics, *Proc. I.R.E.*, vol. 29, pp. 345–349, June, 1941.

[4] JEN, C. K., On the Energy Equation in Electronics at Ultra-high Frequencies, *Proc. I.R.E.*, vol. 29, pp. 464–466, August, 1941.

[5] This relation results from the fact that the charge induced on one of a system of grounded conductors by an electron is eV_n, where V_n is the potential to which the location point of the electron is raised when unit potential is applied to the electrode in question and all other electrodes are grounded. The induced current is then simply the time rate of change of charge. See SHOCKLEY and RAMO, *op. cit.*

current pulse is sufficiently accurate. The area under the current pulse in this case is again $+e$.

Currents Induced in the Electrodes of a Triode. The relations discussed above may be applied to triodes quite successfully to give an indication of the magnitudes and phases of the currents induced in the different electrodes. Between the electrodes the potential fields will resemble those of a diode in that potential will vary linearly with distance except in the immediate vicinity of the grid wires. The gradients of potential will be determined by the electrode voltages and the tube dimensions. The currents induced in any electrode can be calculated from the general relation of Eq. (16.18). To find the current induced in the cathode it is necessary to know the gradient of potential which exists at the elec-

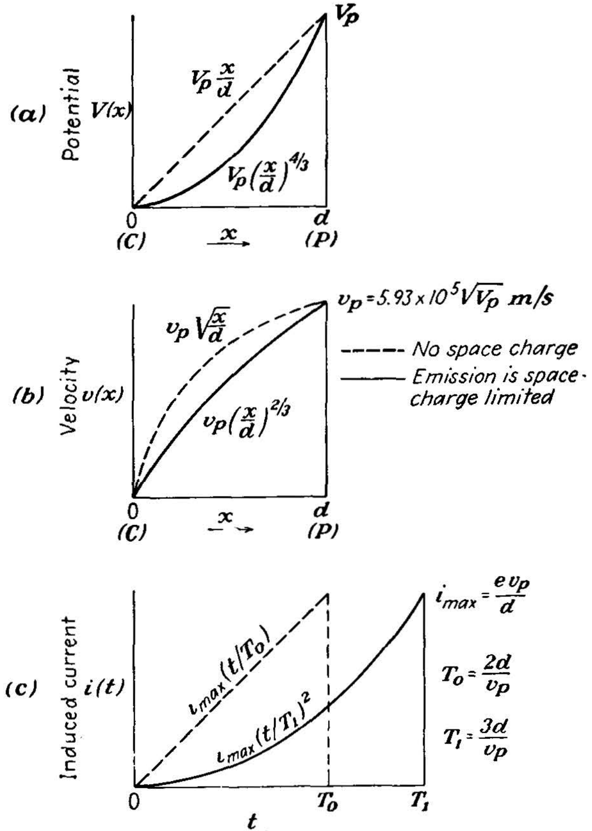


FIG. 16.7.—Induced current resulting from a single electron in transit in a plane diode whose emission is space-charge-limited.

tron's location when the cathode is raised to unit potential and the other electrodes are at zero potential. The potential contours in a triode for this condition are shown in Fig. 16.8*a*. The effects of space charge have been neglected in setting up these profiles. When the gradient of potential is known, the induced current is simply the product of the gradient, the electron charge, and the *actual* velocity. Since the gradient of potential is negative in both the cathode-grid and in the grid-plate region, the induced cathode currents will always be negative. Furthermore, the induced current will be greater in magnitude by approximately the amplification factor of the tube when the electron is in the cathode-grid region than when it is in the grid-plate region.

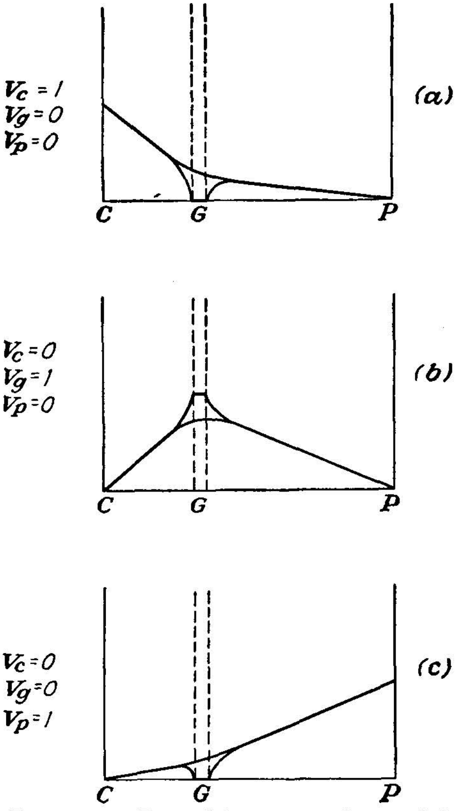


FIG. 16.8.—Potential contours in a triode used in determining the currents induced in the electrodes by the transit of a single electron.

To determine the currents induced in the grid wires, it is necessary to know the potential distribution that results when the grid is at unit positive potential and the cathode and plate are at zero potential. The resulting potential profiles are shown in Fig. 16.8*b*. The induced grid current will be positive when the electron is in the cathode-grid region but negative when the electron is in the grid-plate region. The magnitudes of the currents will be approximately in the inverse ratio of the cathode-grid distance and the grid-plate distance *for a given electron velocity* since the magnitudes of the potential gradient are in this inverse ratio.

To determine the induced plate current it is necessary to consider the potential configuration that results when the plate is at unit positive potential and the grid and cathode are at zero potential. The resulting potential profiles are sketched in Fig. 16.8*c*. Then, by Eq. (16.18), the induced current to any electrode is simply the product of the electron charge, the electron velocity, and the corresponding gradient of potential. The resulting electrode currents are listed in Table X. It will be noted that the induced electrode current is always of the form $\frac{ev}{d}$. For any position of the electron it will also always be true that the sum of the cathode, grid, and plate current is zero.

TABLE X

CURRENTS INDUCED IN THE ELECTRODES OF A PLANE-ELECTRODE TRIODE BY THE PASSAGE OF A SINGLE ELECTRON

	Cathode current I_c	Grid current I_g	Plate current I_p
Electron in cathode-grid region......	$\frac{-ev(1+\mu)}{d_{gp}+(1+\mu)d_{cg}}$ $\cong \frac{-ev}{d_{cg}}$	$\frac{+ev\mu}{d_{gp}+(1+\mu)d_{cg}}$ $\cong \frac{ev\mu}{(1+\mu)d_{cg}}$	$\frac{ev}{d_{gp}+(1+\mu)d_{cg}}$ $\cong \frac{ev}{\mu d_{cg}}$
Electron in grid-plate region......	$\frac{-evd_{gp}}{d_{gp}[d_{gp}+(1+\mu)d_{cg}]}$ $\cong \frac{-ev}{\mu d_{cg}}$	$\frac{-ev\mu d_{cg}}{d_{gp}[d_{gp}+d_{cg}(1+\mu)]}$ $\cong \frac{-ev\mu}{(1+\mu)d_{gp}}$	$\frac{ev(d_{gp}+\mu d_{cg})}{d_{gp}[d_{gp}+(1+\mu)d_{cg}]}$ $\cong \frac{ev\mu}{(1+\mu)d_{gp}}$

Expressions similar to those for the plane-electrode triode can also be worked out for the cylindrical-electrode triode. These expressions will be more involved than those for the plane-electrode triode and will involve the radial position of the electron. This comes about because the gradient of potential is not constant in the interelectrode spaces.

For the cylindrical diode, for instance, the induced cathode and plate currents are

$$I_c = -\frac{ev}{r}\ln\left(\frac{r_p}{r_c}\right) \tag{16.19}$$

and

$$I_p = \frac{ev}{r}\ln\left(\frac{r_p}{r_c}\right) \tag{16.20}$$

16.5. Onset of Transit-time Effects in Triodes. As the frequency of operation of a vacuum tube is raised, there is finally reached a frequency at which electron-transit-time effects make themselves felt. These are evident first with the appearance of a conductive component of the grid input admittance; *i.e.*, a definite amount of power is required to drive the grid even though it does not intercept any electrons. In addition, the mutual conductance and amplification factor become complex and smaller in magnitude, having a negative phase angle that increases in magnitude with frequency. Of these various effects the appearance of a conductive component of grid input admittance is most important. This component is one that at first grows as the square of the frequency. The existence of this component and its dependence upon frequency and other factors can be demonstrated by examining the induced grid currents along the lines indicated in the previous section.

Consider first the grid current induced by the transit of a single electron from cathode to plate. Ordinarily the grid will be negative, but above its cutoff value. The electron, however, passes readily through the space between grid wires where the potential is positive. In the cathode-grid region the gradient of potential is nearly constant at a small positive value determined by the cathode-grid distance and the mean potential of the grid plane. In the grid-plate region the potential gradient is again positive, but at a much higher value. Potential contours for a typical condition are shown in Fig. 16.9*a*. The associated electron velocities will as a first approximation be considered linear with time in both the cathode-grid and the grid-plate region because the potential gradients in these regions are nearly constant. The electron velocity is as shown in Fig. 16.9*b*. It is seen to increase linearly with time at a relatively slow rate in the cathode-grid region and at a relatively faster rate in the grid-plate region.

The corresponding current induced in the grid electrode will be as shown in Fig. 16.9*c*. The induced current has the form of the product of the electron velocity as in Fig. 16.9*b* by the potential as shown in Fig. 16.8*b*. The sign of the current changes as the electron passes the grid plane, for here the electron changes its relative direction with respect to the grid. The induced-current pulse as a result consists of a positive triangular pulse followed by a negative trapezoidal pulse. The areas of the positive and negative pulses will nearly equal plus and minus e respectively, yielding a net zero direct component as expected from physical considerations.

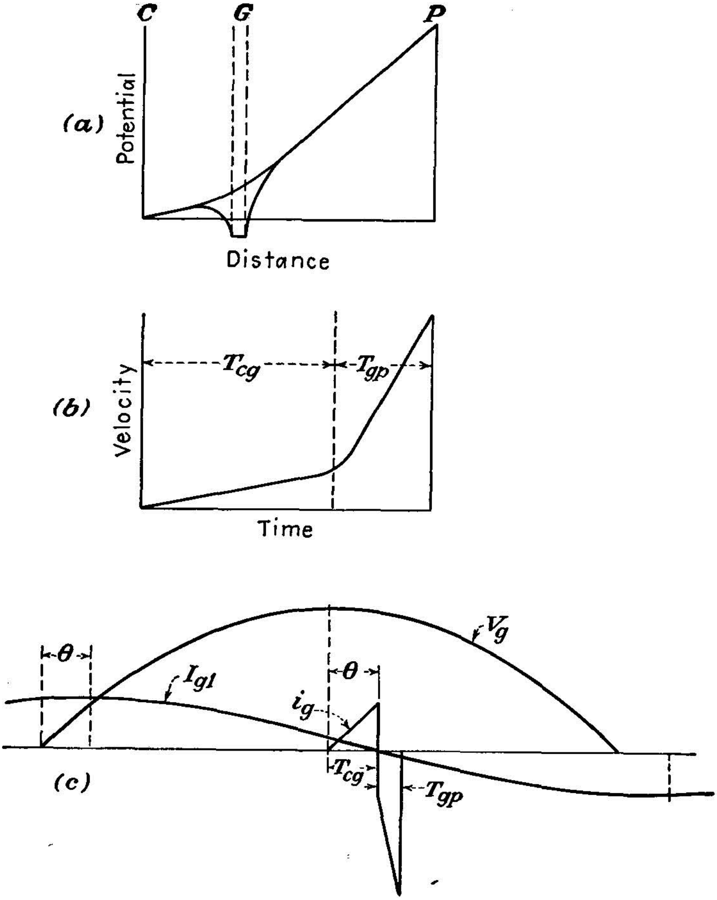


FIG. 16.9.—Factors determining induced grid current in a plane-electrode triode.

If it is assumed that there is one electron liberated per cycle at the same time after the maximum value of grid voltage, then the pulse of Fig. 16.9c will occur once each cycle and will have a fundamental component of current of the frequency of the exciting voltage. The fundamental component of current of the induced-current pulse will have the position shown in Fig. 16.9c. This fundamental component of current will change from positive to negative at about the same time as the induced-current pulse itself changes from positive to negative. As a result, the fundamental component of grid current will lead the grid voltage by 90 deg minus some small angle θ. The total grid current will be made up of the sum of all the induced currents resulting from the total electron flow. Since the electron current will be nearly sinusoidal and in phase with the grid voltage, the resultant fundamental component of grid current will have the same location as that shown for the single electron of Fig. 16.9*c*. This is because most electrons will flow at the peak of the grid voltage, and as a result any summation of pulses will favor those associated with the peak of the grid voltage.

The magnitude of the resulting fundamental component of grid current will be proportional to the product of the mutual conductance, the frequency, the electron transit time, and the grid voltage

$$I_{g1} = kG_m fTV_g \tag{16.21}$$

This occurs because the magnitude of the induced current depends upon the change in the number of electrons in the stream, which in turn depends upon the product of mutual conductance and voltage. The fundamental component of induced grid current depends upon the frequency, for the length of the current pulses induced by the individual electrons relative to the period of the exciting voltage is directly proportional to this factor, as will also be the area of the pulse. The fundamental component of the induced grid current will also depend upon the transit time of the electrons, for this will determine the area of the pulses of current induced by the passage of each electron.

The grid input admittance will be defined as the ratio of the grid current to the grid voltage

$$Y_g = \frac{I_{g1}}{V_g} = k_1 G_m fT \tag{16.22}$$

This admittance will have a conductance component and a susceptance component. If the grid current led the grid voltage by 90 deg, the input admittance would be purely imaginary, corresponding to the susceptance of the cathode-grid capacity, $j\omega C_{cg}$. Actually, this will be the larger component of the input admittance. However, the admittance will have a conductance component of the form

$$G_g = Y_g \sin \theta \tag{16.23}$$

where θ is the angle of Fig. 16.9*c* by which the fundamental component of the induced grid current fails to lead the grid voltage by 90 deg. For small angles, $\sin\theta$ can be replaced by θ. The angle θ itself depends upon the product of the frequency and transit time of the electron,

$$\sin\theta = \theta = k_2 fT \tag{16.24}$$

This is evident from Fig. 16.9*c*, it being remembered that the angle of a full period is 2π radians and that if the fundamental period is changed the angle θ will be changed even though the electron transit time is not changed. As a result of Eqs. (16.23) and (16.24), the input conductance is given approximately by

$$G_g = k_3 G_m f^2 T^2 \tag{16.25}$$

to a high degree of approximation.[1] Equation (16.25) shows that the grid input *conductance increases as the square of the frequency* for a given set of operating conditions. This is to say that the equivalent input *resistance* considered to be in parallel with the input capacity *decreases as the square of the frequency.* Some experimentally determined values are given in Fig. 16.10. The input resistance encountered here is such that the driving power required for a given degree of excitation increases as the square of the frequency. This rapidly becomes a limiting factor of considerable seriousness.

Although space-charge effects have been neglected in the above development, their presence will merely change the numerical constant. If the induced-current pulse of Fig. 16.9*c* had been drawn to include the effect of space charge, the positive part of the pulse would have had the form of the solid curve of Fig. 16.7*c* instead of the triangular form given. The shape of the negative portion of the pulse would not have been much changed. The constant of Eq. (16.25) can be evaluated to include the effect of space charge.[2] The specific form of the grid conductance is

$$G_g \cong \frac{4\pi^2}{180} G_m f^2 T_{cg}^2 \left[9 + 44\frac{T_{gp}}{T_{cg}} + 45\left(\frac{T_{gp}}{T_{cg}}\right)^2 - 2\frac{T_{gp}}{T_{cg}} \frac{\left(17 + 35\frac{T_{gp}}{T_{cg}}\right)}{1 + \frac{v_p}{v_g}} + \frac{20\left(\frac{T_{gp}}{T_{cg}}\right)^2}{\left(1 + \frac{v_p}{v_g}\right)^2} \right] \tag{16.26}$$

where T_{cg} is cathode-grid transit time, T_{gp} is grid-plate transit time, v_p is electron velocity at the plate, and v_g is mean electron velocity in the

---

[1] See FERRIS, W. R., Input Resistance of Vacuum Tubes as Ultra-high Frequency Amplifiers, *Proc. I.R.E.*, vol. 24, pp. 82–105, January, 1936, for an alternative derivation of Eq. (16.25).

[2] NORTH, *op. cit.*

grid plane. The numerical value of the constant given by the first term only of the expression in brackets is approximately 2; that is, k_3 in Eq. (16.25) is approximately 2 when T is the cathode-grid transit time.

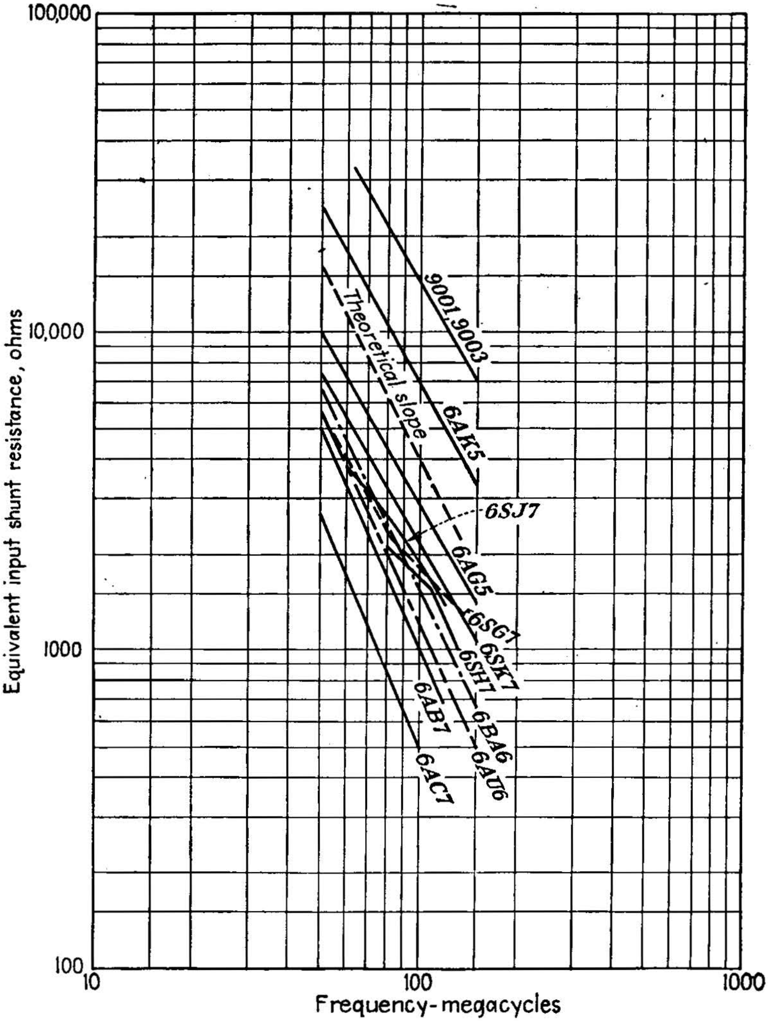


FIG. 16.10.—Input resistance of triodes as a function of frequency.

A number of factors conspire to prevent Eqs. (16.7) and (16.25) from being fulfilled exactly. The actual situation with respect to input conductance is extremely complicated.[1] As a result the above equations

---

[1] "Input Admittance of Receiving Tubes," Tube Department, Radio Corporation of America, Harrison, New Jersey, November, 1946.

indicate only first-order effects. Departures from the simple theory indicated above are due to the following:

1. The input capacity of a tube is nonlinear with transconductance. This is a low-frequency effect due to space charge. It contributes to a nonlinearity between input conductance and tube transconductance.
2. Partial resonance between lead inductance and interelectrode capacity may change apparent input capacity.
3. There may be a negative input-conductance component due to screen lead inductance in pentodes.
4. There are cold-tube input-conductance components due to lead resistance and dielectric losses that obscure lead-inductance and electron-transit-time effects. The lead resistance yields an input-conductance component that increases as the five-halves power of frequency as a result of skin effect and the series combination of resistance and inductance. Dielectric losses yield a component of conductance that increases linearly with frequency.

16.6. Transit-time Effects in the Space-charge-limited Diode. In the discussions thus far, relatively little attention has been paid to the effects of space charge. The effect of space charge may be expected to be considerable, particularly in the vicinity of the cathode, where the space-charge density is very high. Before going into this subject it will be well to emphasize the distinction between the various components of current encountered.

The general form of current involves a combination of conduction current and displacement current.

$$J = \rho v + \varepsilon_0 \frac{\partial E}{\partial t} \qquad (16.27)$$

The first term here is the conduction current and is proportional to the number of electrons arriving per second at any reference plane. The second component of current is the displacement current. This is the current that flows as a result of changes in the electric-field strength. In vacuum-tube problems the resultant current will ordinarily be a combination of conduction and displacement current. At low frequencies the current will be nearly all conduction current, but at sufficiently high frequencies the displacement current will be considerable. This occurs because of the finite transit time required by the electrons to pass from one point to another. Thus, if a group of electrons is liberated at a cathode of a diode, it will be a while before they arrive at the plate. This does not mean that the plate current is zero until the electrons arrive. True, the conduction current will be zero until the electrons arrive, but in the meantime there will be current in the form of displacement current (the induced current of the previous sections). The total current in the general case is the sum of the conduction current and the displacement current and *is the same at any point in the circuit.* Thus in the diode the current at the cathode is virtually all conduction current because the field there is zero. At the plate, in the presence of an alternating voltage, the total current will be the sum of the conduction current at the plate and the displacement current associated with the changing electric field resulting from electrons en route to the plate.

To examine the relations in the plane-electrode diode with space charge it is necessary to know the equation of motion of the electron in addition to the general definition of current flow. The equation of motion is simply

$$-eE = e\frac{dV}{dx} = ma \tag{16.28}$$

where e is the magnitude of the electron charge, $\frac{dV}{dx}$ is the gradient of potential, m is the mass of the electron, and a is its acceleration. In addition, Poisson's equation will be involved to take account of the effect of space charge upon the potential distribution. For the plane-electrode case with the various quantities varying in the x direction only,

$$\operatorname{div} \varepsilon_0 E = \varepsilon_0 \frac{\partial E}{\partial x} = \rho \tag{16.29}$$

where ρ is the space-charge density and ε_0 is the dielectric constant of free space. Combining Eqs. (16.27) and (16.29) gives

$$J = \varepsilon_0\left(\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial t}\right) = \varepsilon_0\frac{dE}{dt} \tag{16.30}$$

Referring back to Eq. (16.28), it is now apparent that

$$J = -\varepsilon_0\frac{m}{e}\frac{da}{dt} \tag{16.31}$$

The previous five equations are the fundamental ones upon which all electron-transit-time studies involving space charge are based. Equation (16.27) is essentially Maxwell's definition of current in its general form. Here it is necessary only to remember that current in general may be either displacement or conduction or a combination of both. Equations (16.28) and (16.29) are relatively well known and deserve no particular comment. Equations (16.30) and (16.31) are the new relations of significance. These give current as a function of time only, directly proportional to the time rate of change of electric field or of acceleration. Since electric field is a function of both time and distance, the total derivative with respect to time has involved partial derivatives with respect to each. Fortunately, the combination of partial derivatives given in Eq. (16.30) is exactly the total derivative of the electric field. It is the last of the above equations that is really new and significant. From this it is seen that if the acceleration or for that matter any of its derivatives be known then the current can be determined. From this equation all the dynamic properties of space-charge flow can be determined.

Let us test the power of Eq. (16.31) by obtaining some basic relations. Let it be assumed that the current density is made up of a constant component plus an alternating component of the form.

$$J = J_0 + J_1\epsilon^{pt} \tag{16.32}$$

where J_0 is a direct component of current density and J_1 is the magnitude of an alternating component, p being equal to $j\omega$, and it is understood that we are dealing with only the real part of the exponential factor ϵ^{pt}. This is a well-known procedure in network theory, and it is used here because it simplifies the writing of the associated equations.

The differential equation corresponding to Eq. (16.31) becomes

$$\frac{da}{dt} = \frac{-e}{m\varepsilon_0}(J_0 + J_1\epsilon^{pt}) \tag{16.33}$$

Let this now be integrated to obtain the acceleration, velocity, and distance in a parallel-plane diode under the assumption that the initial velocity and acceleration of the electrons are zero. With these restrictions, a first integration of Eq. (16.33) gives

$$a = \frac{-e}{m\varepsilon_0}\left[J_0(t - t_a) + \frac{J_1}{p}(\epsilon^{pt} - \epsilon^{pt_a})\right] \tag{16.34}$$

where t_a is the time when the electron leaves the cathode. A second integration gives

$$v = \frac{-e}{m\varepsilon_0}\left[\frac{J_0}{2}(t - t_a)^2 + \frac{J_1}{p^2}(\epsilon^{pt} - \epsilon^{pt_a}) - \frac{J_1}{p}(t - t_a)\epsilon^{pt_a}\right] \tag{16.35}$$

A third integration gives

$$x = \frac{-e}{m\varepsilon_0}\left[\frac{J_0}{6}(t - t_a)^3 + \frac{J_1}{p^3}(\epsilon^{pt} - \epsilon^{pt_a}) - \frac{J_1}{p^2}(t - t_a)\epsilon^{pt_a} - \frac{J_1}{2p}(t - t_a)^2\epsilon^{pt_a}\right] \tag{16.36}$$

These equations give acceleration, velocity, and distance as a function of the time, the starting time, and the current-density components. Let the validity of these equations be tested by examining the direct components. If the above equations are restricted to the case in which the alternating component of current density J_1 is zero and if $t - t_a$ be considered the transit time t_0, then the above equations reduce to

$$a_0 = \frac{-eJ_0t_0}{m\varepsilon_0} \tag{16.37}$$

$$v_0 = \frac{-eJ_0t_0^2}{2m\varepsilon_0} \tag{16.38}$$

$$x_0 = \frac{-eJ_0t_0^3}{6m\varepsilon_0} \tag{16.39}$$

Of these, the last equation will be recognized as giving the proper variation of distance with time. To bring the above equations into a more familiar form it is desirable to obtain an expression of the voltage difference corresponding to the distance x_0. This is readily obtained from the definition

$$V_0 = -\int_0^{x_0} E_0\,dx \tag{16.40}$$

which by virtue of the equation of motion (16.28) is the same as

$$V_0 = \frac{m}{e}\int_0^{t_0} a_0v_0\,dt \tag{16.41}$$

This yields

$$V_0 = \frac{m}{e}\frac{v_0^2}{2} \tag{16.42}$$

to our small surprise. If now the expression for v_0 from (Eq. 16.38) be substituted in this and the value of t_0 as determined from Eq. (16.39) be applied, there results

$$V_0^{3/2} = \frac{9}{4}\sqrt{\frac{m}{2e}}\,\frac{x_0^2J_0}{\varepsilon_0} \tag{16.43}$$

which is Child's law as previously given by Eq. (8.7). Apparently Eqs. (16.33) to (16.35) can be trusted to give some reliable answers if properly interpreted.

In the same way as the direct current was found as a function of the direct voltage, the alternating component of current can be found as a function of the alternating component of voltage. In this case the electron transit time is expected to be involved, and it is. When the voltage and current are known, their ratio gives the equivalent impedance, a factor of great importance in tube application problems. The derivation of the impedance of a diode whose emission is space-charge-limited has been given many times and is much too lengthy to be included in the text.[1-3] The specific formula for the diode impedance is

$$\frac{Z}{R_0} = \frac{2}{\beta} + \frac{12}{\beta^4}\,(2 - 2\epsilon^{-\beta} - \beta - \beta\epsilon^{-\beta}) \tag{16.44}$$

where $\beta = i\theta$, θ being the transit angle from cathode to plate, that is, $\theta = 2\pi fT$, where T is the transit time. R_0 is the low-frequency dynamic plate resistance of the diode as determined by the slope of the voltage-current characteristic. This expression separates readily into real and imaginary parts corresponding to series resistance and reactance components.

$$\frac{R}{R_0} = \frac{12}{\theta^4}\,[2(1 - \cos\theta) - \theta\sin\theta] \tag{16.45}$$

$$\frac{X}{R_0} = -\frac{2}{\theta} - \frac{12}{\theta^4}\,[\theta(1 + \cos\theta) - 2\sin\theta] \tag{16.46}$$

Curves of $\frac{R}{R_0}$ and $\frac{X}{R_0}$ are given in Fig. 16.11 as a function of the transit angle θ. These components are part of the series representation of impedance and indicate that the diode impedance is equivalent to a resistance in series with a capacitive reactance, X being always negative. Also shown as a dashed curve in Fig. 16.11 is the high-frequency asymptote of the reactance curve. This has the form of the reactance curve of a pure capacity. The resistance component drops from a maximum value for zero transit angle to a zero value for a transit angle of 2π. After that, it assumes alternately negative and positive values but never exceeds a few per cent of the maximum value in magnitude. It is interesting to note that the resistance changes from positive to negative at transit angles of 2π, 4π, 6π, etc., whereas the change from negative to positive resistance occurs for transit angles of 3π, 5π, 7π, etc. This means that the region in which the diode resistance is negative is smaller than the region in which the diode resistance is positive. The negative resistance predicted by the form of Fig. 16.11 for transit angles between 2π and 3π is quite real, and special diodes have been made to oscillate by virtue of

---

[1] BENHAM, W. E., A Contribution to Tube and Amplifier Theory, *Proc. I.R.E.*, vol. 26, pp. 1093–1170, September, 1938. This article summarizes work in earlier British publications.

[2] LLEWELLYN, F. B., "Electron Inertia Effects," Cambridge, London, 1941 (distributed in the United States by Macmillan). This tract summarizes the work covered in Llewellyn's numerous papers prior to 1941.

[3] MÜLLER, J., Eletronenschwingungen im Hochvakuum, *Hochfrequenz. und Elektroakustic*, vol. 41, pp. 156–167, May, 1933.

this negative resistance.[1] The finite velocity of emission of the electrons tends to reduce the magnitude of the negative resistance predicted by Eq. (16.45).

The nature of the reactive component of diode impedance is best understood by examining the imaginary component of the reciprocal of impedance, *i.e.*, the admittance. A plot of the real (conductance) and imaginary (susceptance) components of admittance of a plane-electrode diode whose emission is space-charge-limited is shown in Fig. 16.12. From this it is seen that the susceptance of the diode is closely represented by that of a capacity in shunt with a resistance for small transit angles. At low frequencies or small transit angles, the capacitive susceptance ratio is given approximately by $\frac{3\theta}{10}$. The proportionality with

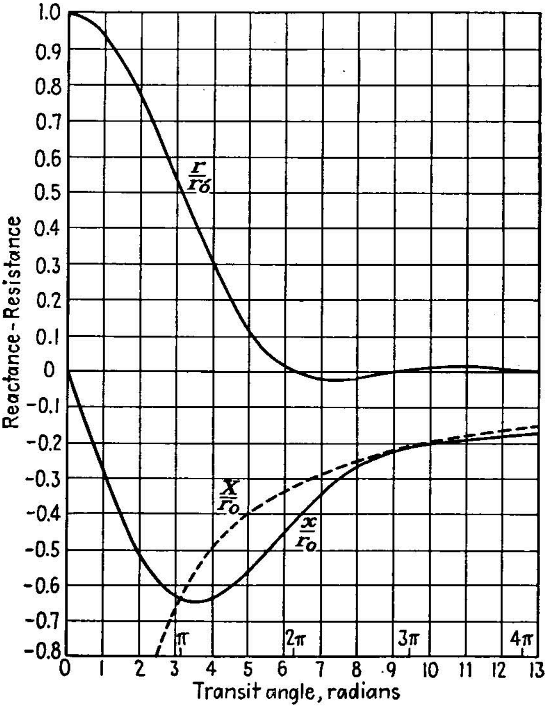


FIG. 16.11.—Components of the equivalent series impedance of a plane diode whose emission is space-charge-limited. (*After Llewellyn.*)

---

[1] LLEWELLYN, F. B., and A. E. BOWEN, The Production of Ultra-high Frequency Oscillations by Means of Diodes, *Bell Sys. Tech. Jour.*, vol. 18, pp. 280–291, April, 1939.

transit angle and frequency means that the susceptance can be represented by a fixed capacity. The size of this capacity happens to be $\frac{3}{5}$ of the cold capacity of the tube. This amounts to saying that the electron charge acts like a dielectric with a dielectric constant of $\frac{3}{5}$. For higher values of frequency and transit angle the susceptance departs from the low-frequency value and finally becomes asymptotic to the value corresponding to the cold capacity.

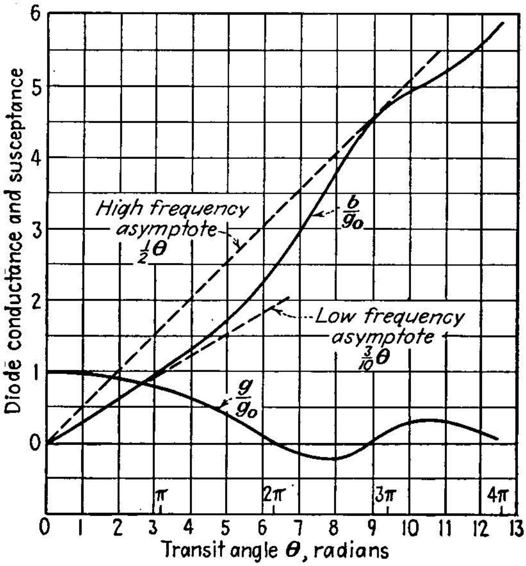


FIG. 16.12.—Components of the equivalent shunt admittance of a plane diode whose emission is space-charge-limited. (*After Llewellyn.*)

It is possible to work out equivalent circuits for the diode admittance over a large range of transit angles. For low frequencies the parallel combination of a resistance equal to the plate resistance in parallel with a capacity equal to $\frac{3}{5}$ of the cold capacity works very well. For frequencies giving rise to transit angles greater than 90 deg it is best to refer to the curves of Figs. 16.11 and 16.12.

16.7. Small-signal Transit-time Effects in the Space-charge-limited Triode. Much of the information obtained in the previous section can be applied to the case of a triode operating with small signal voltages and with its emission space-charge-limited. Here it is expected that there will be something like a diode action between the cathode and grid. This will influence the input impedance of the tube. Further, it is expected that the tube capacities will play an important role. In addition, it is to be expected that the mutual conductance of the tube will be changed by transit-time effects. Various equivalent circuits have been proposed for triodes operating under the above conditions, one of the most successful being that shown in Fig. 16.13.[1] This is a T section of admittances with an internal generator in the plate lead to represent the effect of the voltage applied in the grid circuit. The junction of the admittances occurs, not on any of the electrodes, but in the grid plane between the grid wires. The admittance Y_{11} is the admittance between the cathode and the grid plane and is the same as that given by Fig. 16.12 for a plane-electrode diode. The admittance Y_{22} is simply the admittance of the plate–grid-plane capacity. The admittance Y_g is the capacity from the grid wires to the grid plane and is mu times as big as the plate–grid-plane

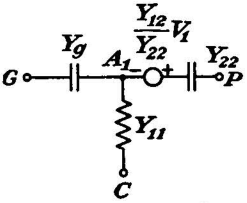


FIG. 16.13.—Equivalent circuit of a triode operating at ultra-high frequencies.

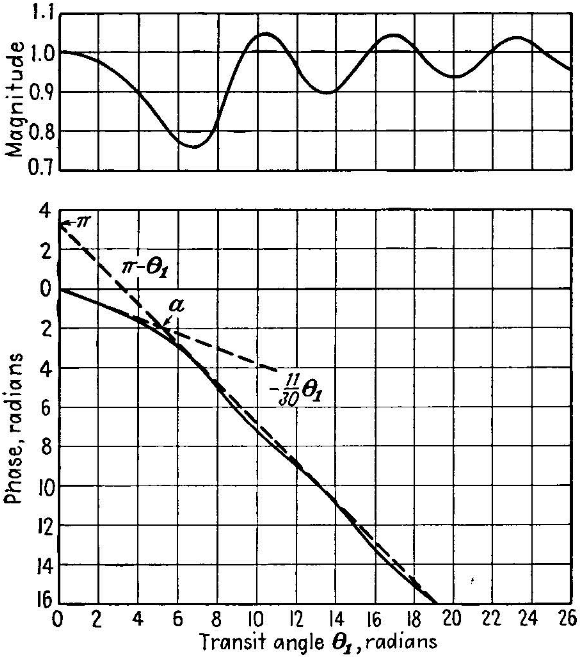


FIG. 16.14.—Small-signal transadmittance of a triode as a function of cathode-grid transit angle. (*After Llewellyn.*)

---

[1] LLEWELLYN, F. B., and L. C. PETERSON, Vacuum Tube Networks, *Proc. I.R.E.*, vol. 32, pp. 144–166, March, 1944.

capacity. The transadmittance Y_{12} replaces the G_m used at low frequencies. The transadmittance can be evaluated by an extension of the arguments used to obtain the diode admittance. On the assumption that the grid-plate transit time is small compared with the cathode-grid transit time the transadmittance is given by the curves of Fig. 16.14.[1] The magnitude of the transadmittance is seen to fluctuate with transit angle, but not excessively. The magnitude never differs from the low-frequency value by more than 25 per cent. This is apparently due to something like the bunching action that occurs in klystrons in the presence of space charge where there is a periodic variation of the effective bunching parameter. The phase of the transadmittance, however, is continuously retarded with transit angle. The first minimum of the magnitude occurs for a transit time of approximately one cycle in the cathode-grid region.

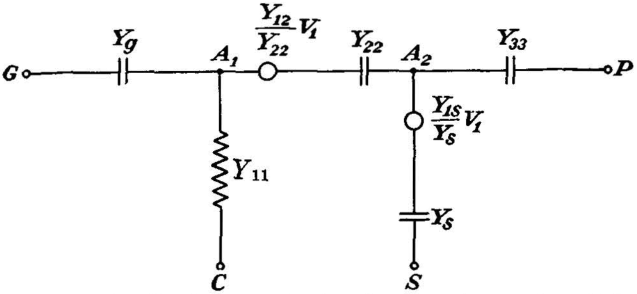


FIG. 16.15.—Equivalent circuit of a tetrode at ultra-high frequencies.

With all the elements of Fig. 16.13 given it is a relatively simple matter to compute the performance of the tube under any conditions. Thus the predictions of Sec. 16.5 on the input impedance of a triode may be verified by inspection. The input impedance of the circuit of Fig. 16.13 is essentially that of the grid-plane capacity in series with the cathode–grid-plane diode impedance. At low frequencies this acts like a capacity in series with a resistance. This is readily shown to be the same as the impedance of a capacity paralleled by a resistance whose magnitude varies inversely as the square of the frequency.

Multielectrode tubes can be treated by an extension of the ideas applied above to the triode. Here it is merely necessary to add another L section for each additional grid to the circuit of Fig. 16.13. Thus the equivalent circuit of a tetrode is as given in Fig. 16.15.[1] Here the first

---

[1] *Ibid*

branch point A_1 is located in the control-grid plane between wires. The second branch point A_2 represents the screen-grid plane between wires. The admittances Y_g, Y_{22}, and Y_{33} are the admittances of the simple capacities between adjacent electrodes. Y_s is the admittance of the screen-grid–screen-grid-plane capacity, which is larger than the screen-grid–plate capacity by the screen-grid mu. Y_{1s} is the transadmittance of the first-grid relative to the second-grid current. Y_{12} is the transadmittance of the plate current relative to the grid potential.

16.8. Similitude and Scaling in Ultra-high-frequency Tubes. It is frequently of interest to consider the effect of changing the size of tubes or of operating given tubes at a different voltage or frequency. A study of such changes is well worth while, for it lays a basis for design and also aids greatly in the understanding of the operation of tubes at ultra-high frequencies.

It is recognized that there is a relation between voltage, frequency, and the distance that an electron must travel in a given length of time. A basic relation between these factors can be obtained from the equation of motion of an electron subjected to an electric field.

$$-eE = ma \tag{16.28}$$

Dimensionally, this is of the form

$$\frac{e}{m} = \frac{d^2}{Vt^2} = \frac{d^2f^2}{V} \tag{16.47}$$

Since $\frac{e}{m}$ is a numerical constant, it follows that $\frac{d^2f^2}{V}$ is also a numerical constant. Essentially, this makes the combination $\frac{m}{e}\frac{d^2f^2}{V}$ a dimensionless parameter that applies to the problems of motion. Thus, as long as the factor $\frac{d^2f^2}{V}$ is constant, no matter what the value of the individual factors it will always be true that an electron will require the same fraction of a cycle to travel corresponding distances. This same conclusion is arrived at by considering transit angle as being equal to $2\pi fT$, where T is the transit time. Since the transit time is proportional to the ratio of distance to velocity or the square root of voltage, transit angle is proportional to the factor $\frac{fd}{V^{1/2}}$, which is simply the square root of the dimensionless factor given above. Hence to get tubes that will have the same impedance at any given frequency, if tube a is twice as big as tube b it must operate at four times the voltage of tube b. Likewise, to keep transit time constant, it is necessary to build tubes smaller in inverse proportion to frequency for operation at a given voltage or operate at voltages that are higher in proportion to the square of frequency for a given size of tube.

Other factors besides the transit angle are involved in high-frequency operation. If the size of tubes is changed, then power-dissipation capacities are changed. So likewise are the actual inductances and capacities of the tube. All these factors may be studied by setting up some scaling factors with respect to the basic equations that determine ultra-high-frequency operation. These equations are two, the equation of motion of an electron and Poisson's equation. Let there be considered two tubes whose dimensions are in the ratio of D operating at wave lengths in the ratio of W. Thus the defining relations for D and W are

$$D = \frac{x_2}{x_1} \tag{16.48}$$

and

$$W = \frac{\lambda_2}{\lambda_1} \tag{16.49}$$

If now an electron moves between corresponding points of two similar tubes in the same fraction of a cycle,

$$dt_2 = W\,dt_1 \tag{16.50}$$

The equation of motion for an electron in the second tube is

$$m\frac{d^2x_2}{dt_2{}^2} = -eE_{x_2} \tag{16.51}$$

The corresponding equation of motion in terms of an electron in the first tube is

$$\frac{mD\,d^2x_1}{W^2dt_1{}^2} = -eE_{x_2} \tag{16.52}$$

For these two equations to yield similar paths with the same dependence upon transit angle at the respective frequencies it is necessary that

$$\frac{E_2}{E_1} = \frac{D}{W^2} \tag{16.53}$$

Referring now to Poisson's equation in the form of Eq. (16.31),

$$\frac{J_2}{J_1} = \frac{\dfrac{dE_2}{dt_2}}{\dfrac{dE_1}{dt_1}} = \frac{D}{W^3} \tag{16.54}$$

In like manner the ratios of all the critical quantities may be obtained in terms of the factors D and W. The resulting relations are summarized in Table XI below.

TABLE XI
SCALING FACTORS FOR ULTRA-HIGH-FREQUENCY TUBES*

Quantity	Ratio	General	Complete scaling $W = D$	Voltage scaling $W = 1$	Wave-length scaling $D = 1$
Inductance	$\frac{L_2}{L_1}$	$\frac{D}{W^0}$	W	D	1
Capacity	$\frac{C_2}{C_1}$	$\frac{D}{W^0}$	W	D	1
Field	$\frac{E_2}{E_1}$	$\frac{D}{W^2}$	$\frac{1}{W}$	D	$\frac{1}{W^2}$
Voltage	$\frac{V_2}{V_1}$	$\frac{D^2}{W^2}$	1	D^2	$\frac{1}{W^2}$
Current density	$\frac{J_2}{J_1}$	$\frac{D}{W^3}$	$\frac{1}{W^2}$	D	$\frac{1}{W^3}$
Current	$\frac{I_2}{I_1}$	$\frac{D^3}{W^3}$	1	D^3	$\frac{1}{W^3}$
Power	$\frac{P_2}{P_1}$	$\frac{D^5}{W^5}$	1	D^5	$\frac{1}{W^5}$
Power density	$\frac{h_2}{h_1}$	$\frac{D^3}{W^5}$	$\frac{1}{W^2}$	D^3	$\frac{1}{W^5}$
Conductance	$\frac{G_2}{G_1}$	$\frac{D}{W}$	1	D	$\frac{1}{W}$

* $D = \frac{x_2}{x_1}$, $W = \frac{\lambda_2}{\lambda_1}$.

It is interesting to note that this same table applies for magnetron tubes, it being necessary only to add a row for the ratio of magnetic-flux densities. The ratios in the column entitled "General" apply for similar tubes operating at different frequencies but with electrons moving between corresponding points in the tubes in the same fraction of a cycle. The ratios in the column entitled "Complete scaling" apply for similar tubes with dimensions proportional to wave length, the usual case. If a tube is simply changed in size and the voltage adjusted accordingly but operation is had on the same frequency, then the values in the "Voltage scaling" column apply. If dimensions are not changed but wave length and voltage are changed to get the same electronic action, then the values in the "Wave-length scaling" column apply.

In the case of complete scaling, increased power output is actually obtained up to the point where one of the requirements indicated by the table is violated. This will usually be either the current-density requirement, which increases as the square of the frequency, or the power dissipation per unit area requirement, which increases as the square of the frequency.

In the case of voltage scaling an excellent gain in performance characteristics is achieved. Tubes scaled on this basis will usually be limited either by voltage breakdown or power dissipation. Note that the power output goes up as the fifth power of the size. The required voltage goes up as the square of the size. Inductance and capacity go up linearly with size but will usually be the same percentage of the associated external values.

Wave-length scaling amounts to operating a given tube at a variable frequency but changing the voltage to compensate for transit-time effects. This requires that the voltage be increased as the square of the frequency. This requires emission-current density that increases as the cube of the frequency and power dissipation per unit area that increases as the fifth power of the frequency. Ordinarily, one cannot go very far in this direction.

16.9. High-frequency Limit of Triode Oscillation. The operation of power oscillator tubes at ultra-high frequencies is considerably more complicated than that of receiving tubes. The increase in complexity results from the fact that the alternating voltages are usually large and therefore current will flow for only part of a cycle. Electrons flowing at different times during the cycle will have widely different behavior as far as transit times are concerned. The general treatment of large signal effects will be left for the next section, and this section will be devoted to some observations that can be made in limiting cases.

It is well known that transmitting tubes whether operating as oscillators or as amplifiers suffer from a loss of output as the frequency is raised. Figure 16.16 gives some curves showing the power output of a number of different oscillator tubes as a function of frequency. All these curves have the same general shape. At low frequencies the output is constant. As frequency is raised, the power drops off, slowly at first, and then very rapidly. Usually the power output will have dropped to zero within a factor of 10 of the frequency at which a decrease in output is first detectable. Of considerable importance is the observation that there is a power-frequency limit for tubes of the same type.[1] This is evident in Fig. 16.16.

Although the curves for different tubes overlap, there is an envelope that can be drawn to the family of curves as a whole. The basic trend

---

[1] Wagener, W. G., The Developmental Problems and Operating Characteristics of Two New Ultra-high Frequency Triodes, *Proc. I.R.E.*, vol. 26, pp. 401–414, April 1938.

is that tubes designed to produce high power are not able to go to as high frequency as tubes designed for a lower power output. A better statement of this effect is that as tubes are designed to operate at higher and

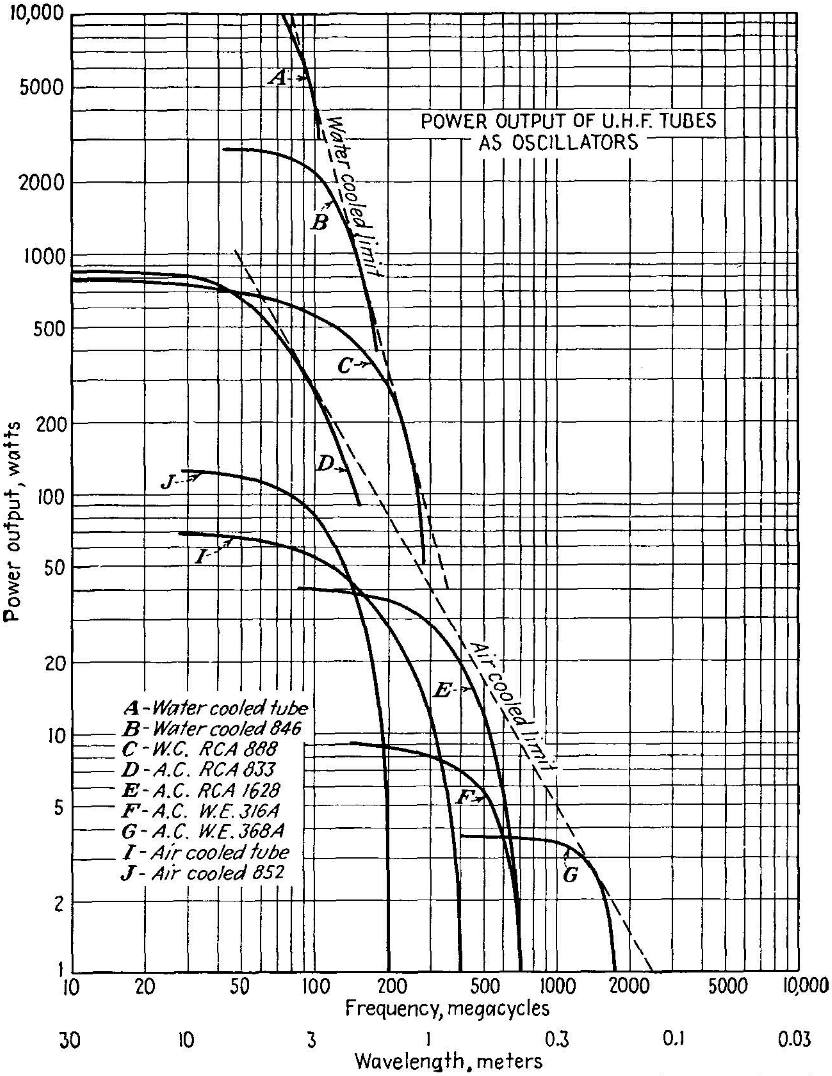


FIG. 16.16.—Power output of ultra-high-frequency oscillator tubes as a function of frequency.

higher frequencies their output is inherently reduced. This is in accord with the observations made in connection with the scaling values of Table XI. The envelope for the water-cooled tubes is approximately a straight line with a slope of −4. This is the proper limit for tubes whose output is limited by a given power dissipation per unit area. From the Complete scaling column in Table XI it is seen that the power density varies as the square of the frequency. Since the actual allowable dissipation is fixed by the cooling system, the power must be decreased inversely as the square of the frequency to keep the dissipation per unit area constant. In addition, the area varies inversely as the square of the frequency, and as a result the power output obtainable with a watercooled tube of optimum design operating at a given fraction of its high-frequency limit is expected to vary as the inverse fourth power of frequency.

The air-cooled tubes have a high-frequency limit that varies approximately as the inverse square of frequency. This is not greatly different from the relation that is expected from complete scaling when the cathode emission is the limiting factor.

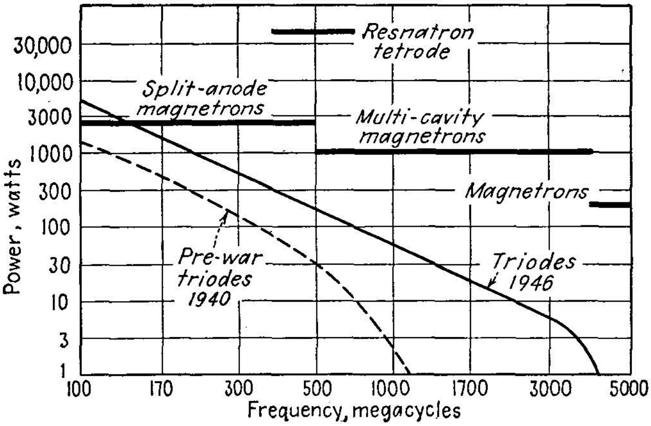


FIG. 16.17.—Power output of continuous-wave oscillators as a function of frequency.

Recent developments in tubes have pushed the high-frequency envelope appreciably to the right. In Fig. 16.17 is shown the power output as a function of frequency of various types of continuous-wave tubes as of early 1946.[1] Undoubtedly, further advances will push these limits still farther to the right, but the big gains in this direction will come from the development of new types of operation rather than from a refinement of conventional tubes.

Not too much is known about the operation of tubes over the complete range of frequencies from a low-frequency range of constant output to a high-frequency limit of extinction. At low frequencies where the

---

[1] BYRNE, JOHN, Power Limits of Continuous Wave Tubes, *Electronics*, vol. 19, p. 91, January, 1946.

transit time of the electrons is negligible the performance is well understood in terms of Class C amplifier theory. As frequency is raised, some phase shifts are encountered as a result of the finite transit time of the electrons and the performance can still be estimated. This takes one out to frequencies where the output has dropped off to about 80 per cent of the low-frequency value. As frequency is increased still further, there are pronounced phase-shift and transit-time effects associated with large alternating voltages and the resulting operation is at best poorly understood. This covers the frequencies from about 80 per cent to about 5 per cent of the low-frequency output. When the output has dropped to about 5 per cent of the low-frequency output, the alternating voltages will be quite small and the small-signal theory of Llewellyn, Benham, and Mueller will apply.

Frequency at Which Efficiency Begins to Fall Off. It is of interest to identify some reference points on the curves of Fig. 16.16. One such reference point is the frequency at which the output has dropped to some given percentage of the low-frequency value, say 90 per cent. This can be done fairly satisfactorily by the application of some simplifying assumptions.[1] Let it be assumed that the oscillator is operating Class C and that the plate-current pulse is a rectangular one which flows for a quarter of a cycle. Thus let $i_p = i_1$ for $-\frac{\pi}{4} < \theta < \frac{\pi}{4}$ and $i_p = 0$ for other angles of the cycle where $\theta = \omega t$. Let the corresponding plate voltage be

$$V_p(t) = V_{p0} - V_{p1} \cos \theta \tag{16.55}$$

The plate power loss for these assumed conditions is as shown in Fig. 16.18.

$$W_{p1} = \frac{1}{2\pi} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} i_1(V_{p0} - V_{p1} \cos \theta)\, d\theta \tag{16.56}$$

$$W_{p1} = \frac{i_1 V_{p0}}{4} - \frac{i_1 V_{p1}}{\sqrt{2}\,\pi} \tag{16.57}$$

Let it now be assumed that electron-transit-time effects set in as a result of an increase in frequency and that the only effect is to cause the plate-current pulse to lag the alternating plate voltage by the angle ωT, where T is the cathode-plate transit time. The plate loss under these conditions corresponding to the dotted curves of Fig. 16.18 will be

---

[1] GAVIN, M. R., Triode Oscillators for Ultra-short Wave Lengths, *Wireless Eng.*, vol. 16, pp. 287–296, June, 1939.

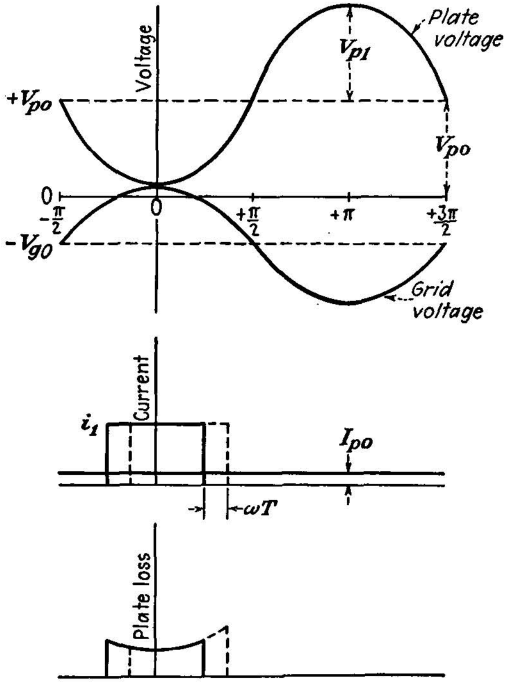


FIG. 16.18.—Effect of transit time upon the plate loss of a Class C oscillator.

$$W_{p2} = \frac{1}{2\pi} \int_{-\frac{\pi}{4}+\omega T}^{\frac{\pi}{4}+\omega T} i_1(V_{p0} - V_{p1} \cos \theta)\, d\theta \tag{16.58}$$

$$W_{p2} = \frac{i_1 V_{p0}}{4} - \frac{i_1 V_{p1}}{\pi \sqrt{2}} \cos \omega T \tag{16.59}$$

or for small transit angles

$$W_{p2} = \frac{i_1 V_{p0}}{4} - \frac{i_1 V_{p1}}{\pi \sqrt{2}} + i_1 \frac{V_{p1}}{2\pi \sqrt{2}} \omega^2 T^2 \tag{16.60}$$

The efficiencies for the two cases cited above are

$$\eta_1 = \frac{W_0 - W_{p1}}{W_0} \tag{16.61}$$

and

$$\eta_2 = \frac{W_0 - W_{p2}}{W_0} \tag{16.62}$$

where W_0 is the input power and W_{p1} and W_{p2} are the plate losses for the two cases. Accordingly, the difference in efficiency is

$$\eta_1 - \eta_2 = \frac{i_1 V_{p1} \omega^2 T^2}{2\pi \sqrt{2}\, W_0} = \frac{\sqrt{2}}{\pi} \frac{V_{p1}}{V_{p0}} \omega^2 T^2 \tag{16.63}$$

which indicates that the decrease in efficiency is proportional to the square of the frequency, to the square of the transit time, and to the ratio of peak alternating to direct plate voltage. A further assumption that is reasonable is that the ratio of peak alternating to direct plate voltage is 0.9. This corresponds closely to the operating condition for maximum efficiency over a wide range of conditions of voltage, load, and tube selection. With this assumption Eq. (16.63) reduces simply to

$$\eta_1 - \eta_2 \cong 0.4 \omega^2 T^2 \tag{16.64}$$

Let now some voltages be assumed so that the transit time T can be determined. A representative operating condition is that

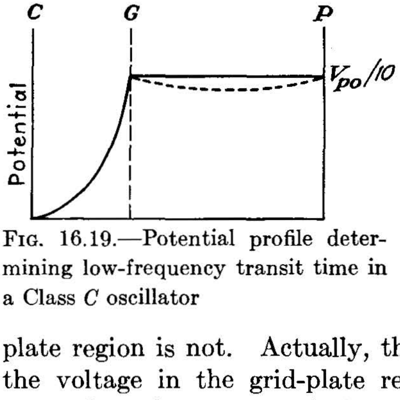


FIG. 16.19.—Potential profile determining low-frequency transit time in a Class C oscillator

$$V_{g\max} = V_{p\min} = \frac{V_{p0}}{10} \tag{16.65}$$

This means that the transit time will be determined for a potential profile like that shown in Fig. 16.19. Here it is assumed that the current flow in the cathode-grid region is space-charge-limited while that in the grid-plate region is not. Actually, the presence of space charge will depress the voltage in the grid-plate region slightly as shown by the dotted curve, but the error made in assuming that there is no space charge present in this region will not be great. The cathode-plate transit time for this condition is

$$T_{cp} \cong \frac{3d_{cg} + d_{gp}}{5.93 \times 10^7 \sqrt{\dfrac{V_{p0}}{10}}} \quad \text{sec} \tag{16.66}$$

where d_{cg} and d_{gp} are cathode-grid and grid-plate distances in centimeters, respectively, and V_{p0} is the direct plate voltage in volts.

If now it is desired to determine the wave length at which the efficiency has dropped 10 per cent from the low-frequency value, then $\eta_2 - \eta_1$ is set equal to 0.1, as a result of which

$$\lambda_2 = 4\pi c T_{cp} \tag{16.67}$$

where $c = 3 \times 10^{10}$ cm per sec is the velocity of light. When now the value of T_{cp} from Eq. (16.66) is substituted, it is found that

$$\lambda_2 \cong \frac{20{,}200(3d_{cg} + d_{gp})}{\sqrt{V_{p0}}} \quad \text{cm} \qquad (16.68)$$

This is the relation that has been sought. It gives the wave length at which the efficiency of a Class C oscillator will have dropped 10 per cent from its low-frequency value. The assumptions made were that the plate current was a rectangular pulse of a quarter-cycle duration which was shifted in phase but not changed in shape by electron-transit-time effects and that the grid and plate voltage at the peak of the alternating grid voltage were each one-tenth of the direct plate voltage. While these assumptions are somewhat rough, the answer depends upon the difference between the two quantities arrived at by making the same assumptions and so the errors involved tend to cancel. The largest error probably lies in the assumption that the shape of the plate current does not change. The formula is probably accurate only within 10 per cent but is still useful in estimating ultra-high-frequency behavior. Inspection of Eq. (16.68) shows that the lower wave-length limit of tubes may be extended by reducing the interelectrode spacings, with the cathode-grid distance more critical than the grid-plate distance, or by increasing the plate voltage.

Frequency at Which Oscillation Ceases. Another reference point on the curves of Fig. 16.16 is the frequency at which the tube ceases to oscillate. This is determined by circuit as well as electron-transit-time considerations, but with proper design of tubes it is always the electron-transit-time effects that finally dominate in reducing the output. It may therefore be expected that whatever mechanism reduces the tube output is some function of the total transit time from cathode to plate. If this can be specified in terms of operating conditions at the limiting frequency, then the extinction frequency can be related to the cathode-plate transit time by experimental observations.[1]

Since most oscillators derive their grid bias from a resistor in the grid circuit, it is expected that as frequency is raised and the oscillations become weaker until finally they cease, the grid-bias voltage will be reduced until at the extinction frequency it has become zero. Under this condition the potential profiles along which the electrons must move will be as shown in Fig. 16.20. The electrons will prefer to move between the grid wires, taking the path that has the most positive potential. For the case under discussion it will be assumed that the current flow in the cathode-grid region is space-charge-limited while that in the grid-

---

[1] *Ibid.*

plate region is not. This, as we shall see later by a comparison of transit-time formulas, is a reasonable approximation to the true condition. As a result of these assumptions the potential will vary as the four-thirds power of the distance from the cathode in the cathode-grid region and linearly in the grid-plate region.

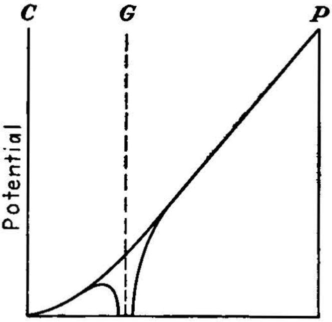


FIG. 16.20.—Potential profile determining the high-frequency limit of triode oscillation.

At the extinction frequency, with the grid electrode at zero voltage, the average potential of the grid plane will be $\frac{V_p}{\mu}$, where V_p is the plate potential and μ is the amplification factor of the tube. With this information the transit times can now be calculated. The transit time in the cathode-grid region will be

$$T_{cg} = \frac{3d_{cg}}{5.93 \times 10^7 \sqrt{\frac{V_p}{\mu}}} \quad \text{sec} \tag{16.69}$$

since the transit time in the absence of space charge is the distance divided by the average velocity and with full space charge is 50 per cent greater. The transit time for the grid-plate region is

$$T_{gp} = \frac{2d_{gp}\sqrt{\mu}}{5.93 \times 10^7 \sqrt{V_p}\,(\sqrt{\mu} + 1)} \tag{16.70}$$

since the transit time is the distance divided by the average velocity. Adding the results of Eqs. (16.69) and (16.70), there is obtained

$$T_{cp} = \frac{1}{5.93 \times 10^7} \sqrt{\frac{\mu}{V_p}} \left(3d_{cg} + \frac{2d_{gp}}{\sqrt{\mu} + 1}\right) \tag{16.71}$$

If the oscillator ceases to oscillate when this transit time is some fraction k of a cycle, then the limiting wave length of oscillation is

$$\lambda_0 = \frac{cT_{cp}}{k} \tag{16.72}$$

where $c = 3 \times 10^{10}$ cm per sec is the velocity of light. In terms of the specific value of T_{cp} this becomes

$$\lambda_0 = \frac{506}{k} \sqrt{\frac{\mu}{V_p}} \left(3d_{cg} + \frac{2d_{gp}}{\sqrt{\mu} + 1}\right) \quad \text{cm} \tag{16.73}$$

This is the relation for the limiting wave length of oscillation that has been sought. It gives the limiting wave length in terms of the plate voltage of the tube in volts, the interelectrode distances in centimeters, the amplification factor, and the fraction of period required for an electron to travel from cathode to plate, k. Gavin found that the limiting total electron transit time was approximately half a cycle, $k = 0.5$, for a series of tubes of the radiation-cooled type with single grid and plate leads brought out the top of the tube. For tubes of the lighthouse type, to be described, the limiting fraction of the cycle required by an electron to travel from cathode to plate is greater, of the order of ¾.

It is possible to evaluate the transit time in the grid-plate region more accurately than was done in Eq. (16.70). The ratio of the grid-plate to the cathode-grid transit time as obtained from the assumption that there is full space limitation of emission in the cathode-grid region and no space charge in the grid-plate region is

$$\frac{T_{gp}}{T_{cg}} = \frac{\dfrac{2}{3}\dfrac{d_{gp}}{d_{cg}}}{1 + \sqrt{\dfrac{V_p}{V_a}}} \tag{16.74}$$

where V_a is the effective potential of the grid plane, V_p is the plate potential, d_{gp} is grid-plate distance, d_{cg} is cathode-grid distance, T_{gp} is grid-plate transit time, and T_{cg} is cathode-grid transit time. This expression is accurate within a few per cent for values of $\frac{T_{gp}}{T_{cg}}$ less than ½. For cases in which the effective grid potential is relatively large compared with the plate potential a more accurate expression that considers space-charge effects in the grid-plate as well as in the cathode-grid region is needed. Such an expression has the form[1]

$$\sqrt{\frac{V_p}{V_a}} = \frac{2\dfrac{d_{gp}}{d_{cg}}}{3\dfrac{T_{gp}}{T_{cg}}} - 1 + \frac{1}{3}\left(\frac{T_{gp}}{T_{cg}}\right)^2 \tag{16.75}$$

This is a cubic equation, which is a little inconvenient to solve, but the relation between the different variables is represented by the nomographic chart of Fig. 16.21. It will be recognized that Eq. (16.75) reduces approximately to Eq. (16.74) when the ratio $\frac{T_{gp}}{T_{cg}}$ is small enough so that the third term on the right-hand side of Eq. (16.75) may be neglected.

---

[1] LLEWELLYN, *op. cit.*, p. 36.

16.10. Large-signal Effects. The analysis of the previous sections has been mostly restricted to small signal voltages, with the attendant assumption that none of the electrons were ever turned back. In actual tubes these assumptions will often not apply because of the large signal voltages developed. When large signal voltages are developed, a new set of considerations apply and it is of some interest to examine these briefly. Unfortunately, the analysis of large-signal effects is so complicated that only relatively simple cases can be solved.[1]

Transit-time Effects in Diodes. The simplest case of large-signal effects that can be handled yielding some generally useful information is that of the unbiased diode. Even here, it is not possible to take into account the effect of space charge because of attendant complications of the analysis. Accordingly, let it be assumed that the emission is temperature-limited. This means that the same number of electrons per second will be liberated whenever the potential gradient at the cathode is positive. This assumed condition is often realized in pulsed oscillators, where the voltages are so extremely high. Of principal interest is the behavior of the electrons with regard to such matters as their transit time, conditions for traveling a certain distance before turning around, and so on. The voltage will be assumed to be of the form

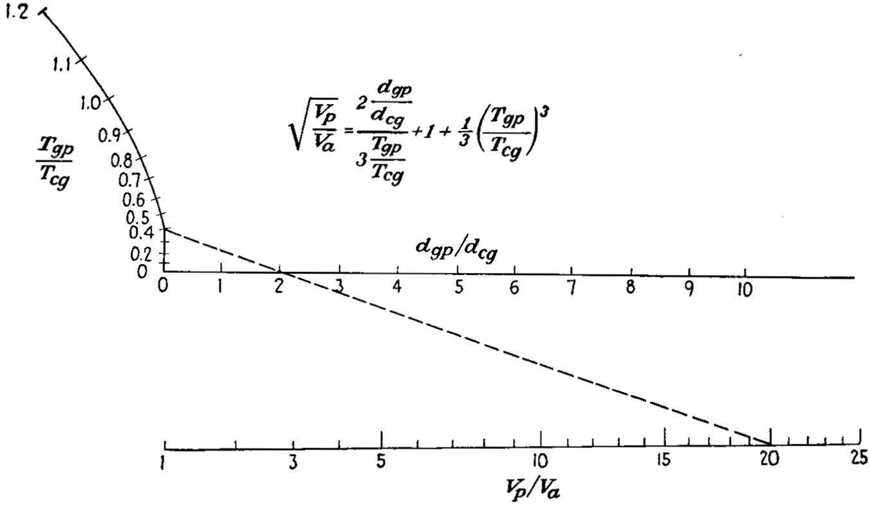


FIG. 16.21.—Nomographic chart relating triode transit times in the presence of space charge to electrode dimensions and voltages as given by Eq. (16.75).

$$V(t) = V \sin \omega t \tag{16.76}$$

---

[1] WANG, C. C., Large Signal High Frequency Electronics of Thermionic Vacuum Tubes, *Proc. I.R.E.*, vol. 29, pp. 200–214, April, 1941.

The corresponding differential equation of motion is

$$e\frac{V}{s}\sin\omega t = m\frac{d^2x}{dt^2} \tag{16.77}$$

where s is the distance between cathode and plate of the diode. The starting conditions are that the initial velocity and acceleration of the electron are zero. As a result, a first integration of Eq. (16.77) gives

$$v = \frac{eV}{m\omega s}(\cos\omega t_1 - \cos\omega t) \tag{16.78}$$

where v is the velocity of the electron and t_1 is the starting time. A second integration gives

$$\frac{x}{s} = \frac{e}{m}\frac{V}{\omega^2 s^2}(\theta\cos\theta_1 - \cos\theta_1\,\theta_1 + \sin\theta_1 - \sin\theta) \tag{16.79}$$

where $\theta = \omega t$ is a transit angle. This equation gives the fractional distance from cathode to plate in terms of the elapsed transit angle and the starting angle. Note that as a coefficient of the right side of the equation there appears the dimensionless parameter of Eq. (16.47). It is convenient to express the distance s in centimeters and the frequency in megacycles, in which case Eq. (16.79) takes the form

$$\frac{x}{s} = \frac{0.232V}{f_{mc}{}^2 s_{cm}{}^2}(\theta\cos\theta_1 - \theta_1\cos\theta_1 + \sin\theta_1 - \sin\theta) \tag{16.80}$$

The behavior of electrons in an unbiased diode is best studied by plotting their position as a function of time from Eq. (16.79). Such a plot is given in Fig. 16.22.[1] It looks different from the more commonly presented figure that results when the voltage is a square wave, but it is the true representation for the unbiased diode without space charge with an applied sine wave of voltage. This figure contains a great store of useful information from which many interesting properties of the electron trajectories may be observed. Curves are shown for electrons emitted every 30 deg of the plate-voltage cycle. The most important observation is that electrons will flow only when the plate voltage is positive or between 0 and 180 deg for the sine wave of voltage assumed. In the second place, all electron curves consist of a straight line with a superimposed sinusoidal component. The slope of the straight-line portion of the curve is proportional to the rate of change of voltage with time at the instant of emission. This makes the slope maximum at the

---

[1] Curves such as those of Fig. 16.22 are readily plotted by graphical means. See KOMPFNER, RUDOLF, Transit-time Phenomena in Electronic Tubes, *Wireless Eng.*, vol. 19, pp. 2–7, 1942.

beginning of the sine-wave cycle, zero at the positive peak of voltage, and negative for the rest of the positive half cycle. Any electrons emitted during the first half of the positive cycle will eventually reach the plate, no matter how far distant. Electrons emitted during the second half of the positive half cycle of voltage may return to the cathode if they do not strike the plate electrode first. This means that the plate, no matter how situated, will always receive at least half the emitted electrons.

The curves of Fig. 16.22 are universal because of the fact that the distance and the time are expressed in units of frequency, cathode-plate distance, and voltage. Increasing the frequency increases the time parameter in direct proportion and the distance parameter in proportion to the square of the separation. Increasing the voltage decreases the distance parameter inversely as the voltage.

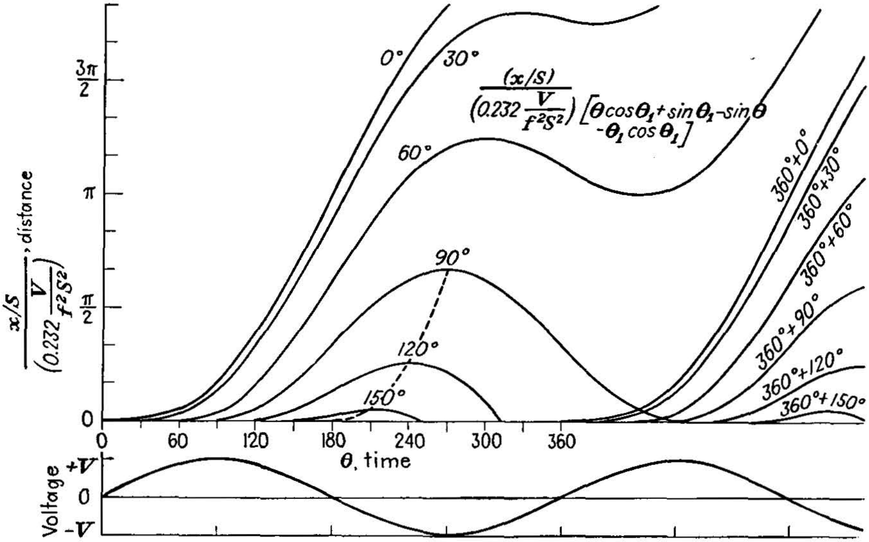


FIG. 16.22.—Distance-time behavior of electrons in an unbiased diode without space charge.

The point at which any curve of Fig. 16.22 reverses direction is given by

$$\theta = 360° - \theta_1 \tag{16.81}$$

The locus of the reversal points is shown by a dashed line in Fig. 16.22 up to the curve for the 90 deg electron, beyond which no electrons will return to the cathode. The starting time of a grazing electron for any plate distance, voltage, and frequency may be calculated from the above relation.

The current associated with the electron movements of Fig. 16.22 may be computed by summing the quantity $\frac{ev}{s}$ for all electrons in transit over a cycle of voltage. This is rather difficult to do analytically because different conditions hold for different parts of the cycle. Suppose, for instance, that frequency, voltage, and electrode separation are such that the electron liberated at the peak of the voltage wave ($\theta = 90$ deg) just grazes the plate. Then in evaluating the current the first electron requires 150 deg of transit angle to arrive at the plate. During this time the current increases rapidly. As soon as the first electron strikes the plate, the current drops, quite rapidly at first, and then more slowly because the contribution to the total induced current made by the electrons striking the plate is much greater than that of the new electrons liberated at the cathode. The induced current then drops because the velocity of the electrons drops progressively from the beginning of the cycle. After the voltage reverses, some of the electrons reverse direction and return to the cathode so that some electrons are inducing positive current while the returning electrons are inducing negative current. Eventually, the current will become negative but will reach a finite magnitude and then decrease. Although analytical treatment of the current is difficult, the shape of the induced-current pulses is readily obtained by graphical methods.[1] Some of the resultant shapes of the induced-current pulses are shown in Fig. 16.23. Curves are labeled with values of the distance parameter of Fig. 16.22, with x set equal to s, that is, values of $\frac{f_{mc}{}^2 s_{cm}{}^2}{0.232V}$. It must be remembered that the diode emission is temperature-limited, which means that for small transit angles the current pulse is expected to be square.

The curves of Fig. 16.23 show the degeneration of the square pulse of current, which exists for short transit times, into a nearly triangular pulse with a negative tail as the transit angle increases. For very short values of the determining parameter the current pulse is very nearly square except for a sharp spike at the front of the pulse, which rises to twice the height of the rest of the pulse. This occurs because the initial induced current is made up of the contributions of a large number of high-velocity electrons, which are bunched at the front of the electron stream. When these are retired from action on striking the plate, the current drops very rapidly because the successive electrons come along at a lower velocity and are not so strongly bunched. This bunching

---

[1] Kompfner, Rudolf, Current Induced in an External Circuit by Electrons Moving between Two Plane Electrodes, *Wireless Eng.*, vol. 19, pp. 52–55, February, 1942. Figure 16.23 is from this paper.

action is evident from the curves of Fig. 16.22, where it is seen that the electrons liberated at 0 and 30 deg are separated by a time interval, over most of their path, which is less than half that between any two successive curves corresponding to electrons liberated at adjacent 30 deg intervals. As the transit angle increases, the peak is reduced somewhat and the subsequent current falls off more gradually. At the same time a negative pulse of current forms, due to electrons falling back on the cathode as the plate voltage becomes negative.

Transmit-time Effects in Triodes. Although the above remarks have been restricted to the diode, they are readily extrapolated to cover the behavior of a triode. For a triode operating Class B the electron behavior may be expected to be very similar to that of the diode under the conditions just discussed. As a result, the distance-time picture for the cathode-grid region will be very similar to that of the corresponding portion of Fig. 16.22. On passing through the grid plane the electrons will encounter a positive gradient of potential that is quite large and varying sinusoidally with time. The plate voltage will adjust itself in an amplifier so that it will be minimum when the fundamental component of plate current is a maximum. This means that the plate voltage lags the negative grid voltage and the first electrons passing through the grid plane will encounter a voltage gradient which is higher than the minimum. As a result, the first electrons passing through the grid will be accelerated more than the electrons immediately following. The resultant distance-

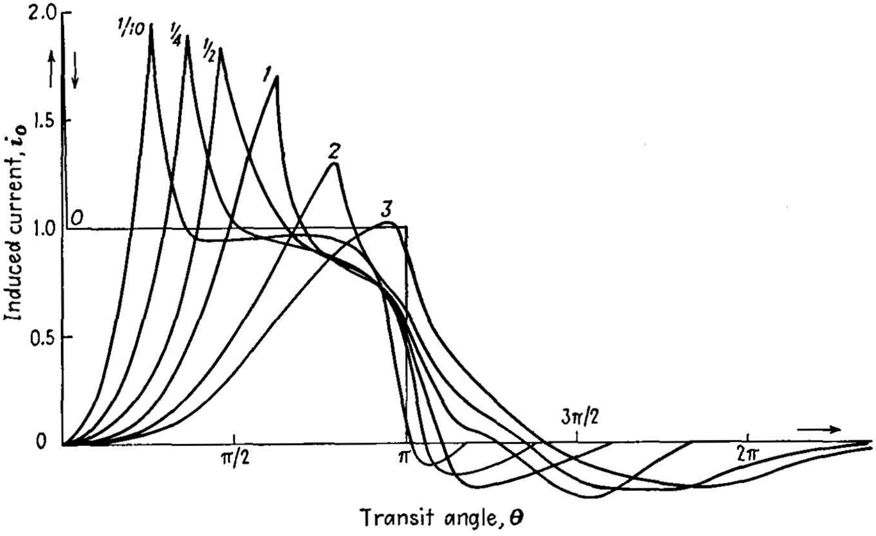


FIG. 16.23.—Induced-current pulses in an unbiased diode without space charge.

time diagram will take the form shown in Fig. 16.24. Here it is seen that the plate-current pulse has been stretched out considerably by what is essentially a debunching action in the grid-plate space. The length of the plate pulse is determined by the interval between the times when the first electron enters the grid-plate space till the last electron leaves it. The corresponding plate-current pulse is shown in Fig. 16.25. For comparison, there are shown in this figure the plate current and plate voltage that would exist at low frequencies for a given grid driving voltage. The plate-current pulse is seen to be displaced and distorted. The displacement takes the form of a phase lag, due both to the grid-plate transit time and to the debunching action of the field, which causes the electrons to be progressively retarded throughout the current pulse. The distortion is due primarily to the debunching action.

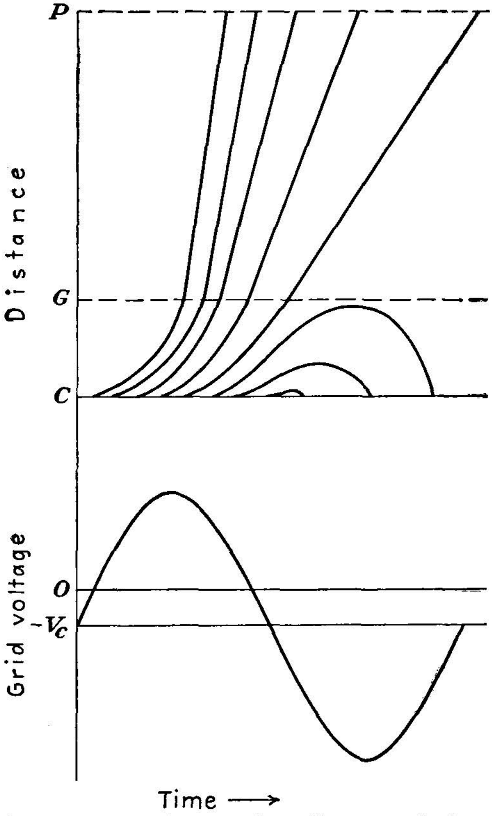


FIG. 16.24.—Distance-time diagram of electrons in a triode.

The phase shifts which occur in the plate current because of transit-time effects are sufficiently pronounced so that there is a considerable difference between the operation of amplifiers and oscillators. In an amplifier the plate voltage will adjust itself to the phase lag of the plate current In an oscillator the plate is coupled to the grid so that the two electrode voltages are ordinarily 180 deg out of phase. As a result, the output of an oscillator falls off more rapidly with frequency than does that of an amplifier. This is because in the amplifier the output is decreased only as the reduction in the fundamental component of plate current arising from pulse distortion but is independent of the phase of the plate current. In the oscillator the output is reduced because in addition to the reduction of the fundamental component of plate current the phase of the current relative to a plate voltage of fixed phase causes a further lowering of the output power. As a result, the general experience is that amplifiers will give output when the transit angle is increased 50 per cent beyond that at which oscillators cease to operate.[1]

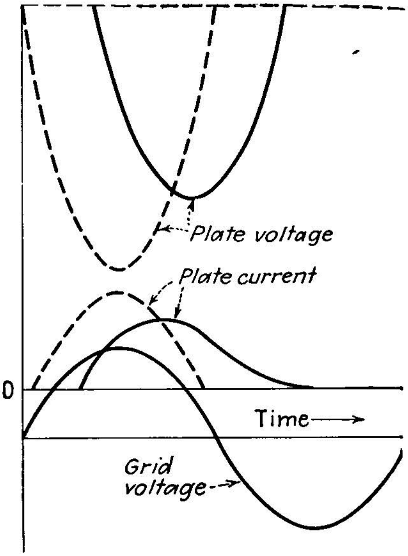


FIG. 16.25.—Distortion of the plate-current pulse in a Class *C* triode amplifier caused by transit-time effects.

Transit-time Effects in Tetrodes. Tetrodes have inherently better operating characteristics than triodes as far as transit-time effects are concerned. This is because the control grid is followed by a positive screen grid maintained at a fairly high potential. As a result, the electrons are accelerated fairly uniformly as they pass the control grid and the attendant debunching action is much less than is the case with the triode. In addition, the over-all transit time from control grid to plate of the screen-grid tube may actually be less than is the case for the triode because the electron is moving in regions of higher potential most of the time. In the screen-plate region of the tetrode the electrons will encounter a retarding potential gradient that will exert some debunching action but that will not be as strong as is the case with the grid-plate region of the triode. A typical set of distance-time curves of a tetrode is shown in Fig. 16.26. These curves exhibit all the properties mentioned above.

---

[1] WAGENER, *op. cit.*

To be a satisfactory tube, a tetrode should be built on the principle of the beam-power tubes, *i.e.*, with aligned control and screen grids. This alignment, along with proper interelectrode dimensions, serves two purposes. (1) It reduces the direct current to the screen. (2) It produces a strong enough potential minimum by virtue of space-charge effects to suppress secondary emission from the plate. The beam tetrode has a number of advantages for ultra-high-frequency operation in addition to the favorable transit-time characteristics mentioned above. In the first place it is possible to attach separate resonant circuits to the cathode and control grid on the one hand and to the screen grid and plate on the other hand. By means of concentric lines or cavity resonators it is possible to separate almost completely the fields of what are then the input and output resonators. The only interaction that exists is through the medium of the cathode-plate capacity, which is inherently small.

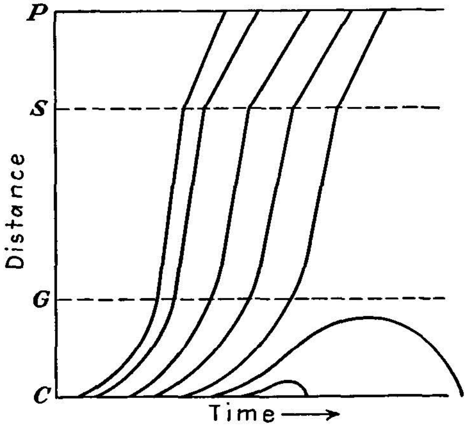


FIG. 16.26.—Distance-time behavior of electrons in a tetrode.

Beam tetrodes built so that they may be connected to concentric-line resonators have been very successful as ultra-high-frequency oscillators. Such tubes, known as "resnatrons,"[1] have been built to give continuous power outputs of 60 kw at frequencies of 500 mc.[2,3] In the form of

---

[1] The resnatron, also known as the "Sloan-Marshall tube," was developed at the University of California. It underwent further development both at the Westinghouse Laboratories and at the Radio Research Laboratory during the Second World War.

[2] SALISBURY, W. W., The Resnatron, *Electronics*, vol. 19, pp. 92–97, 1946.

[3] DOW, W. G., G. HOK, and H. W. WALSH, "Very High-Frequency Techniques" (report of the Radio Research Laboratory), Chaps. XVIII, XIX, McGraw-Hill, New York, 1947.

a very large tube operating continuously at voltages of 10 to 15 kv the resnatron takes advantage of the inherent benefits of voltage scaling. The large size makes possible water-cooled screen grids, thus removing what might otherwise be a limiting factor in the tube design. In addition, the high voltage reduces secondary emission since at high enough voltages the ratio of secondary- to primary-electron currents goes down again. Evidence of this is found in the fact that scaled-down tubes designed to give about 1 kw of continuous power have shown efficiencies of only about 20 per cent, whereas the large tubes have given efficiencies of the order of 50 to 60 per cent.

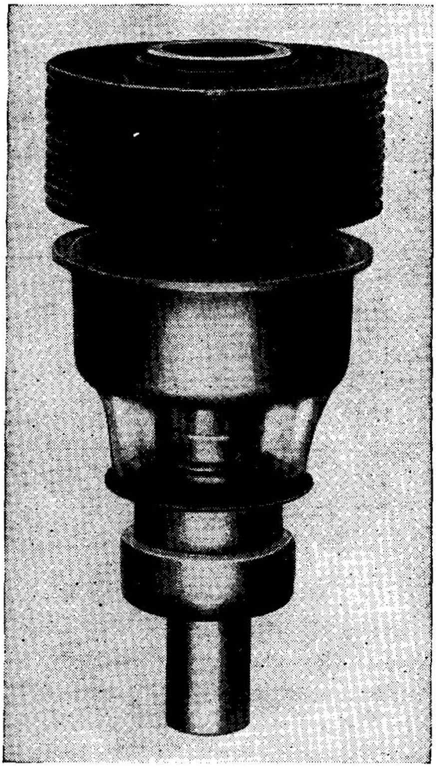


FIG. 16.27.—Lighthouse tube—external view. Type 2C39—plate at top.

16.11. Disk-seal Tubes. There have recently been developed a number of tubes known as "disk-seal" or "lighthouse tubes."[1,2] Essen-

---

[1] Disc Seal Tubes, *Gen. Elec. Rev.*, vol. 48, pp. 50–51, January, 1945.

[2] McArthur, E. D., Disc Seal Tubes, *Electronics*, vol. 18, pp. 98–102, February, 1945.

tially, the lighthouse tube is a tube with so many leads brought out from a single electrode that the leads become a disk. The external form of the tube is shown in Fig. 16.27. In small power tubes the plate is brought out through a cap at the top of the tube. The grid is brought out through a disk at the center of the active tube structure, and the cathode is brought out through a cylinder at the base of the tube. The electrodes are separated by cylindrical sections of glass, which are butt-sealed to the metal disks with which they are in contact. The internal electrodes are of a plane-parallel design. The grid is of parallel wires supported over the hole in a disk in which currents flow radially to the outside circuit. The cathode and plate are the ends of small-diameter cylinders, which are supported by the glass tubing. A cutaway view of a low-power lighthouse tube is shown in Fig. 16.28. In high-power lighthouse tubes having plate dissipations between 20 and 100 watts, the position of the plate and cathode is reversed, with the result that the plate is at the large end of the tube. This permits radiating fins to be attached to the plate for air cooling.

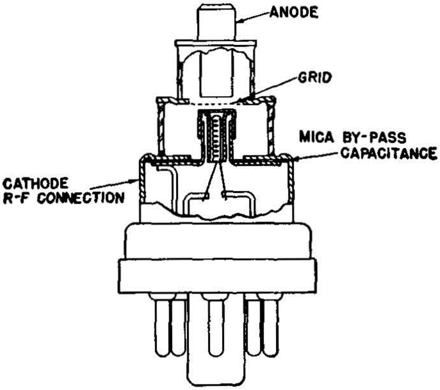


FIG. 16.28.—Lighthouse tube—cutaway view. Type 2C40.

By virtue of the electrode arrangement of lighthouse tubes, lead inductance is cut to a minimum. Therefore, operation is possible to much higher frequencies than with tubes having single or even double wire leads. Lighthouse oscillators have been made that will operate to 3,000 mc (1946).

A further advantage of the electrode arrangement is that it makes the tube suitable for use in a double concentric-line structure such as is shown in Fig. 16.29. The line type of resonator makes it possible to operate at higher frequencies than the natural resonant frequency of the shorted tube. This is possible with resonator operation on a three-quarter wave-length mode, for which it is possible to go to such high frequencies that the voltage node is pushed inside the tube. In addition, it is possible to gang the cathode and plate resonators for broad-band operation. This has been done successfully over a 3 to 1 band of frequencies. If minor trimming adjustments are permitted, it is possible to produce an oscillator that will operate at 300 to 3,000 mc. Not shown in Fig. 16.29 are the input and output coupling devices and the intercavity coupling. These, however, are usually loops or probes of conventional form and can readily be imagined.

In addition to the somewhat conventional resonator arrangement of Fig. 16.29, various special methods of coupling the cathode and plate lines for oscillator operation may be used.[1-4]

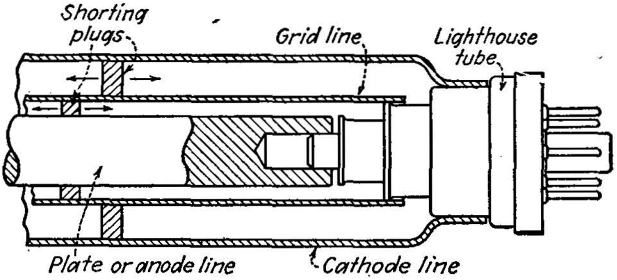


FIG. 16.29.—Double concentric-line oscillator utilizing a lighthouse tube.

Lighthouse-tube amplifiers have found some use in the ultra-high-frequency band. Here the operation is that of a grounded-grid amplifier, with the attendant advantages of low input impedance, high output impedance, low interaction between input and output circuit, and the relatively low noise associated with a triode.[5,6] At frequencies below 1,200 mc an amplifier-converter combination using lighthouse tubes is superior in its noise figure to a crystal mixer. There will undoubtedly be advances in tube design, which will extend appreciably the present limits of such tubes.

---

[1] General Electric Company, *Electronic Tube Eng. Bull.* ET-Bl, June, 1945.

[2] GUREWITSCH, A. M., Cavity Oscillator Circuits, *Electronics*, vol. 19, pp. 135–137, February, 1946.

[3] GUARRERA, J. J., Tunable Microwave Cavity Resonators, *Electronic Ind.*, vol. 5, pp. 80–82, March, 1946.

[4] GUREWITSCH, A. M., and J. R. WHINNERY, Microwave Oscillators using Disk Seal Tubes, *Proc. I.R.E.*, vol. 35, pp. 462–473, May, 1947.

[5] DISHAL, MILTON, Gain and Noise of Grounded Grid Amplifier at Ultra-high Frequencies, *Proc. I.R.E.*, vol. 32, pp. 276–284, May, 1944.

[6] JONES, M. C., Grounded-grid Radio Frequency Voltage Amplifiers, *Proc. I.R.E.*, vol. 32, pp. 423–429, July, 1944.

CHAPTER 17

VELOCITY-MODULATED TUBES, OR KLYSTRONS

17.1. The Bunching Principle. We have seen in the last chapter that there are some severe limitations on conventional tubes which conspire to make their operation relatively poor at ultra-high frequencies. The principal limitations arise from electron-transit time, lumped electrical reactances, and low-Q resonant circuits. With negative-grid tubes each of these factors has been pushed considerably beyond conventional form, and yet the performance characteristics of these tubes leave much to be desired at the ultra-high frequencies. It was not strange, therefore, that various investigators sought means of efficiently generating and amplifying power at ultra-high frequencies by a totally new attack on the utilization of electronic principles. This new attack, which resulted in the modern klystron, involved a combination of the electronic-bunching principle and the cavity resonator. Both were necessary for the production of a successful tube. The bunching principle overcame the transit-time difficulties, and the use of cavity resonators largely eliminated lumped reactances and produced high-Q resonant circuits.

In negative-grid tubes the transit-time difficulties encountered arise largely because the electrons in the cathode-grid space start at zero velocity, hence inherently move slowly, and thus take a large fraction of a cycle to get from cathode to grid as the frequencies get up into the ultra-high region. Since the method of producing variations in plate current is inextricably associated with the large cathode-grid transit angle, the negative-grid tube always operates poorly if the frequency is raised high enough. Means are therefore sought for producing variations in current that are not limited by transit time. Such means were independently conceived by the Heil brothers and the Varian brothers.[1,2] Both these pairs of men proposed devices utilizing an electron beam

---

[1] Heil, A. A., and O. Heil, Eine neue Methode zur Erzeugung kurzer ungedampfter elektromagnetishen Wellen von grosser Intensität (A New Method of Generating Short Undamped Electromagnetic Waves of High Intensity), *Zeit. für Phys.*, vol. 95, pp. 752–773, July, 1935.

[2] Varian, R. H., and S. F. Varian, A High Frequency Oscillator and Amplifier, *Jour. Appl. Phys.*, vol. 10, pp. 321–327, May, 1939.

similar to that in a cathode-ray tube and then obtaining current pulses by periodically varying the beam-electron velocity a small amount about its average value. When the velocity of the beam electrons is varied, those which have been speeded up will subsequently overtake those which have been slowed down. The result is that a short distance beyond the point where the electron velocity is varied there will appear bunches of current from which power can be extracted. This in its essence is the bunching principle. The formation of electron bunches is illustrated in Fig. 17.1. In this figure there is shown the behavior of a series of electrons, represented by dots, released at uniform intervals through a cycle of alternating voltage, which is applied between two closely spaced grids of an input resonator known as the *buncher*. The voltage between the grids of the buncher serves to modify the velocity of the electrons as they arrive from the cathode. Some electrons are speeded up a little, and some are slowed down a little. The bunching action resulting from the regrouping of the electrons of different velocity is evident from the figure. Thus, the application of the bunching principle *utilizes* transit-time effects, whereas in negative-grid tubes transit-time effects are detrimental.

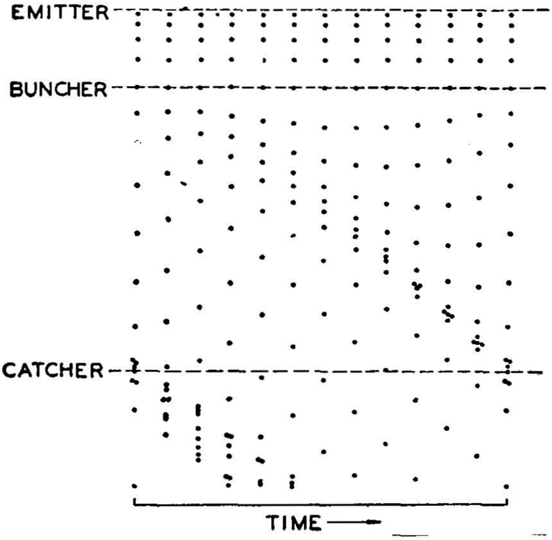


FIG. 17.1.—Elementary representation of bunching action.

To utilize the current bunches that are formed along the beam of electrons it is necessary to extract energy from this current stream. This is done by passing the bunched beam through the grids of an *output resonator or catcher.* As the electrons pass through the grids, charges are induced on the grids that change in magnitude and sign as each electron passes through. In effect, this causes the induced charge to flow through the resonator to produce a current flow that delivers power to the resonator by passing through its equivalent resistance. Tubes utilizing the velocity-modulation principle are generally referred to as klystrons after the Greek verb "klyzein," expressing the breaking of waves on a beach.

The physical form of the klystron has been described briefly in Sec. 2.7. Further information with specific reference to klystron amplifiers is given in Sec. 17.5. The klystron differs from negative-grid tube amplifiers and oscillators in two respects. First, the current pulses are produced by a velocity variation rather than by an intensity variation. Second, energy is extracted from the current pulses by the charges induced on passing the beam through a short region of varying field instead of a long one. Furthermore, the extraction of energy does not require the electrons to strike the electrodes attached to the resonator. It is not always recognized that energy is extracted from electrons in a negative-grid tube by forcing the electron to move against an alternating component of electric field, but this is the case. Electrons in a negative-grid tube will arrive at the plate with velocities which are on the average less than those which they would have had if no alternating component of electric field were present. The residual energy represents a loss and appears as heat liberated at the plate electrode. The difference between the direct power input to the tube and the heat liberated at the plate appears as useful output. In the klystron the electrons that have passed through the catcher grids emerge with less energy on the average than they would have had if the beam had been unbunched. The difference in energy goes into useful r-f power. The residual energy appears as heat on a collector electrode.

17.2. Cavity Resonators. The desirability of extracting energy from electrons by passing them through a short region of alternating electric field, which as we shall see leads to greater efficiency of conversion of energy, requires the use of cavity resonators. The outstanding characteristic of these devices is that current flow and associated alternating components of field are entirely internal to the resonator.[1,2] A concentric-line resonator in which the inner conductor is shorted to the outer at one end and which is coupled by a small capacity gap to the outer

---

[1] HANSEN, W. W., A Type of Electrical Resonator, *Jour. Appl. Phys.*, vol. 9, pp. 654–663, October, 1938.

[2] HANSEN, W. W., and R. D. RICHTMYER, On Resonators Suitable for Klystron Oscillators, *Jour. Appl. Phys.*, vol. 10, pp. 189–199, March, 1939.

conductor at the other is one form of cavity resonator. Such resonators have already been referred to in the discussion of grounded-grid amplifiers and oscillators utilizing lighthouse tubes. Such a shorted concentric-line resonator will resonate when the capacitive reactance of the gap equals the inductive reactance of the line and hence if the gap is small will oscillate at lengths somewhat less than $\frac{1}{4}$, $\frac{3}{4}$, $\frac{5}{4}$, etc., of a wave length. In such a resonator the electric and magnetic fields will be totally confined to the interior of the resonator. In addition, if the wall thickness is large compared with the skin depth,[1] as it usually is, the currents associated with the fields will flow in a thin layer on the interior conducting surfaces of the resonator—no currents will flow on the outside of the resonator. The electric and magnetic fields in such a resonator will be established 90 deg out of time phase. As a result, when the magnetic field is a maximum, the electric field is zero, and vice versa. The total energy stored in magnetic and electric fields at any point on the cycle is very nearly constant over a period of a cycle or two. The total stored *energy* of a freely oscillating resonator decreases exponentially over long periods of time and drops by a factor of 2.718 in a time of $\frac{Q}{2\pi}$ cycles. Electric- and magnetic-field components and associated voltages and current likewise decrease exponentially with time in a freely oscillating resonator. *All the currents, voltages, and field components* will decrease

---

[1] "Skin effect" is a term applied to the tendency of ultra-high-frequency currents to flow in a layer on the surface of a conductor. This comes about because of the tendency of the current to flow in such a way that it is encircled by the fewest number of magnetic-flux lines. Thus with circular conductors the current tends to flow on the surface, and hollow tubes are just as good conductors at sufficiently high frequencies as are solid conductors. Since the penetration of current at 6.5 mc is only 0.001 in. in copper and is less at higher frequencies, most skin-effect problems for ultra-high frequencies can be solved by assuming that the surfaces are plane, *i.e.*, that the radius of curvature of the surface is much greater than the skin depth. For plane-surface conductors the relations are relatively simple (see WHEELER, H. A., Formulas for the Skin Effect, *Proc. I.R.E.*, vol. 30, pp. 412–424, September, 1942). The current density drops off exponentially into the conductor, and the effective skin depth is defined as that depth at which the current density is $\frac{1}{2.718}$ of the surface current density. The formula for skin depth is $d = (\pi f \mu \sigma)^{-\frac{1}{2}}$ meters, where μ is the permeability in mks units and σ is the conductivity of the material. For copper this reduces to $2.57 \times 10^{-3} f^{-\frac{1}{2}}$ in. The corresponding surface resistivity is $R = (\pi f \mu \rho)^{\frac{1}{2}}$ ohms per unit square, where ρ is the volume resistivity in ohm-meters and other units are mks. For copper this reduces to $R = 2.61 \times 10^{-4} f_{mc}^{\frac{1}{2}}$ ohms. The direction of current flow is always parallel to the surface and directly proportional to the strength of the tangential component of magnetic-flux density at the surface.

by a factor of 2.718 in a time of $\frac{Q}{\pi}$ cycles. Another characteristic of a concentric-line cavity resonator is that the magnetic-flux lines always encircle current, whether this be in the form of conduction or displacement current (displacement current is equal to the time rate of change of electric field multiplied by the dielectric constant).

So far, all the remarks on closed resonators have been confined to concentric-line resonators. Many other closed or cavity resonators are possible. It is possible to get electromagnetic-field resonances that exhibit all the above-mentioned characteristics in simple cavities such as cubes or cylinders or spheres. These have limited usefulness for electronic purposes, for it is not possible to shoot an electron through such pure cavities in a sufficiently small fraction of a cycle to secure an efficient energy exchange between the field and the electron. This is because the dimensions of pure cavities are relatively large compared with a wave length. The diagonal of a cubical resonator, for instance, is equal to the wave length of oscillation for operation on its lowest resonant frequency. For this reason, the pure cavity resonators find their principal application in such devices as wavemeters and filter elements rather than in vacuum tubes. For tube applications, cavity resonators that are reentrant, *i.e.*, have internal projections from the walls, are of most use because this form produces a very intense electric field, concentrated in a small region, through which it is convenient to shoot electrons. Some typical resonators of this kind are shown in Fig. 17.2. The resonators *a*, *b*, and *c*, shown in this figure have the same resonant frequency. Extreme forms such as *a* and *c* may be considered equivalent to coaxial and radial lines, respectively, with capacity loading[1] and may be studied

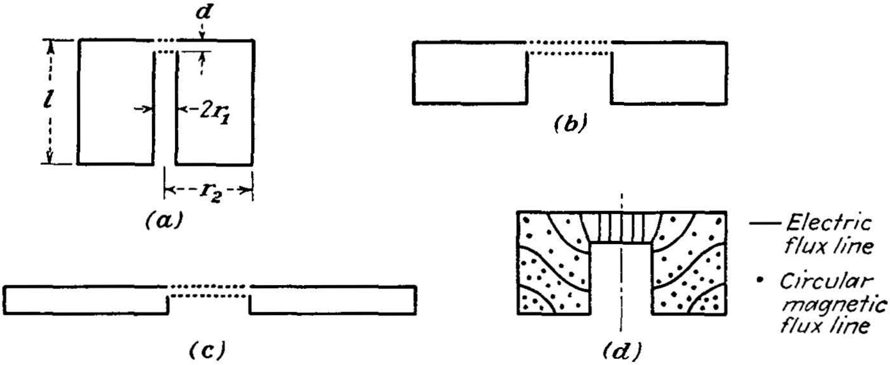


FIG. 17.2.—Reentrant cavity resonators.

---

[1] RAMO, S., and J. R. WHINNERY, "Fields and Waves in Modern Radio," pp. 404–411, Wiley, New York, 1944.

by conventional transmission-line formulas. Intermediate forms such as are shown in Fig. 17.2*b* can be analyzed only by more powerful methods.[1,2] The electric field is almost entirely confined to the resonator gap. The shape and location of the electric- and magnetic-flux lines are shown in Fig. 17.2*d*.

Every cavity resonator has an infinite number of resonant frequencies. Of these the lowest frequency of resonance is usually that of most interest. In terms of response to a sinusoidal excitation, resonance occurs when equal amounts of energy are stored in the electric and magnetic fields on successive quarter cycles. At the frequencies for which this occurs the impedance, or ratio of equivalent voltage to equivalent current, will be a maximum at any point in the resonator. In terms of the transient response to a shock excitation, currents and voltages will occur as a combination of exponentially damped sine waves, provided only that the losses are not excessive. The resonant frequencies are the frequencies of the individual damped-sine-wave components. In terms of field theory there will be certain solutions of the wave equation that fit the resonator shape at distinct frequencies. These frequencies are the resonant frequencies. For certain simple cavities the shapes of these fields are readily found, but in general they are difficult to find.

The longest resonant wave length of a resonator such as that shown in Fig. 17.2*a* may be determined quite closely by solving for the frequency for which the capacitive reactance of the gap equals the inductive reactance of the shorted transmission line formed by the rest of the resonator. The formula for the resonant wave length is approximately

$$\lambda = 2\pi r_1 \sqrt{\frac{l}{2d} \ln \frac{r_2}{r_1}} \tag{17.1}$$

for dimensions as in Fig. 17.2*a*. This formula gives the resonant wave length to within about 5 per cent for resonators of the shape shown in Fig. 17.2*a* but will give values that range from 60 to 80 per cent of the true value for resonators of the shape shown in Fig. 17.2*b*.[3] It will be observed from Eq. (17.1) that the resonant wave length is proportional to the linear dimension of the resonator. This proves to be a general property so that the resonant wave lengths of geometrically similar cavity reso-

---

[1] HANSEN, W. W., On the Resonant Frequency of Closed Concentric Lines, *Jour. Appl. Phys.*, vol. 10, pp. 38–45, January, 1939.

[2] HAHN, W. C., A New Method for the Calculation of Cavity Resonators, *Jour. Appl. Phys.*, vol. 12, pp. 62–28, January, 1941.

[3] Curves giving the resonant wave length of resonators having the approximate shape of that in Fig. 17.2*b* are given in "Microwave Transmission Design Data," pp. 200–204, Sperry Gyroscope Company, Brooklyn, 1944.

nators of different sizes are directly proportional to the size of the resonator. Resonators of the general shape shown in Fig. 17.2c may be treated by radial-transmission-line theory.[1]

In cavity resonators it is difficult to identify the lumped reactance elements that are apparently involved. This occurs because the fields are distributed more than they are grouped; *i.e.*, there are not specific regions within the cavity within which the electric field exists alone and there is no magnetic field, and vice versa. As a result, it is more convenient to express the resonator characteristics in terms of its Q, shunt resistance, and resonant wave length instead of its inductance, capacity, and resistance. Some more general definitions of circuit parameters are therefore required. The Q of a cavity resonator, or reciprocal sharpness of resonance, is most conveniently defined in terms of the transient response to shock excitation. As previously mentioned, the fields within a resonator decay by a factor of 2.718 in a time of $\frac{Q}{\pi}$ cycles. Thus the time variation of any component of field is given by

$$E(t) = E_1 \epsilon^{-\frac{\pi t}{QT_0}} \sin\left(\frac{2\pi t}{T_0}\right) \tag{17.3}$$

where T_0 is the period of oscillation frequency. This is seen to correspond to the equation for the voltage decay in a high-Q series resonant circuit that has the form

$$V(t) = V_1 \epsilon^{-\frac{Rt}{2L}} \sin\left(\frac{2\pi t}{T_0}\right) \tag{17.4}$$

Upon substitution of $\frac{2\pi L}{RT_0}$ for Q, Eq. (17.4) may be obtained from Eq. (17.3) with the further recognition that the equivalent voltage of a cavity resonator is commonly taken as the line integral of the electric field along the line of maximum field strength. The stored electrical energy associated with a transient decay in a resonator will vary as the square of Eq. (17.3) since the energy stored in the electric field is obtained by integrating the square of the electric field throughout the volume of the resonator.

$$\mathcal{E}_e(t) = \mathcal{E}_{e1} \epsilon^{-\frac{2\pi t}{QT_0}} \sin^2\left(\frac{2\pi t}{T_0}\right) \tag{17.5}$$

This is seen to have twice the frequency and to decay exponentially at twice the rate of the field. Likewise, the energy stored in the magnetic field will be similar in form but shifted 180 deg in phase.

---

[1] See RAMO and WHINNERY, *loc. cit.*

$$\mathcal{E}_m(t) = \mathcal{E}_{m1}\epsilon^{-\frac{2\pi t}{QT_0}} \cos^2\left(\frac{2\pi t}{T_0}\right) \tag{17.6}$$

The total stored energy is the sum of the energy stored in the electric and magnetic fields and is given by

$$\mathcal{E}(t) = \mathcal{E}_e(t) + \mathcal{E}_m(t) = \mathcal{E}_1\epsilon^{-\frac{2\pi t}{QT_0}} \tag{17.7}$$

from which it is seen that the total energy stored in the fields decays by a factor of 2.718 in a time of $\frac{Q}{2\pi}$ cycles. The decay in this case is a simple exponential one. If the rate of change of stored energy with time be obtained by differentiating Eq. (17.7) with respect to time and solving for Q, there results

$$Q = -\frac{2\pi\mathcal{E}(t)}{T_0\frac{d\mathcal{E}}{dt}} \tag{17.8}$$

which may be written in words as

$$Q = \frac{2\pi \times \text{energy stored}}{\text{energy loss per cycle}} \tag{17.9}$$

This last is probably the most fundamental definition of Q that can be written and serves as a basis for the calculation of the Q of cavity resonators.[1] The energy stored in the field is most readily calculated from the peak value of the energy stored in the magnetic field. Likewise, the loss per cycle can be calculated from the ohmic losses associated with current flow, which is directly proportional to the tangential component of magnetic field at the inner surface of the resonator. The unloaded Q's of cavity resonators will be quite high, for the current flow associated with the fields is distributed over a large surface. The Q's of pure cavities (about 25,000 at 3,000 mc) are about ten times as high as those of reentrant cavities as shown in Fig. 17.2. The Q's of reentrant cavities, in turn, are at least ten times as high as those of resonant circuits consisting of lumped inductances and capacities. The Q's of loaded cavities, *i.e.*, cavities supplying power to an external load, may be calculated from Eq. (17.9) if the energy loss per cycle be considered as the sum of the energies delivered to the walls of the resonator and to the external load. In most applications the energy per cycle supplied to the external load will be many times that to the cavity itself, and as a result, the Q of a loaded cavity is much lower than that of the cavity when not loaded.

---

[1] HANSEN, W. W., A Type of Electrical Resonator, *Jour. Appl. Phys.*, vol. 9, pp. 654–663, October, 1938.

Another circuit parameter that is convenient in describing cavity-resonator characteristics is the equivalent shunt, or parallel, resistance. It is possible to talk about equivalent shunt resistance in terms of an equivalent voltage and the power supplied to the cavity walls and load by the fields. This is done in preference to the usual procedure of defining resistance as a ratio of voltage to current, for it is relatively more difficult to define an equivalent current than to determine the power consumed. The equivalent voltage is logically taken as the product of the negative electric field in the reentrant-cavity gap and the gap spacing. It should be pointed out that this is not a true voltage but merely an equivalent voltage, for the energy interchange between an electron crossing the gap and the field would be the same as for the low-frequency or direct-voltage case only if the electron were able to cross the gap in zero time. Since the frequencies involved in microwave generators are extremely high and the velocity of an electron is ordinarily only a fraction of the velocity of light, the electron will generally take an appreciable part of a cycle to cross the gap and the energy change of the electron will be somewhat less than the corresponding direct-voltage value. Nevertheless, the concept of an equivalent voltage defined by

$$V = -Ed \tag{17.10}$$

where E is electric intensity in the gap and d is the gap spacing, is an extremely useful one. The shunt resistance of a reentrant cavity resonator is given by

$$R_{sh} = \frac{V^2}{2 \times \text{power consumed}} \tag{17.11}$$

$$R_{sh} = \frac{V^2}{2P} \tag{17.12}$$

from the usual power relation, where P is the power consumed by the resonator walls and load and the factor 2 results from the use of peak rather than rms-voltage values. Equation (17.12) is a fundamental definition of shunt resistance that is consistent with lumped-reactance-circuit formulas. The shunt resistance may also be calculated from the fields for a cavity resonator. As with the Q, the shunt resistance of a loaded cavity is lower than that of the unloaded cavity because of the fact that the power loss includes the power delivered to the external circuit as well as that consumed in the cavity walls. The shunt resistance of pure cavities at frequencies of 3,000 mc is of the order of megohms. The shunt resistance of unloaded reentrant cavity resonators is of the order of hundreds of thousands of ohms at the same frequency. The shunt resistance of a loaded reentrant cavity resonator is likely to be of the order of tens of thousands of ohms, depending upon the degree of loading. These values are much higher than can be achieved with lumped reactance circuits at this frequency.

When the resonant frequency, Q, and shunt resistance of a cavity are known, its behavior in the vicinity of resonance is completely determined. The impedance of a resonant cavity in the vicinity of resonance is given approximately by

$$Z(\omega) = \frac{R_{sh}}{1 + 2j\,\delta Q} \tag{17.13}$$

where R_{sh} is the shunt resistance at resonance and δ is the fractional deviation from resonance.[1] It is sometimes of interest to determine the equivalent *series* resistance, inductance, and capacity from the resonant frequency, Q, and shunt resistance, though too much significance should not be attached to these equivalents. By analogy with the low-frequency relations in a closed series R,L,C circuit, the equivalent series elements are

Equivalent series resistance

$$R = \frac{R_{sh}}{Q^2} \tag{17.14}$$

Equivalent inductance

$$L = \frac{R_{sh}}{\omega_0 Q} \tag{17.15}$$

Equivalent capacity

$$C = \frac{Q}{\omega_0 R_{sh}} \tag{17.16}$$

where ω_0 is the equivalent resonant angular frequency. Because of the fact that the circuit elements are not lumped, equivalent values calculated by all the methods possible will not agree. For pure cavities it is found

---

[1] This is arrived at by assuming that the circuit is equivalent to the parallel combination of a resistance equal to the shunt resistance, a lossless inductance, and a lossless capacity whose resonant frequency is the same as that of the cavity. The impedance of this parallel combination is $\dfrac{1}{\dfrac{1}{R_{sh}} + j\omega C + \dfrac{1}{j\omega L}}$, which may be written as

$$\frac{R_{sh}}{1 + jR_{sh}\left(\omega C - \dfrac{1}{\omega L}\right)}.$$

Since $\omega_0^2 = \dfrac{1}{LC}$ and $Q = R_{sh}\omega C$, the impedance can be written as $\dfrac{R_{sh}}{1 + jQ\left[1 - \left(\dfrac{\omega_0}{\omega}\right)^2\right]}$.

If now $\dfrac{\omega}{\omega_0}$ be replaced by $1 + \delta$, the denominator expanded into a series, and only the first-power term of δ retained, then Eq. (17.13) results.

that the equivalent capacities as given by various arbitrary definitions differ between themselves by 50 per cent from a mean value which is about 50 per cent of the low-frequency value.[1] In reentrant cavities the difference between the different values possible will be less, say 10 per cent deviation from a mean value that is approximately 80 per cent of the low-frequency value.

This has been something of a digression on the subject of cavity resonators, but it has been desirable because of the necessity of using klystron circuits in which the electric field appears only between two closely spaced surfaces. Since only cavity resonators exhibit this property in anything approaching its ideal form, an understanding of the principal properties of such resonators is necessary before undertaking a complete discussion of klystron principles.

17.3. Mechanism of Energy Interchange between Electrons and Cavity Resonators. In klystron amplifiers and oscillators, resonators of the reentrant type are most extensively used. In such resonators, the gap surfaces are made as grids instead of solid conducting material, and electrons are shot through the spaces in the grids, with the result that the electrons will interact with the electric field which exists between the grids. The grid structures are shown in Fig. 17.2. The grids may consist of a fine mesh of wire that will give about 80 per cent electron transmission. Those electrons which hit grid wires will be retired from operation and give up their kinetic energy in the form of heat. In high-power tubes, where the heating from intercepted electrons may be appreciable, grids are sometimes made of copper strips arranged like the spokes of a wheel but with the center of the wheel cut out so that the strips are supported only from the outside of the grid aperture. For minimum interception of electrons such strips should present their thin edge to the oncoming electrons.

When an electron enters the space between grids, the lines of flux associated with the electron charge will terminate almost entirely on the grid conductors. As an electron moves from the first to the second grid, at first most of its flux lines will terminate on the first grid, where they will induce a positive charge. This situation is shown in Fig. 17.3*a*. As the electron advances toward the second grid, relatively less charge will be induced on the first grid and relatively more induced charge will appear on the second grid, as shown in Fig. 17.3*b*. In effect, the passage of an electron between the two grids causes a positive charge equal in magnitude to the electron charge to move from the first to the second grid. *This transfer of charge must occur through the resonator circuit.*

---

[1] RAMO and WHINNERY, *op. cit.*

In the previous chapter it was shown that the charge induced on one of two parallel plates between which an electron is passing is

$$q_2 = \frac{ex}{d} \tag{17.17}$$

In the case of the electron passing through the grids of a cavity resonator this is the charge induced on the second grid when the distance between

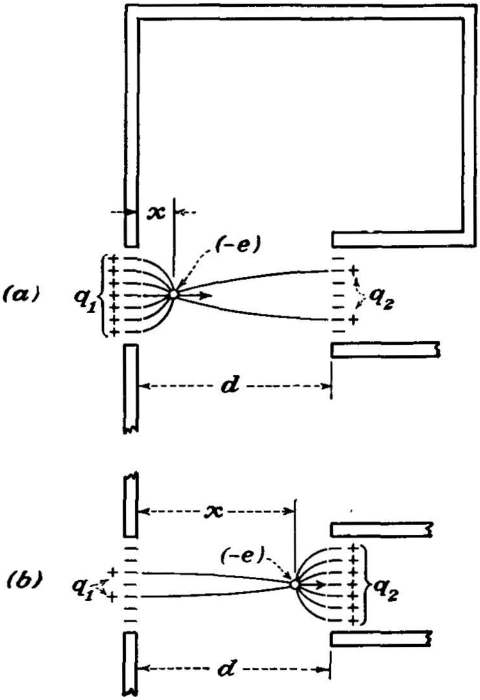


FIG. 17.3.—Charges induced by an electron moving between resonator grids.

grids is d and the distance from the first grid to the electron is x, as shown in Fig. 17.3. The charge induced on the first grid is

$$q_1 = e\left(1 - \frac{x}{d}\right) \tag{17.18}$$

Since the induced charge results from electric-flux lines of the electron terminating on the grids, it must be true that

$$q_1 + q_2 = e \tag{17.19}$$

which it obviously does, as may be seen from the previous two equations.

Since the electron will move between the grids with a nearly constant velocity, the charge induced on the second grid will increase uniformly with time, as shown in Fig. 17.4. It was also shown in the previous chapter that the current associated with the transfer of induced charge from one plate to the other has the value

$$i(t) = \frac{ev}{d} \tag{17.20}$$

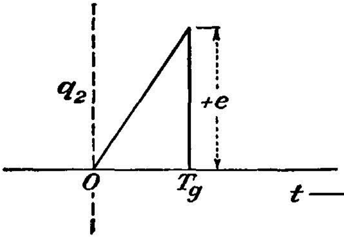


FIG. 17.4.—Charge induced by an electron moving between resonator grids as a function of time.

This same value is obtained if the current is defined as $\frac{dq_2}{dt}$ and the value of this derivative is obtained from Eq. (17.17). The current given in Eq. (17.20) represents a current flowing from the first to the second grid since the associated charge transferred is positive. Curves of $i(t)$ as a function of time are given in Fig. 17.5, in which the time required for the electron to move between grids is represented by T_g. As far as current production goes, the result is the same as though an electron initially at zero velocity suddenly acquired a velocity v and traveled to the second grid, where it was stopped. The current form of Fig. 17.5 is also seen to be that determined by the slope of the charge function of Fig. 17.4. The *shape* of the induced-current pulse is independent of the intergrid transit time and the voltage between grids, provided that this is not excessively high. Thus a slow electron will induce a rectangular pulse of current that is relatively small in magnitude but long in duration. A

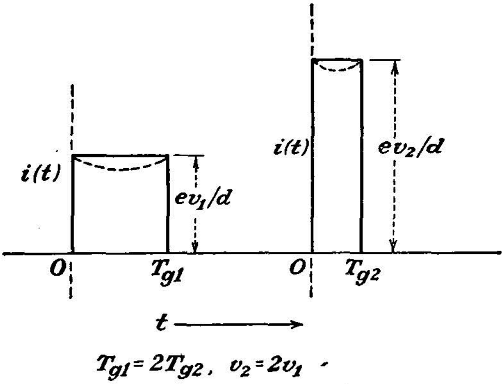


FIG. 17.5.—Induced current resulting from passage of an electron between resonator grids.

fast electron will induce a rectangular pulse of current that is large in magnitude but short in duration. The area under the induced-current pulse is numerically equal to the charge of the electron and hence must be the same for electrons of any speed.

Since the electron stream has a current density that is periodic with time in a period T_1, there will be a fundamental component of resonator current associated with it. If one electron arrives at the same point in every cycle, it will produce rectangular pulses such as those of Fig. 17.5, which are periodic in a time T_1 corresponding to the r-f cycle. The fundamental component of this current is given by Fourier series analysis as

$$I_{r1} = \frac{2e}{T_1} \frac{\sin\left(\dfrac{\pi T_g}{T_1}\right)}{\dfrac{\pi T_g}{T_1}} \tag{17.21}$$

where I_{r1} is the fundamental component of current flowing through the resonator and the other symbols have their previous significance. Let

$$A = \frac{\sin\left(\dfrac{\pi T_g}{T_1}\right)}{\dfrac{\pi T_g}{T_1}} \tag{17.22}$$

$$A = \frac{\sin\left(\dfrac{\theta_g}{2}\right)}{\dfrac{\theta_g}{2}} \qquad (17.23$$

be the ratio of the fundamental component of current for a finite transit angle θ_g to that for a zero transit angle for the case of a pulse created by the passage of a single electron each cycle between the grids, where θ_g is the intergrid *transit angle* of the electron in radians, 2π radians corresponding to the period T_1 of the radio frequency considered. The factor A is the function encountered in Eq. (15.89) and plotted in Fig. 15.40 for the ratio of d-c to r-f deflection sensitivity of electrostatic-deflection plates in a cathode-ray tube and will not be replotted here. It has a maximum value of unity for zero transit angle and first falls to zero for a transit angle of 2π radians.

The power delivered to the resonator by the periodic transit of a single electron when the resonator is tuned to resonance at the frequency corresponding to the period T_1 is

$$P = \frac{I_{r1}V_1}{2} \qquad \text{watts} \tag{17.24}$$

when the voltage gradient is in a direction opposite to that of the electron flow, *i.e.*, when the voltage exerts a force on the electron in the direction opposite to its motion, and for which condition the fundamental component of induced resonator current and the resonator voltage are 180 deg out of phase. In Eq. (17.24) the values of current and voltage are peak rather than rms. The above value of power has the value

$$P = \frac{eV_1}{T_1} A \tag{17.25}$$

When the grid transit angle is negligible, the factor A is unity and the power supplied to the resonator is $\frac{eV_1}{T_1}$. It is seen that the factor A is therefore one which measures the efficiency of energy transfer; as such, it will be extensively used in subsequent analysis. It is frequently referred to as the *beam coupling coefficient.* It should be noted that the value of A given in Eq. (17.22) is only a first-order approximation which has assumed that the velocity of the electron has not changed in moving between the grids. Actually, the velocity of the electron will change as energy is extracted from it.[1] A more rigorous analysis leads to the same first-order results as those given above.[2]

In an actual tube a fairly continuous stream of electrons passes through the resonator grids. Each electron of this stream induces a rectangular pulse of current that flows through the resonator. The resonator current will therefore have the same form as the beam current as a function of time except that the magnitude of any component will be reduced by the factor A computed for the corresponding frequency.

17.4. First-order Bunching Theory. The general picture of the bunching principle has been given in the first section of this chapter. It now remains to give a quantitative analysis of the effects associated

---

[1] Actually, the velocity of an electron while crossing the resonator gap will be a constant plus a sinusoidal variation in accordance with

$$v = v_0 \left[1 + \frac{T_1 V_1}{4\pi T_g V_0} (\sin \omega t_x - \sin \omega t) \right]$$

where t_x is the time at which the electron enters the gap, the gap voltage being assumed to be $V_1 \cos \omega t$, T_1 is the period of the r-f gap voltage, and T_g is the gap transit time of an unmodulated electron. Maximum energy will be extracted from the electron when it enters the gap at a time $\frac{T_g}{2}$ before the negative peak of the gap voltage. Under these conditions the induced current will have the form shown by the dotted curves of Fig. 17.5.

[2] BLACK, L. J., and L. P. MORTON, Current and Power in Velocity Modulated Tubes, *Proc. I.R.E.*, vol. 32, pp. 477–482, August, 1944.

with this principle. The bunching principle was studied by the Heil brothers, but their work was confined to numerical and graphical computations on certain special kinds of operation.[1] The first satisfactory analysis of bunching was made by Webster, whose work has formed the basis for virtually all the subsequent work in this field.[2,3] Webster's attack on the subject has been considerably enlarged by Hansen, Hahn, and Feenberg in this country, by Benham, Hartree, Petrie, Strachey, and Wallis in England, and by Hollman, Brüche, and Recknagel in Germany.

The general picture of the bunching action of a set of resonator grids is well illustrated by a distance-time diagram (attributed to L. M. Appelgate). Let the problem under consideration be formulated as follows: A beam of parallel electrons which have been accelerated through a potential of V_0 volts is passed through the grids of a resonator across which there appears a voltage $V_1 \sin \omega t$. Let the resulting electric field be parallel to the electron motion. Those electrons which pass through the resonator gap at the time the alternating voltage has its maximum value will emerge with an energy corresponding to $V_0 + V_1$ volts if the grid transit angle is sufficiently small. More exactly, they will emerge with an energy corresponding to $V_0 + AV_1$ if the grid transit angle is appreciable. This occurs because the beam coupling coefficient A given by Eq. (17.22) applies whether energy is transferred from the electron to the resonator, or vice versa. Electrons passing through the grids when the r-f voltage opposes the electron motion will emerge with an energy corresponding to $V_0 - AV_1$ volts. In general, they will emerge with an energy corresponding to $V_0 + AV_1 \sin \omega t_a$, where t_a is the time at which the electron passes the midplane of the resonator gap

Since the velocity of an electron is proportional to the square root of the voltage through which it has been accelerated, the velocity with which an electron emerges from the first, or bunching, resonator of a two-resonator klystron will be

$$v_a = v_0 \sqrt{1 + \frac{AV_1}{V_0} \sin \omega t} \tag{17.26}$$

In all subsequent work the numerical subscripts will be associated with the corresponding frequency components; thus v_0 is the d-c component of velocity, and V_1 is the fundamental r-f component of voltage. The letter subscript a will refer to the first-resonator gap transit, and the

---

[1] HEIL and HEIL, *op. cit.*

[2] WEBSTER, D. L., Cathode-ray Bunching, *Jour. Appl. Phys.*, vol. 7, pp. 501–508, July, 1939.

[3] WEBSTER, D. L., Theory of Klystron Oscillations, *Jour. Appl. Phys.*, vol. 10, pp. 864–872, December, 1939.

letter subscript b will refer to the second-resonator gap transit. Let the symbol α be used to designate the *excitation-voltage ratio* $\frac{V_1}{V_0}$, which will ordinarily be less than unity. The product $A\alpha$ is known as the *depth of modulation,* since it is the ratio of the peak amplitude of velocity modulation in volts to the beam voltage. Then if α is small, say less than 0.2, the radical of Eq. (17.26) is represented within a few per cent by the first two terms of its binomial-series expansion,

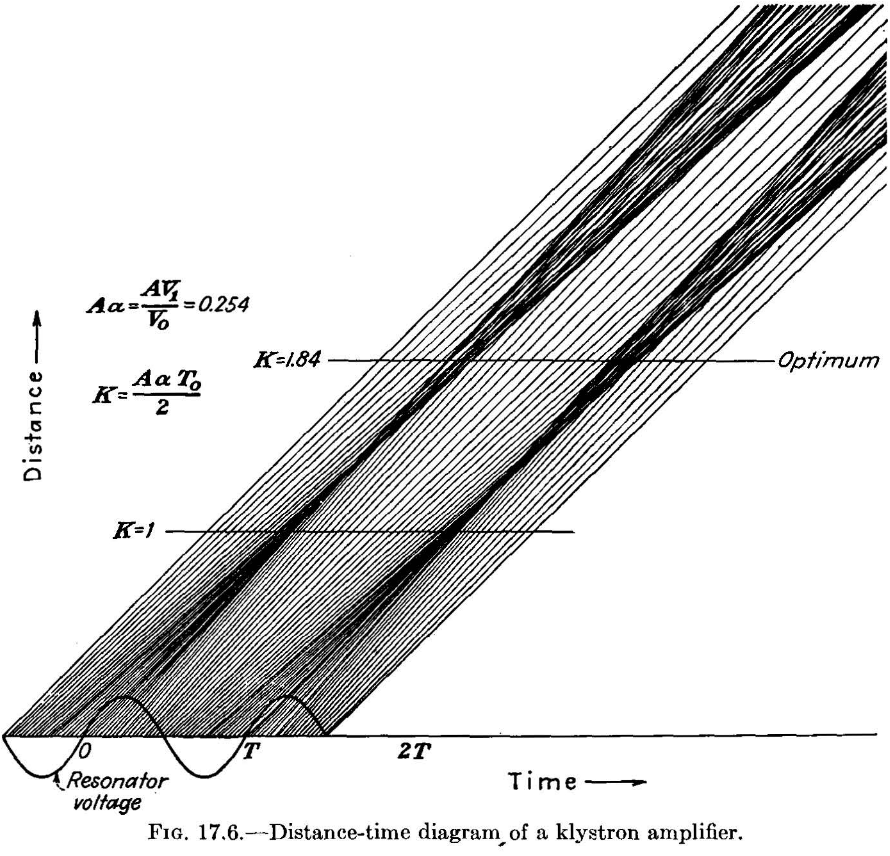


FIG. 17.6.—Distance-time diagram of a klystron amplifier.

$$v_a \cong v_0 \left(1 + \frac{A\alpha}{2} \sin \omega t\right) \tag{17.27}$$

From this equation it is seen that for a small excitation-voltage ratio the velocities of the electrons emerging from a bunching resonator have a value which is a constant plus a factor which is sinusoidal with time.

It is instructive to make a chart that shows how this variation in first-resonator velocity affects the subsequent grouping of electrons. Such a chart is shown in Fig. 17.6. In this chart, distance is plotted vertically, the time is plotted horizontally. The distance-time representation of any electron moving in the field-free region outside of the bunching-resonator grids will be a straight line whose slope is proportional to its velocity. The horizontal axis of this chart corresponds to zero distance from the first-resonator grid, and electrons are assumed to leave this grid at the uniform rate of 40 electrons per cycle. Shown along the horizontal axis is a sine wave corresponding to the r-f voltage between the resonator grids. Electrons that pass through the grids when this voltage is zero will be represented by lines that have a slope corresponding to the original velocity of the electrons, which has been undisturbed by passage through the resonator at this point of the cycle. All other electrons will have either greater or smaller slopes (velocities) than the undisturbed electrons. Those electrons which pass through the resonator when the r-f voltage is negative, *i.e.*, has its gradient in the direction opposite to the electron velocity, will be slowed down and will have slopes smaller than those of the undisturbed electrons. Correspondingly, those electrons which pass through the resonator when the r-f voltage is positive, *i.e.*, has its gradient in the direction of the electron velocity, will be speeded up and will be represented by lines whose slopes are greater than those of the undisturbed electrons. In the resulting set of lines the density of lines along any horizontal line corresponds to the magnitude of the current as a function of time. The bunching action that results is quite evident from the diagram. *A bunch forms about the electron that passes through the resonator at the instant the r-f voltage is changing from retarding to accelerating.* As the electrons move along the beam in what is commonly called the *drift space,* there is first formed a bunch that is very narrow and has a high current associated with it. Farther down the beam, the bunch becomes wider and has the highest current associated with its edges. The corresponding picture of current as a function of time for any position on the beam is shown in Fig. 17.7. The particular shapes that the bunches of electrons give to the beam current will be demonstrated analytically.

The time it takes any electron to move a certain distance along the beam depends upon the point on the cycle at which it passed through the resonator gap and also upon the magnitude of the gap voltage. For travel a distance l from the first resonator gap,

$$t_b - t_a = \frac{l}{v_a} \qquad (17.28)$$

where t_a is the time at which the electron leaves the first resonator, t_b is the time at which the electron has moved a distance l along the beam, and v_a is the velocity with which the electron leaves the first resonator. Substituting the value of v_a from Eq. (17.27),

$$t_b = t_a + \frac{l}{v_0\left(1 + \dfrac{A\alpha}{2}\sin \omega t_a\right)} \tag{17.29}$$

If the depth of modulation factor α is small compared with unity, the fractional term is closely represented by the first two terms of its series expansion.

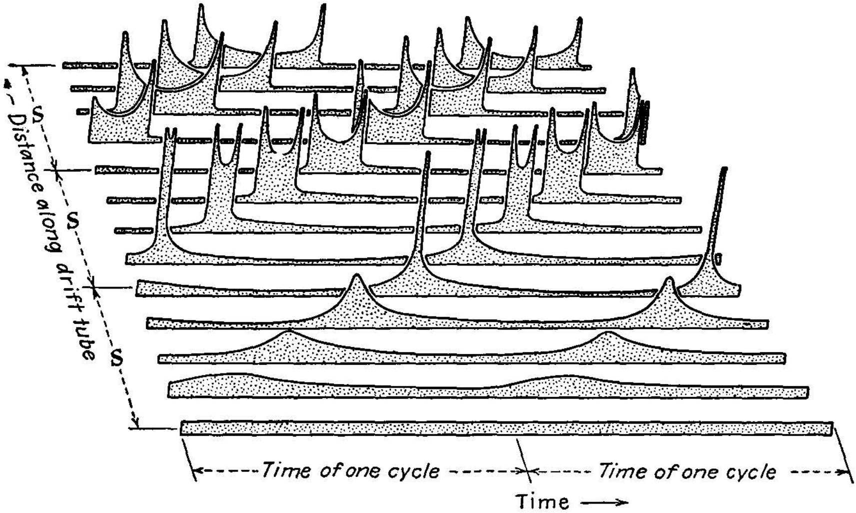


FIG. 17.7.—Current as a function of time for various distances along a bunched beam.

$$t_b \cong t_a + t_0\left(1 - \frac{A\alpha}{2}\sin \omega t_a\right) \tag{17.30}$$

where the transit time of an undisturbed electron, t_0, has been written for $\frac{l}{v_0}$. This expression will be accurate within 5 per cent if $A\alpha$ is less than 0.2. In subsequent analysis it is convenient to deal with transit angles instead of transit times. Transit angle is simply the transit time multiplied by the angular frequency,

$$\tau = \omega t \tag{17.31}$$

where τ is the symbol that will be used for transit angle. Other times as well as the transit time are also conveniently represented by the corresponding angle found by multiplying the time by the angular frequency. Accordingly, Eq. (17.30) may be rewritten

$$\tau_b = \tau_a + \tau_0 - \frac{A\alpha\tau_0}{2} \sin \tau_a \tag{17.32}$$

The factor $\frac{A\alpha\tau_0}{2}$ occurs so frequently in subsequent work that it will be

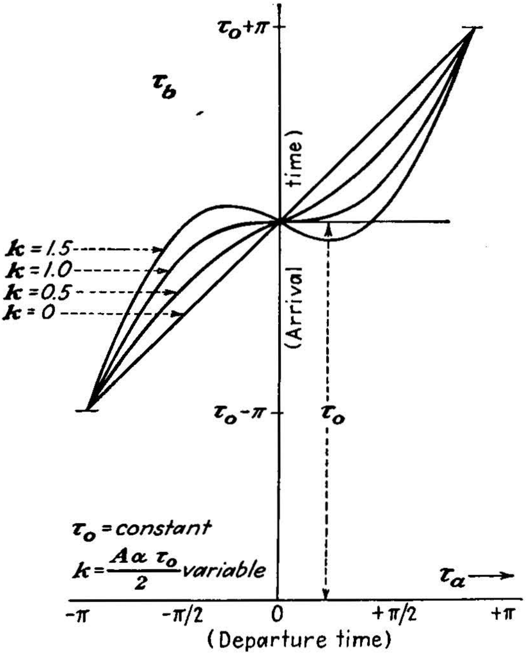


FIG. 17.8.—Electron arrival time as a function of departure time in a bunched beam.

designated by the symbol k and called the *bunching parameter*. With this notation,

$$k = \frac{A\alpha\tau_0}{2} \tag{17.33}$$

and

$$\tau_b = \tau_a + \tau_0 - k \sin \tau_a \tag{17.34}$$

This equation gives the arrival angle τ_b with respect to travel of a distance l in terms of the departure angle τ_a and the bunching parameter k. It is instructive to plot some curves of arrival time in terms of departure time. This is done in Fig. 17.8, in which there are shown curves of τ_b as a function of τ_a and k for values of the latter of 0, 0.5, 1.0, and 1.5.

The bunching parameter k has its value determined by half the product of the beam coupling coefficient A, the excitation-voltage ratio α, and the d-c transit angle τ_0. To discuss the curves of Fig. 17.8, let it be considered that the value of k is varied by increasing the bunching voltage V_1. For a value of the bunching parameter of zero, *i.e.*, no bunching, the arrival angle (time) is a straight-line function of the departure angle (time). As the bunching voltage is raised from zero, the arrival angle as a function of departure angle will be a straight-line function with a superimposed sinusoidal variation whose phase is such that electrons leaving the resonator gap slightly before the reference departure time of zero will have a greater transit time than for no-bunching voltage. Likewise, electrons leaving after the reference departure time of zero will have a smaller transit time than for no-bunching voltage. Those conditions are evident for the value of k equal to $\frac{1}{2}$. For a value of k equal to unity the properties observed above still hold but are accentuated to the point where the slope of the curve of arrival angle as a function of departure angle has a zero value at the reference departure angle of zero. It will be seen later that this has a special significance. Up to a value of k equal to unity the arrival angle is a single-valued function of departure angle, and vice versa. The point for which k equals unity is marked on the distance-time diagram of Fig. 17.6. At this value of the bunching parameter there is evident a strong bunching action. At this value of k the electrons that left just before and after the electron leaving at time zero arrive together.

As the bunching parameter is increased still further, the curve of Fig. 17.8 exhibits a negative slope at the departure time zero, and at this point the departure time is a triple-valued function of arrival time. Furthermore, it will be noted that the electrons near the center of the bunch in the distance-time diagram have crossed, and over an appreciable region it will be true that electrons leaving *after* a time zero arrive *before* electrons which have left earlier, and vice versa. For still larger values of k this property continues to hold. It should be noted that although the departure time is a triple-valued function of the arrival time, the arrival time is always a single-valued function of the departure time. This is to say that, if the arrival time near the center of the bunch for k greater than unity is specified, there will be three different electrons which have left at different times in the vicinity of zero arriving simultaneously at this time. On the other hand, each departure time has a single value of arrival time associated with it.

To find the current associated with the electron bunches it must first be observed that the principle of conservation of charge applies to any corresponding departure- and arrival-time intervals. The electron stream can always be broken up into increments such that all electrons which depart from the bunching resonator between two particular electrons m and n will arrive at a distance l between the times at which these two electrons arrive. Mathematically, this is written

$$|\, dq_b \,| = |\, dq_a \,| \tag{17.35}$$

or

$$|\, I_b \, dt_b \,| = |\, I_a \, dt_a \,| \tag{17.36}$$

Only the magnitudes of the charge increments are of interest, for it is these which will determine the current. Even though electrons may arrive in a reverse order from that in which they left the bunching resonator, their effect in producing output-resonator current is the same since they are traveling through the output resonator in the same direction. From Eq. (17.36) the current at a distance l along the beam is related to the current at a distance zero by

$$I_b = I_a \left| \frac{dt_a}{dt_b} \right| \tag{17.37}$$

$$I_b = I_0 \left| \frac{dt_a}{dt_b} \right| \tag{17.38}$$

since I_a equals I_0, the direct current through the bunching resonator.

From Eq. (17.34), I_b, the current a distance l along the beam, as a function of t_b may be obtained by making use of the fact that $\frac{dt_a}{dt_b}$ equals $\frac{d\tau_a}{d\tau_b}$, which in turn is equal to $\frac{1}{\frac{d\tau_b}{d\tau_a}}$. This has the value

$$I_b(\tau_a) = I_0 \left| \frac{1}{1 - k \cos \tau_a} \right| \tag{17.39}$$

Curves of I_b as a function of τ_a have no great significance. Curves of I_b as a function of τ_b as determined by invoking Eq. (17.34) are shown in Fig. 17.9 (top) for values of k of 0, 0.5, 1.0, and 1.50. For k equal to zero the current is constant. This corresponds to an undisturbed beam. For k equal to ½ a current pulse is seen to begin to form. For k equal to unity the current exhibits an infinite peak corresponding to the simultaneous arrival of several electrons. For k equal to 1½ the curve is double peaked. Infinite current peaks will appear at points corresponding to arrival times for which the slope of arrival time as a function of departure time shown in Fig. 17.8 is zero. As k is still further increased, the double peaks will spread farther apart and the magnitude of the current midway between them will decrease, as shown in Fig. 17.7.

The area under any of the curves shown in Fig. 17.9 is the same, regardless of the value of k, since the current distribution resulting from the bunching action always involves the same number of electrons per cycle. The distance s shown in Fig. 17.7 corresponds to a value of k of unity for which a single infinite peak of current first appears. It has a value, which may be obtained from Eq. (17.33), of

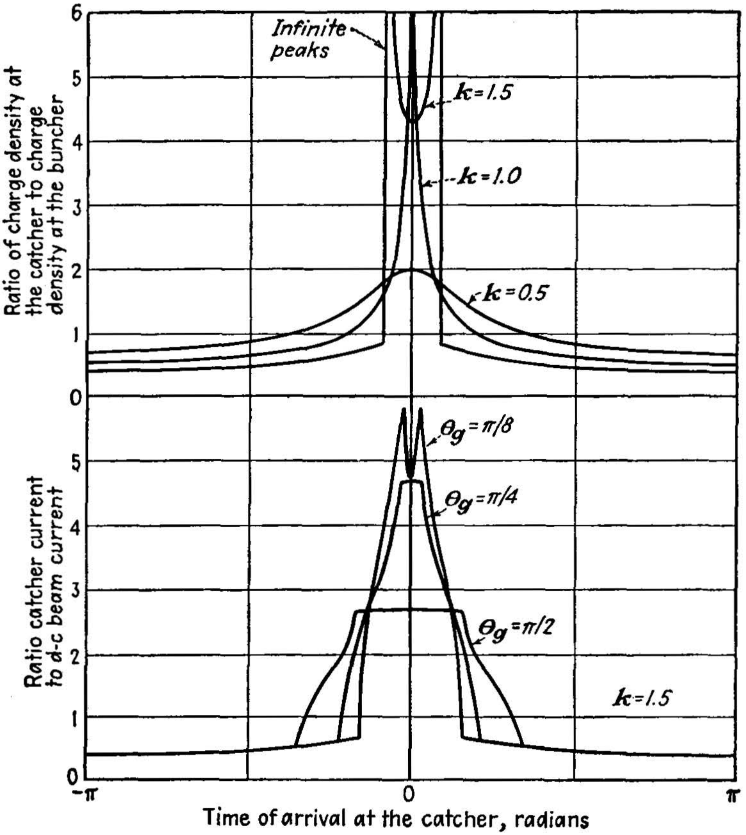


FIG. 17.9.—*a* (top) Current as a function of time along a bunched beam, for different degrees of bunching. *b* (bottom) Induced resonator current as a function of time for different intergrid transit angles. (*After Black and Morton.*)

$$s = \frac{2v_0V_0}{AV_1\omega} \tag{17.40}$$

The curves shown in Fig. 17.9*a* are curves of beam current as a function of arrival time. The current induced in a catcher resonator by the passage of the beam will not have exactly the same form because of the finite time required for electrons to pass between the resonator grids. This finite time will have the effect of integrating the current at different points on the beam between the resonator grids. The integrating effect may be simulated by making a mask with a vertical slit of width corresponding to the intergrid transit angle and sliding this along the curves of Fig. 17.9*a*. Actual resonator currents will be proportional to the area under the beam-current curve revealed through the slit in the scanning mask. Curves of actual induced resonator current obtained by this method are shown in Fig. 17.9*b*. Here it is seen that the infinite peaks do not appear in the actual resonator current and that as the intergrid transit angle is increased the resonator-current pulse becomes less sharp.[1] Examination of the fundamental components of current corresponding to each of the curves of Fig. 17.9*b* will reveal that they are in the ratio of the corresponding A factors as given by Eq. (17.23) to the fundamental components of the corresponding curves of Fig. 17.9*a*.

Several of the parameters which have been used in the preceding analysis are used frequently enough so that it is convenient to have charts giving their magnitude. One of these factors is the d-c transit angle corresponding to a distance l. This has the value

$$\tau_0 = \frac{\omega l}{v_0} = \frac{1{,}000\pi l}{\sqrt{V_0}\,\lambda} \tag{17.41}$$

A chart of this factor as a function of the variables upon which it depends is given in Fig. 17.10. Another factor of importance is the bunching parameter k given by Eq. (17.33). A chart of this factor as a function of the variables upon which it depends is given in Fig. 17.11.

The curves of beam current shown in Fig. 17.9 give the bunched beam current as a function of time. To find out how this current may be used it is necessary to determine its various components, of which the fundamental component is the most important. The fundamental component of the periodic beam current for any degree of bunching may be determined from the Fourier series coefficient formula

$$I_{b1} = \frac{1}{\pi} \int_{-\pi}^{\pi} I_b \cos \tau_b \, d\tau_b \tag{17.42}$$

where I_{b1} is the fundamental component of I_b whose value is given by Eq. (17.39). Since the curves of I_b are symmetrical about the bunch center, the resultant terms in its frequency composition will be cosine terms if the current is arbitrarily centered at the bunch center by neglect-

---

[1] BLACK, L. J., and P. L. MORTON, Current and Power in Velocity-modulation Tubes, *Proc. I.R.E.*, vol. 32, pp. 477–482, August, 1944.

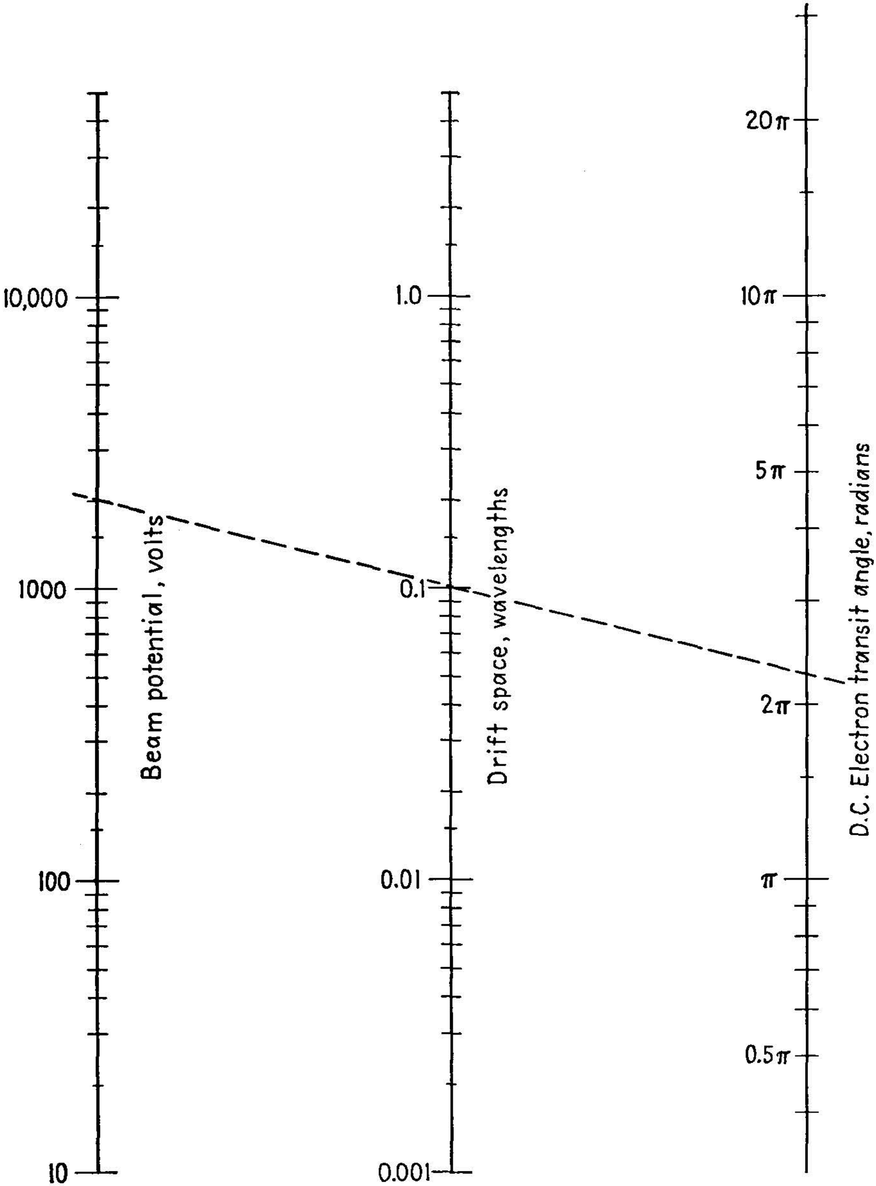


FIG. 17.10.—Nomographic chart giving transit angle as a function of drift distance and beam voltage.

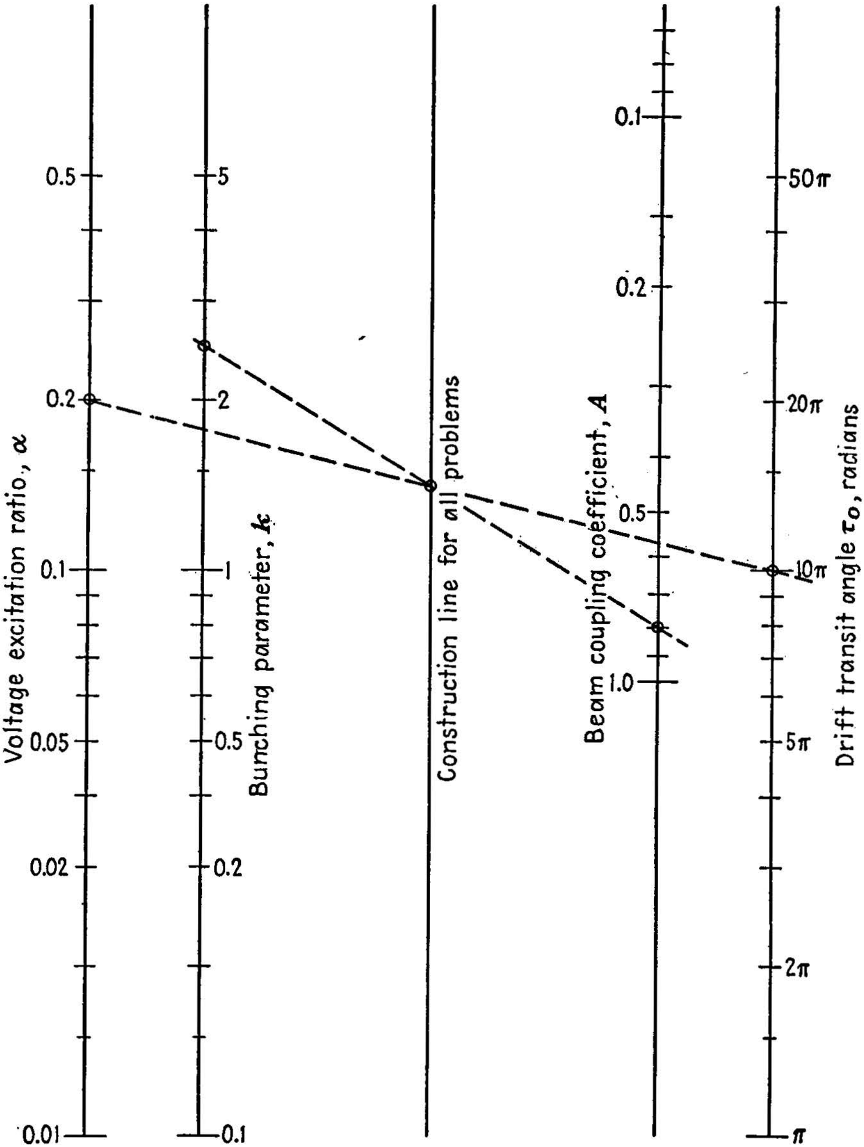


FIG. 17.11.—Nomographic chart giving bunching parameter as a function of beam coupling coefficient, excitation-voltage ratio, and transit angle.

ing the d-c transit-angle term. This amounts to neglecting the phase factor, but this can be put in by inspection later. The fundamental component indicated by Eq. (17.42) is not possible of determination in the form given, for I_b is given as a function of τ_a, whereas the integral is in terms of τ_b. To follow the elegant method proposed by Webster, the integral of Eq. (17.42) can be placed entirely in terms of τ_a by using the relations of Eqs. (17.34) and (17.38). This method eliminates all the apparent difficulties associated with electron crossovers, for τ_b is a single-valued function of τ_a. When this is done,

$$I_{b1} = \frac{1}{\pi} \int_{-\pi}^{\pi} I_b \cos\,(\tau_a - k \sin \tau_a) \frac{I_0}{I_b}\, d\tau_a \qquad (17.43)$$

which reduces to

$$I_{b1} = \frac{I_0}{\pi} \int_{-\pi}^{\pi} \cos\,(\tau_a - k \sin \tau_a)\, d\tau_a \qquad (17.44)$$

in which it is seen that the τ_0 term of τ_b has been dropped. When the integrand is expanded, the integral assumes the form

$$I_{b1} = \frac{I_0}{\pi} \int_{-\pi}^{\pi} [\cos \tau_a \cos\,(k \sin \tau_a) + \sin \tau_a \sin\,(k \sin \tau_a)]\, d\tau_a \qquad (17.45)$$

This is a somewhat formidable integral involving cosines and sines of a sine function. Such terms are encountered in frequency-modulation studies where the frequency of a wave varies periodically with time. In the frequency-modulation problem it is found that terms such as the above correspond to a carrier and a doubly infinite set of side bands whose magnitude is expressed in terms of Bessel functions. The same situation applies here. Each term of the integrand contains an infinite number of terms according to the relations

$$\cos\,(k \sin x) = J_0(k) + 2[J_2(k) \cos 2x + J_4(k) \cos 4x + \cdots] \qquad (17.46)$$

and

$$\sin\,(k \sin x) = 2[J_1(k) \sin x + J_3(k) \sin 3x + \cdots]^* \qquad (17.47)$$

If the above series are substituted into the integrand of Eq. (17.45), the integral is readily evaluated term by term, all but the $\sin^2 \tau_a$ term yielding zero. The result is

$$I_{b1} = 2I_0J_1(k) \qquad (17.48)$$

which is the most important equation in the first-order bunching theory.

---

* These relations are developed in WOODS, F. S., "Advanced Calculus," p. 281, Ginn, Boston, 1932.

If a curve of I_{b1} as a function of k be plotted, the form shown in Fig. 17.12 results.[1] The ratio of the fundamental component of beam current to the d-c beam current is simply a curve of the first-order Bessel function multiplied by 2. This curve starts out as a straight line for small values of k of the form

$$\frac{I_{b1}}{I_0} \cong k \qquad \text{(for small } k\text{)} \tag{17.49}$$

The maximum value of the current ratio occurs for a value of k equal to 1.84 and is equal to 1.16. The curve falls to zero for a value of k equal to 3.83. The significance of the maximum value is that the fundamental component of current will have its maximum value for k equal to 1.84, which is marked on the distance-time diagram of Fig. 17.6. In this figure, k varies directly as distance. The maximum value of fundamental current is obtained not when the bunched beam has its first infinite peak but rather when the double peaks have appeared and spread apart

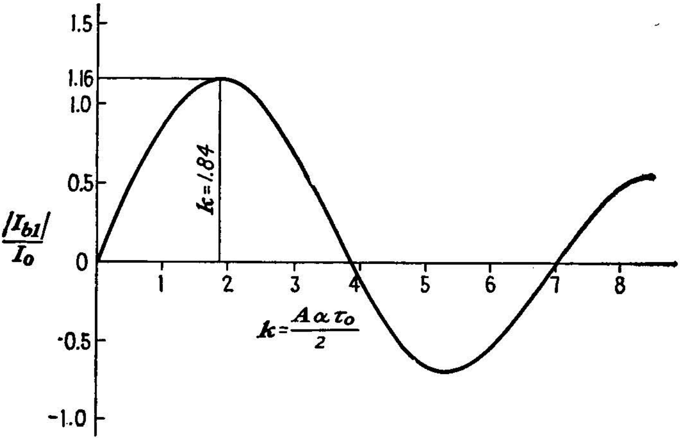


FIG. 17.12.—Fundamental component of current in a bunched beam as a function of bunching parameter. $I_{b1} = 2I_0J_1(k)$.

---

[1] The Bessel functions resemble damped sine waves except that they are not exactly periodic and that the damping is geometric instead of exponential. The order of the Bessel function indicated by the subscript tells what the small-value nature of the function is. For small values of x, $J_n(x) = \frac{x^n}{n!2^n}$, which is to say that the first-order Bessel function starts like a straight line, the second-order function starts like a parabola, etc. The functions soon reverse curvature and have a zero value, after which they approximate damped sine waves and the distinction between the orders appears merely as a phase factor. For a compilation of the principal properties of Bessel functions, see Appendix VI.

appreciably. The maximum value of the fundamental component of current occurs for a value of the bunching parameter for which the area under the product of the beam current as a function of time *multiplied by a cosine wave* is a maximum.

The phase of the fundamental component of current relative to the peak of the bunching voltage may be determined by inspection from the distance-time diagram of Fig. 17.6. Here it is seen that the bunch center, or peak of the fundamental component of current, forms about an electron which leaves the bunching resonator a quarter of a cycle prior to the peak of the bunching voltage. Accordingly, the fundamental component of current lags the bunching voltage by τ_0 *minus* $\pi/2$ radians. The fundamental component of beam current can therefore be written as

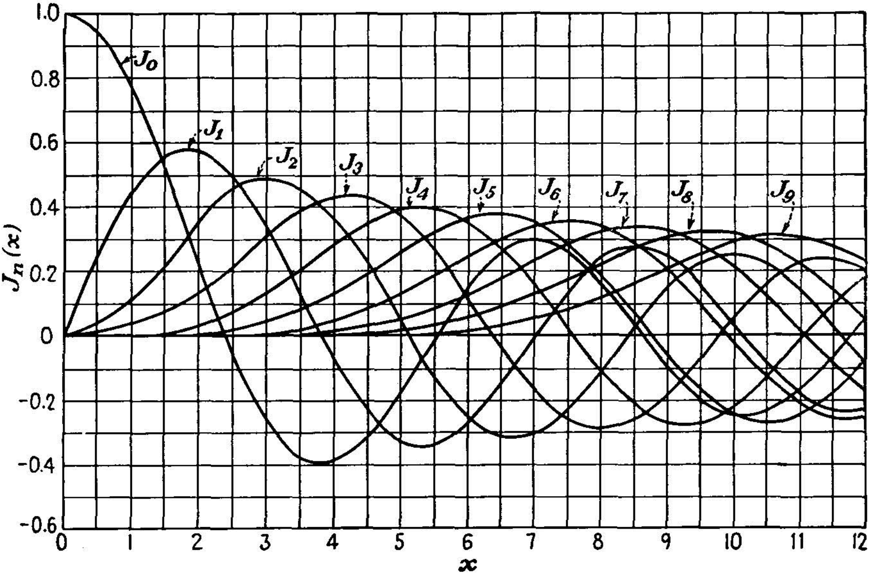


FIG. 17.13.—Curves of the higher order Bessel functions.

$$I_{b1} = 2I_0J_1(k)\epsilon^{-j\left(\tau_0-\frac{\pi}{2}\right)} \tag{17.50}$$

in which the exponential factor is one that has unit magnitude and a phase factor of $-\left(\tau_0 - \dfrac{\pi}{2}\right)$.

If the harmonics of the beam current were evaluated by the method used to determine the fundamental, it would be found that

$$I_{bn} = 2I_0J_n(nk)\epsilon^{-jn\left(\tau_0-\frac{\pi}{2}\right)} \tag{17.51}$$

Curves of the higher-order Bessel functions are shown in Fig. 17.13. Curves of $J_n(nk)$ as a function of k are shown in Fig. 17.14. The peaks of the various harmonics are smaller as the magnitude of the harmonic increases and occur for values of the bunching parameter closer to unity but not less than unity. The locus of the peaks of $J_n(nk)$ is shown dotted in Fig. 17.14. The magnitude of the harmonics as given by Eq. (17.51) drops off very slowly as the order of the harmonic increases, indicating that the klystron should make a good frequency multiplier. This is expected from the shape of the bunched beam current as a function of time, for a current pulse with infinite peaks is rich in harmonics. Maximum values of $J_n(nk)$, along with the corresponding values of k that produce this maximum for different orders of n, are shown in Fig. 17.15. Maximum ratios of harmonic component to d-c component of current fit the empirical function

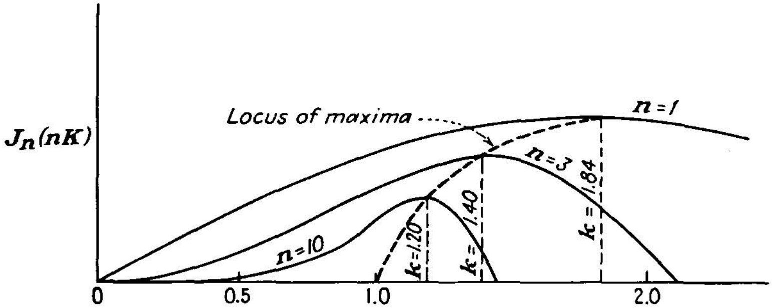


FIG. 17.14.—Curves of $J_n(nk)$ as a function of k.

$$\text{Max}\ \frac{I_{bn}}{I_0} \cong 1.16n^{-0.269} \tag{17.52}$$

within a few per cent out to the tenth harmonic. Values of the bunching parameter for which harmonic currents are maximum are given very closely by[1]

$$k \cong 1 + 0.808n^{-2/3} \qquad (\text{for max } I_{bn}) \tag{17.53}$$

In actual frequency-multiplier tubes the output drops off much more rapidly than is indicated by the relation of Eq. (17.51). This is because of various deficiencies in the first-order bunching theory that have not yet been considered.

17.5. The Klystron Amplifier. Historically, the bunching principle was first applied to produce an oscillator tube. Some amplifier tubes

---

[1] HANSEN, W. W., and J. R. WOODYARD, A New Principle in Directional Antenna Design, *Proc. I.R.E.*, vol. 26, p. 338, March, 1938.

were then built, using the velocity-modulation principle; and later a special kind of klystron oscillator, known as the "reflex-klystron oscillator," was extensively used. In this exposition these three kinds of tubes will be discussed in the order, klystron amplifier, reflex-klystron oscillator, and two-resonator klystron oscillator. This order is used because it makes the explanation of the operation of these tubes much easier. The amplifier is readily described in terms of principles already discussed. The reflex-klystron oscillator is the simplest kind of klystron oscillator to discuss. The methods of analysis used in describing the reflex-klystron oscillator are readily applied to the two-resonator klystron oscillator.

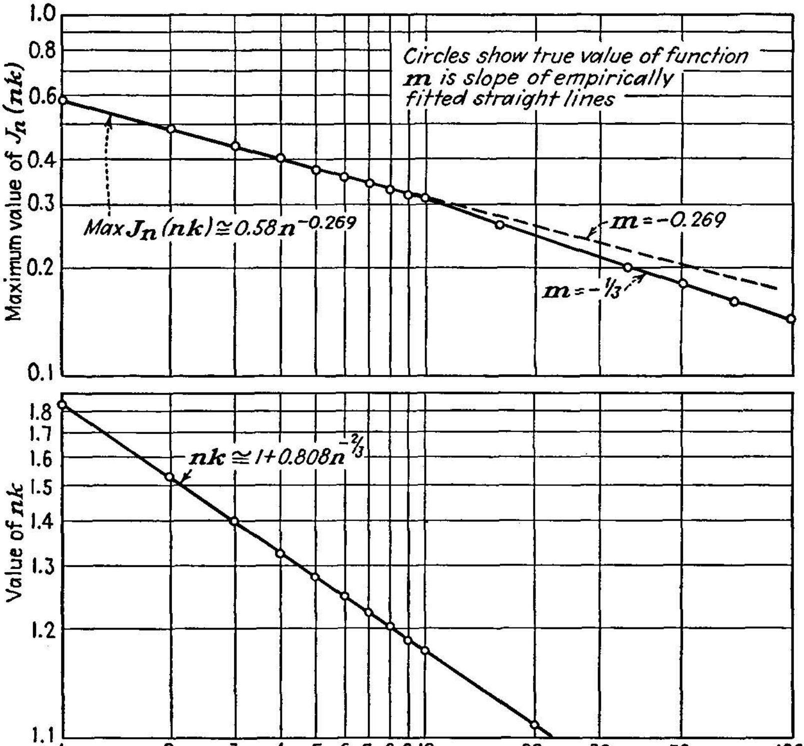


FIG. 17.15.—Maximum values of $J_n(nk)$ as a function of the order n and corresponding argument at which the maximum occurs.

Structure of the Klystron Amplifier. The structure of a klystron amplifier is shown schematically in Fig. 17.16. This type of tube has evacuated reentrant cavity resonators, which are tuned by squeezing the cavities mechanically so that the resonator gap spacings and hence the associated capacities are changed. The tuning range made available by this means is small, being of the order of 10 per cent of the mid-frequency. The tube contains a cathode gun, which may be of the form of the unipotential cathode structures described in the chapter on Cathode-ray Tubes or which may involve a control or focusing electrode. The input and output resonators are usually identical and are placed back to back so that there is a relatively short drift space between the resonator gaps. The length of this gap is moderately critical. If it is too short, there will not be enough time for the electrons to bunch sufficiently. If it is made too long, the bunches are found to deteriorate instead of improve. Power is transferred in and out of the resonators by means of coaxial lines, which terminate in small loops that provide inductive coupling to the magnetic field of the resonator. Electrons that have passed through both resonators impinge on a collector electrode, which returns them to the cathode.

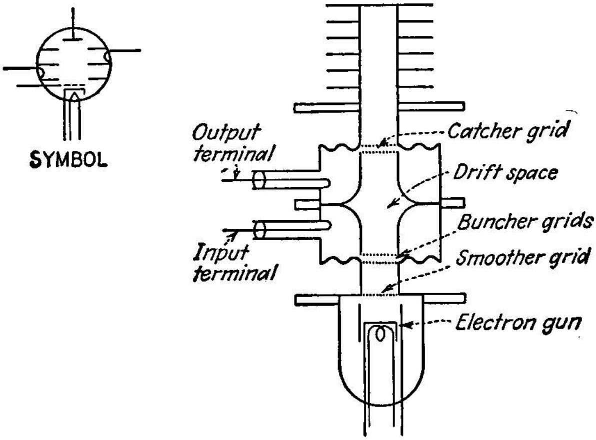


Fig. 17.16.—Structure of the two-resonator klystron amplifier.

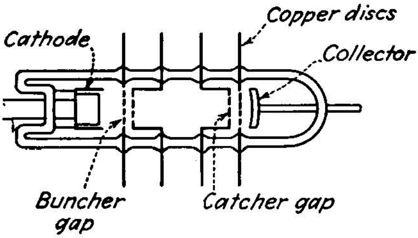


Fig. 17.17.—External-cavity klystron amplifier.

Another form of the klystron amplifier tube is shown in Fig. 17.17. In this form of the tube the resonant cavities are attached externally to the evacuated tube and are tuned by plugs inserted into the cavity.

This form is easier to build but is not as stable as the first type described.

Operation of the Klystron Amplifier. The operation of the klystron amplifier is very simple. Radio-frequency power is fed into the input resonator through the coaxial-line connection, where it produces fields that bunch the electron beam. The bunched electron beam sets up alternating fields in the output resonator. Power is extracted from the output resonator through the coaxial line to a load. The amplifier may be operated at virtually any beam voltage or current. It is necessary only that the input and output resonators be tuned exactly to the frequency of the excitation power.

Output Power of the Klystron Amplifier. If the electron beam passing through the grids of the output resonator has a fundamental component of value I_{b1}, then the output power will be the same as though a current of value AI_{b1} were passed through the resonator, where A is the beam coupling coefficient defined by Eq. (17.23) and it is assumed that the A factor is the same for both buncher and catcher. This follows because the current resulting from the stream of electrons is the simple summation of the currents corresponding to the individual electrons. Accordingly,

$$P_{b1} = \frac{(AI_{b1})^2}{2} R_{sh} \tag{17.54}$$

where P_{b1} is the power delivered to the resonator and R_{sh} is the equivalent shunt resistance of the resonator. In terms of the equivalent output-resonator gap voltage

$$P_{b1} = \frac{AI_{b1}V_{b1}}{2} \tag{17.55}$$

in which V_{b1} is the fundamental component of the second-resonator gap voltage. In both the above equations it must be remembered that I_{b1} is the fundamental component of *beam* current, whereas AI_{b1} is the corresponding effective component of *resonator* current.

Efficiency of the Klystron Amplifier. The efficiency of the klystron amplifier is defined as the ratio of the output power to the input power,

$$\text{Efficiency} = \frac{P_{\text{out}}}{P_{\text{in}}} \tag{17.56}$$

$$\text{Efficiency} = \frac{AI_{b1}V_{b1}}{2I_0V_0} \tag{17.57}$$

not including the power required to bunch the beam. The maximum theoretical efficiency is obtained when each of the factors in Eq. (17.57) assumes its maximum value. The factor A has a maximum value of unity when the gap transit angle is zero. The maximum value of I_{b1}

is $1.16I_0$. The maximum value of V_{b1} is expected to be about V_0; for if this value is exceeded, electrons will be thrown back toward the cathode by the output-resonator voltage and the effective output-resonator resistance will drop very sharply because of the associated increased losses. Upon substituting these maximum values into Eq. (17.57) it is found that the maximum *theoretical* efficiency is 58 per cent. Practical efficiencies are much lower than this because of various second-order bunching effects to be described and are of the order of 20 per cent.

Equivalent Circuit of the Klystron Amplifier. The equivalent circuit of the klystron amplifier is shown in Fig. 17.18. The circuit is the same as that of an ideal pentode r-f amplifier with a delay circuit between the pentode and the output circuit. The hypothetical pentode involved has a virtually infinite plate resistance, for there is no electronic or electromagnetic interaction between the input and output circuits. The delay circuit is one that produces a phase shift of $-\left(\tau_0 - \frac{\pi}{2}\right)$, corresponding to the phase shift between the output current and the input voltage.

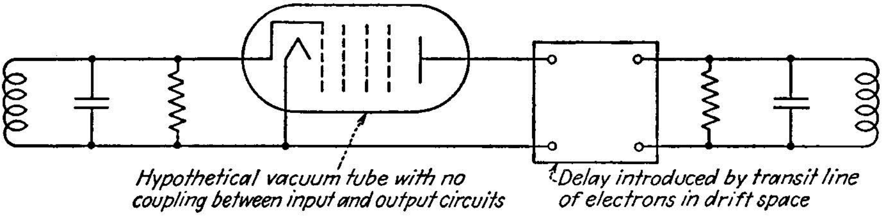


FIG. 17.18.—Equivalent circuit of the klystron amplifier.

Mutual Conductance of the Klystron Amplifier. An equivalent mutual conductance of the klystron amplifier may be defined as the ratio of the induced output current to input voltage V_{a1}. From Eq. (17.50) this may be written as

$$|G_m| = \frac{2I_0}{V_{a1}} AJ_1(k) \tag{17.58}$$

which is readily rearranged by the application of Eq. (17.33) to give

$$|G_m| = G_0 A^2 \tau_0 \frac{J_1(k)}{k} \tag{17.59}$$

where $G_0 \left(= \frac{I_0}{V_0}\right)$ is a factor that may be called the d-c beam conductance. The above assumes that the beam coupling coefficient A is the same for input and output resonators. The mutual conductance is not a constant but rather is a factor that decreases as the magnitude of the excitation voltage increases in accordance with the curve of $\frac{J_1(k)}{k}$. A curve of mutual conductance as a function of the factor k is given in Fig. 17.19. Since, for the zero value of k, $\frac{J_1(k)}{k}$ is ½, the small-signal value of mutual conductance is the maximum and has the value

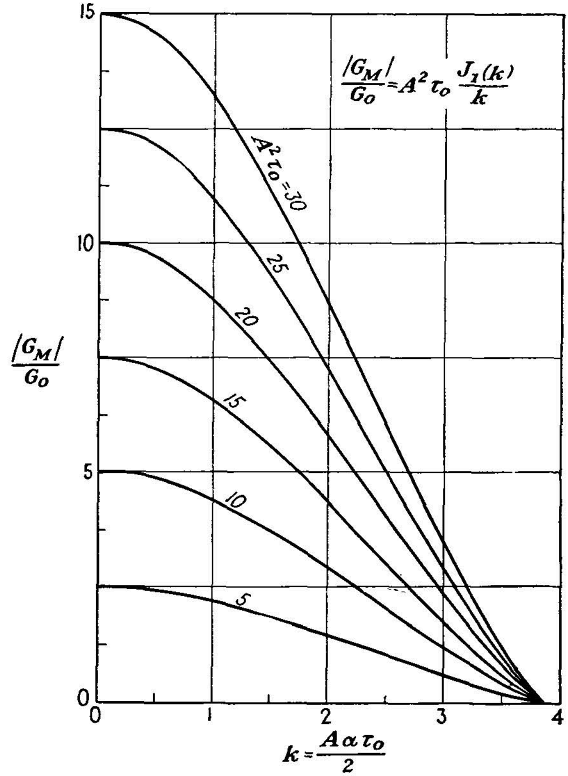


FIG. 17.19.—Transconductance of a klystron amplifier as a function of the bunching parameter.

$$\frac{|G_m|}{G_0} = \frac{A^2\tau_0}{2} \qquad \text{for small signals} \qquad (17.60)$$

The value of mutual conductance for maximum current ($k = 1.84$) is

$$\frac{|G_m|}{G_0} = 0.316A^2\tau_0 \quad \text{for maximum output} \qquad (17.61)$$

The mutual conductance as used above has a phase angle associated with it that is $-\left(\tau_0 - \frac{\pi}{2}\right)$; accordingly, we may speak of the *transadmittance* of the amplifier tube as the product of the mutual conductance and the phase factor

$$\frac{Y_m}{G_0} = A^2\tau_0 \frac{J_1(k)}{k} \epsilon^{-j\left(\tau_0 - \frac{\pi}{2}\right)} \tag{17.62}$$

$$\frac{Y_m}{G_0} = A^2\tau_0 \frac{J_1(k)}{k} (\sin \tau_0 + j \cos \tau_0) \tag{17.63}$$

where Y_m, the transadmittance, is the ratio of the output current to the input voltage in phase and magnitude.

Power Required to Bunch the Beam. In the input resonator of a klystron amplifier the bunching action speeds up electrons over half the cycle and slows them down over the other half. When the intergrid transit angle is small, the average energy of electrons leaving the bunching resonator over a cycle will be nearly equal to the energy with which they enter. However, as the intergrid transit angle increases, the average energy of electrons leaving the resonator will be greater than the entering energy and as a result the bunching resonator must supply power to bunch the beam. Therefore, there is an equivalent resistance that can be attributed to the power required to bunch the beam.

The calculation of the power needed to produce bunching action requires extensive manipulation of second-order effects and will only be indicated here.[1] It is a simple matter to calculate the velocity of any electron passing through the bunching resonator as a function of the point on the cycle at which it enters the resonator and the subsequent time interval. Likewise, the distance-time behavior can be calculated. The resulting expressions give velocity and distance as a function of time. However, it is desired to know the velocity of an electron as it leaves the resonator, which requires that the above expressions be inverted so that exit velocity is given as a function of d-c transit angle. An equation for this relation can be obtained in terms of a series in powers of the d-c transit angle. When this is obtained, the average exit energy can be calculated from the square of the velocity. The difference between the average exit energy and the entrance energy is a measure of the bunching power required. The ratio of the power required to produce bunching to the d-c power required to produce the beam has the form

---

[1] See FEENBERG, E., Notes on Velocity Modulation, *Sperry Gyroscope Laboratories Rept.*, 5521-1043, Chap. I, pp. 41–44.

$$\frac{P_1}{P_0} = \frac{V_{a1}^2}{2V_0^2}\,\frac{1}{2}\,\frac{\sin\left(\frac{\theta_g}{2}\right)}{\frac{\theta_g}{2}}\left[\frac{\sin\left(\frac{\theta_g}{2}\right)}{\frac{\theta_g}{2}} - \cos\left(\frac{\theta_g}{2}\right)\right] \tag{17.64}$$

$$\frac{P_1}{P_0} = \frac{V_{a1}^2}{2V_0^2}F(\theta_g) \tag{17.65}$$

for values of $\frac{V_{a1}}{V_0}$ less than ½, where P_0 equals I_0V_0 and the definition

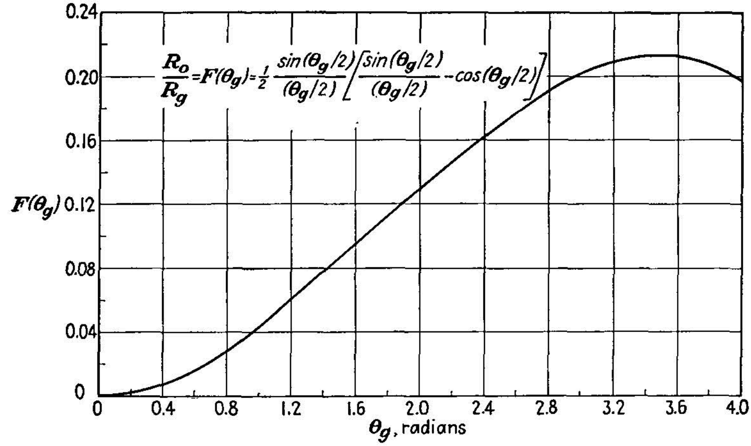


FIG. 17.20.—Theoretical equivalent bunching resistance of a beam as a function of the intergrid transit angle. (*After Feenberg.*)

of $F(\theta_g)$ is apparent. The ratio of equivalent beam resistance to bunching resistance may be defined as

$$\frac{R_0}{R_g} = \frac{V_0^2}{P_0}\,\frac{2P_1}{V_{a1}^2} \tag{17.66}$$

Accordingly,

$$\frac{R_0}{R_g} = F(\theta_g) \tag{17.67}$$

A curve of $\frac{R_0}{R_g}$ as a function of θ_g is given in Fig. 17.20. The justification for defining an equivalent bunching resistance as $\frac{2P_1}{V_{a1}^2}$ is that the power required to produce bunching is proportional to V_{a1}^2 provided that the excitation-voltage ratio is not excessive. This means that the power required to produce bunching is the same as would be consumed by a resistance R_g, as defined in Eq. (17.67), in parallel with the shunt resistance of the resonator. Examination of Fig. 17.20 shows that the equivalent bunching resistance is ten times the beam resistance for an intergrid transit angle of 1.66 radians (95 deg). For transit angles less than a quarter of a cycle the equivalent bunching resistance will be greater than ten times the beam resistance, and for transit angles greater than a quarter of a cycle it will be less than ten times the beam resistance. The power consumed by the equivalent bunching resistance is by no means negligible and will ordinarily be of the same order of magnitude as the ohmic power loss in the resonator itself. Measured values of equivalent bunching resistance range from 20 to 50 per cent of the theoretical values and are not independent of the excitation-voltage ratio.[1]

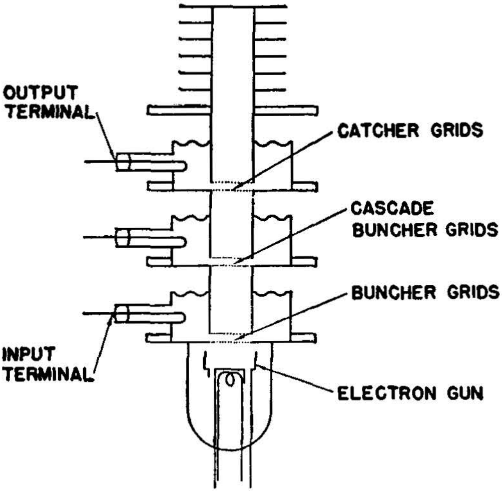


FIG. 17.21.—Structure of the cascade amplifier.

17.6. The Cascade Amplifier. If a three-resonator klystron amplifier be made with the first and third resonators used as input and output resonators, respectively, but the middle resonator be left unloaded and simply tuned to the frequency of the input signal, very large power amplifications are obtained. Such an amplifier has been termed a "cascade amplifier" and has the structure shown schematically in Fig. 17.21. If a small input signal to the first resonator is assumed, there is produced at the second resonator a fundamental component of current that, though small, is appreciable. The second, or cascade, resonator, being unloaded, will have a very high effective resistance, which is determined by the parallel combination of its shunt resistance and its effective

---

[1] HADLEY, C. F., Velocity Distribution of Velocity Modulated Beams, Ph.D. Dissertation, Stanford, 1944.

bunching resistance, as discussed in the previous section. Both these components of resistance can be made quite high. As a result, a large voltage will be produced in the cascade resonator by a small fundamental component of beam current. Since the exciting current that produces the voltage in the second resonator and the resulting current that is produced are 90 deg out of phase, the bunching actions before and after the second resonator can be considered independently to a fair degree of approximation. The ratio of bunching parameters of the first and second resonators relative to the second and third resonators will be proportional to the square of the transit angle between resonators. The ratio of the output power of the third resonator to the input power to the first

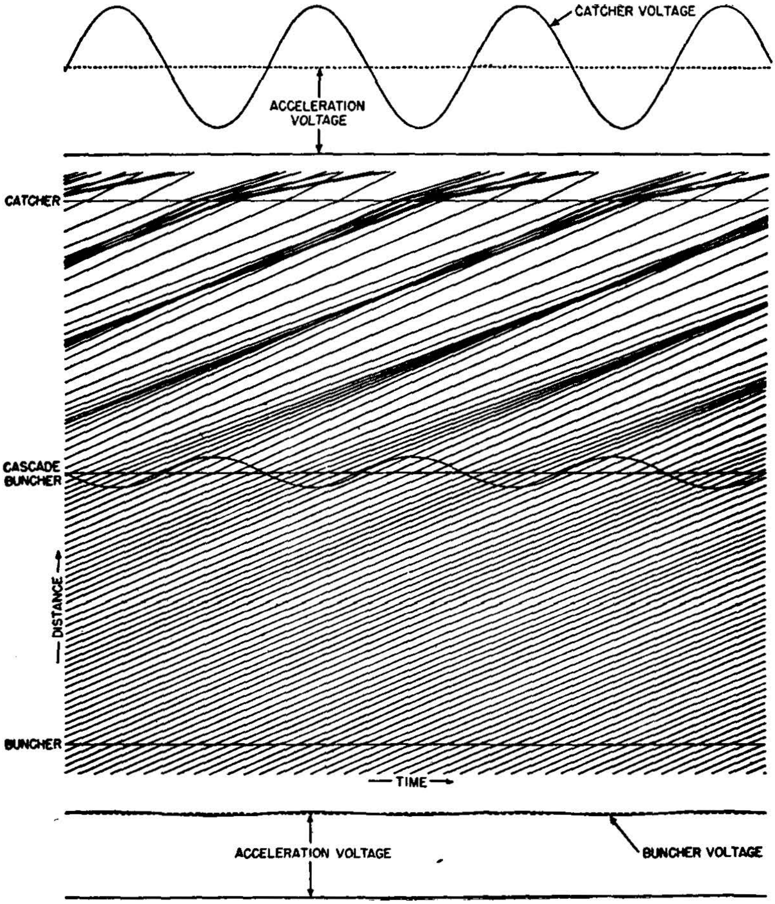


FIG. 17.22.—Distance-time diagram of a cascade amplifier. (*After Harrison.*)

resonator will be proportional to the fourth power of the transit angle between resonators.

A distance-time diagram of a cascade amplifier is shown in Fig. 17.22. In this diagram it is seen that a very small bunching action between the first and second resonators gives rise to an appreciable voltage on the second resonator. This in turn gives rise to a higher degree of bunching. The resultant bunching action is not due to that of the second resonator alone. Those electrons which pass through the second resonator at times when the second-resonator voltage is zero form the center of its bunching action, while those which pass through the second resonator when its voltage is maximum are at the center of the bunch formed by the first resonator. The combined action is much better than could be achieved by a single bunching resonator and approximates that which would result from a single resonator which had a saw-toothed instead of a sinusoidal gap voltage. Maximum theoretical efficiencies are 74 per cent, though actual efficiencies are much less. Cascade-amplifier tubes have given power amplifications of the order of 1,000 to 5,000 times. Tuning of such an amplifier is quite critical since all three resonators must be tuned to exactly the same frequency. Such amplifiers are essentially single-frequency devices. Unfortunately, the internal noise of klystron amplifiers is so high that the improvement in signal-to-noise ratio is much less than the actual power amplification.

17.7. Frequency-multiplier Klystrons. Because the harmonic content of a bunched beam is relatively high, the klystron makes a good microwave frequency multiplier. The frequency-multiplier klystron is similar to the amplifier except that the output resonator is designed to be tuned to a harmonic of the input frequency. To get a tube with an input resonator that tunes to a low frequency it is necessary to use a resonator in the form of a concentric line that is heavily loaded with capacity. The structure of such a tube is shown schematically in Fig. 17.23. Such tubes are critical of excitation and beam voltage. This is because the maximum value of harmonic current for a large frequency-multiplication factor is relatively critical with respect to the bunching parameter, as may be seen from Fig. 17.14. As input excitation is increased for a given beam voltage, a frequency-multiplier tube will pass through its maximum output rather sharply and is easily overdriven. For a given input excitation, the power output as the beam voltage is changed will follow a curve such as is shown in Fig. 17.24. This curve is like the curves of Fig. 17.13 squared, but with the x axis inverted. Maximum theoretical efficiencies are equal to half the ratio of maximum value of harmonic current to d-c beam current as given in Sec. 17.4. Actual efficiencies run about one-tenth of the maximum theoretical efficiencies. Efficiencies are further found to drop off much more rapidly with the order of the harmonic than the inverse one-third power that is expected from the theoretical maximum values of harmonic components of current. Frequency multiplication by a factor of 10 in a single tube is entirely practical in the range of 300 to 10,000 mc and makes possible crystal-controlled microwave signals.

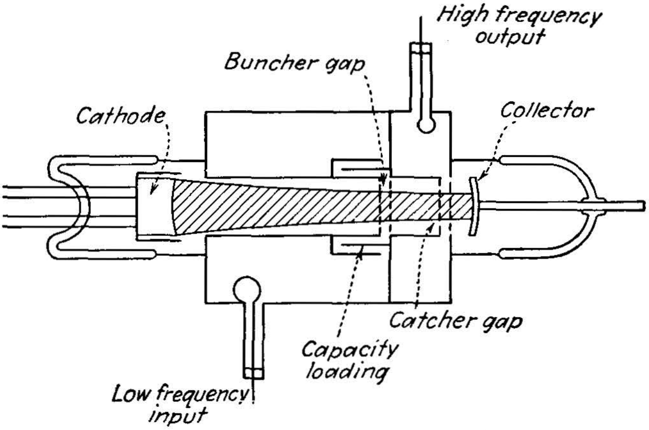


Fig. 17.23.—Structure of the frequency-multiplier klystron.

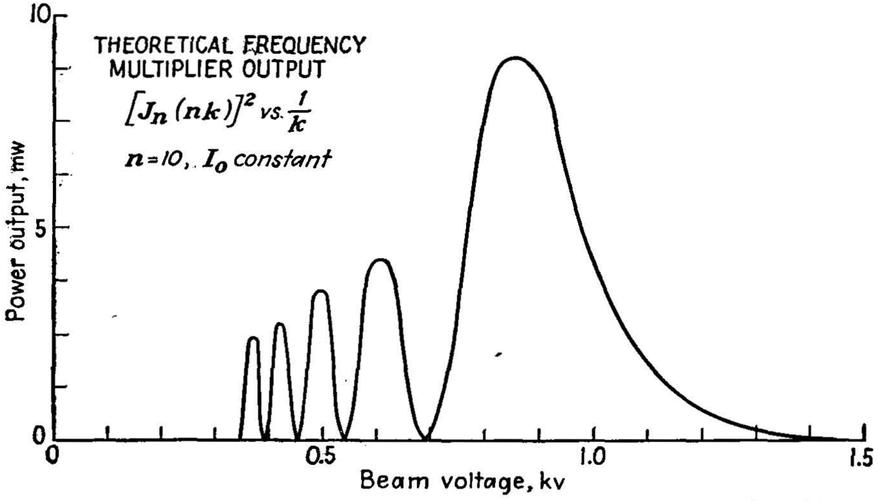


Fig. 17.24.—Power output of a frequency-multiplier klystron as a function of beam voltage.

17.8. Second-order Bunching Effects. The theory of bunching that has been presented up to this point is what may be called the "simple" or "first-order theory." It gives a correct picture of the mechanism, but appreciable departures from it are encountered in actual tubes.

One of the limitations of the first-order theory is that it has entirely neglected the mutual electrostatic repulsion forces between the electrons associated with the electron charge. This is to say that space-charge effects have been neglected. The analytical treatment of such effects becomes sufficiently involved so that only a description of the principal effects will be undertaken here.[1]

Space-charge effects will in general have the effect of reducing the degree of bunching that would exist if there were no space charge. For a beam that is large in diameter the axial repulsion forces of the electrons will be greater than the radial forces except at the edge of the beam. For such a beam, as electrons tend to come together there will build up repulsion forces that oppose the bunching action. Near the center of a bunch there will develop forces that are proportional to the distance of a particular electron from the center of the bunch. The resultant electron action is similar to that observed in mechanical compression problems. Imagine an observer riding along with an electron at the center of a bunch. He observes electrons approaching the center of the bunch in both directions. These electrons will be progressively slowed down as they approach the center of the bunch because the electrostatic repulsion forces will build up. As a result, the individual electrons before and after the center of the bunch will approach to within a given distance of the center of the bunch and will then turn and move away from it. This action is illustrated roughly in the distance-time diagram of Fig. 17.25. Here electrons near the center of the bunch are seen to approach each other and then diverge without crossing.[2–4] To a first order of approximation the velocity of an electron near the center of the beam and the center of the bunch is a constant with a superimposed sinusoidal variation. Webster has termed this action "longitudinal debunching." From the distance-time diagram it is seen that the maximum degree of bunching occurs considerably farther along the beam than the distance corresponding to the formation of the first infinite peak of current in the absence of space charge. Furthermore, the maximum degree of

---

[1] The most complete treatment of bunching theory in all its aspects yet published appears in *Sperry Research Laboratories Rept.* 5221-1043 by E. Feenberg, 1945.

[2] WEBSTER, D. L., Cathode-ray Bunching, *Jour. Appl. Phys.*, vol. **7**, 501–502, July, 1939.

[3] HAHN, W. C., Small Signal Theory of Velocity-modulated Electron Beams, *Gen. Elec. Rev.*, vol. 42, pp. 258–270, June, 1939.

[4] WARE, L. A., Electron Repulsion Effects in a Klystron, *Proc. I.R.E.*, vol. 33, pp. 591–596, September, 1945.

actual bunching will be equal only to that which would occur considerably before the first infinite peak of current in the absence of space charge. Infinite peaks of current are, of course, a physical impossibility, and it is doubtful whether or not even double peaks occur where high beam currents are involved. Since the space-charge repulsion forces near the edge of the beam are less than at the center, the bunches will tend to form sooner near the edge and be more intense. As a result,

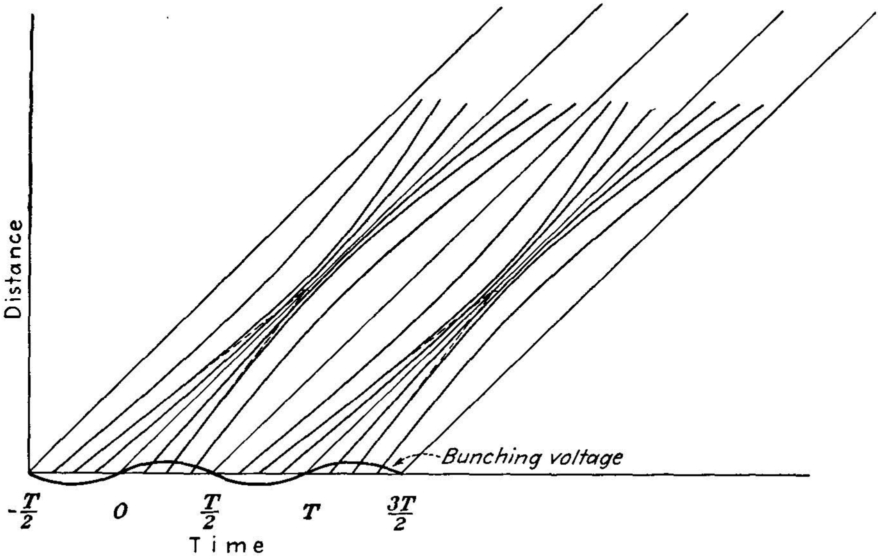


FIG. 17.25.—Distance-time diagram of a high-current-density beam showing space-charge debunching effects.

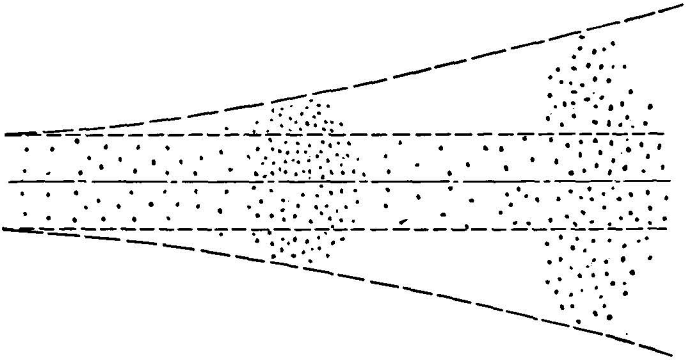


FIG. 17.26.—Picture of a bunched beam showing radial dispersion due to space charge.

the bunches will tend to be crescent-shaped in a plane through the axis of the beam, with their concave side toward the cathode.

In Chap. 15 it was shown that an unbunched beam would tend to spread owing to radial electrostatic repulsion forces. The same action occurs in a bunched beam except that it is accentuated by the bunching action. As bunches tend to form, the space-charge density in that region will increase and the radial expansion will be greatest about a bunch center. Portions of the beam between bunches will have their space-charge density reduced and will not spread so much. The resultant action is that the bunches will tend to form, achieve a certain degree of grouping of electrons, and then literally explode radially. The actual picture of a bunched beam may be expected to look something like Fig. 17.26. As a result of this action, the current density associated with any bunch will first increase as the electrons move along the beam and then decrease.

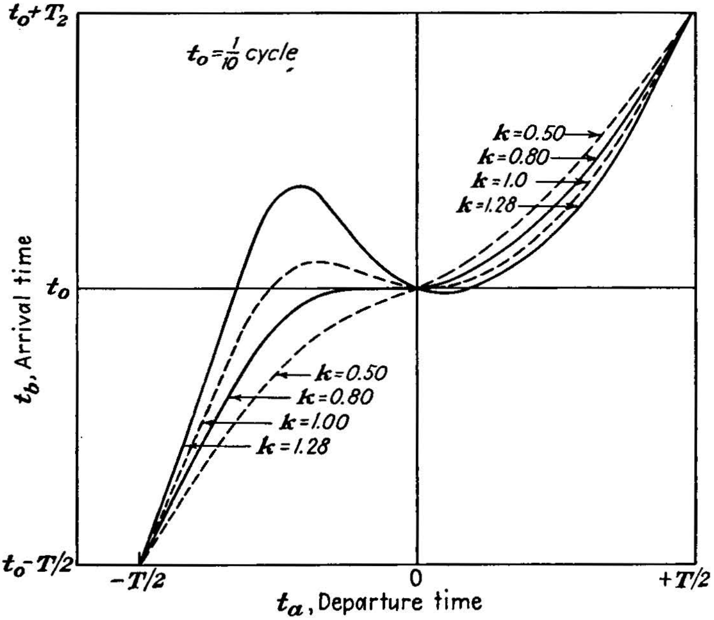


FIG. 17.27.—Electron arrival time as a function of departure time for a bunched beam with a short drift distance and relatively large values of r-f voltage.

The general observations made about space-charge effects above are borne out in the operation of actual tubes. Studies of the velocity distribution of bunched beams made with a special tube incorporating a velocity spectrograph where the collector electrode is normally placed reveal that the actual velocity distribution is quite different from that expected from first-order bunching theory neglecting space charge.[1] Furthermore, it is found that amplifier and oscillator klystrons built with relatively short drift spaces give higher output and efficiency than do those with long ones, this being particularly true of frequency-multiplier tubes. Likewise, the output of almost any klystron amplifier or oscillator can be increased by applying an axial magnetic field even when the cathode design is good.

Since space-charge effects dictate short drift spaces, larger driving voltages are required to produce a given degree of bunching. Accordingly, the limiting value of $\frac{V_1}{V_0}$ of 0.2 assumed in the first-order theory is generally exceeded. In general, the resulting action is the same as before except that the degree of bunching is less than that predicted by the first-order theory. Curves of arrival versus departure time for a very short drift distance but relatively large values of $\frac{V_1}{V_0}$ are shown in Fig. 17.27 (space-charge effects neglected). According to such a set of curves, double peaks occur at smaller values of k than unity, and maximum output results at a value of k appreciably smaller than 1.84. This tendency is frustrated by the space-charge effects.

17.9. The Reflex-klystron Oscillator. The reflex-klystron oscillator is a single-resonator klystron with a reflector electrode, operated below cathode potential and located so that electrons are reversed in direction after a first passage of the resonator and made to return through the same resonator.[2,3] The electron stream is velocity-modulated by its first passage through the resonator gaps, and power is extracted from the bunched beam current upon the second passage of the electrons through the resonator. The structure of some typical 10-cm reflex-klystron oscillators is shown in Fig. 17.28. Tubes are of two types, those in which the resonant cavity is sealed to the tube and evacuated and those in which the resonant cavity is attached externally to the tube. Both types utilize a cathode for the production of the beam. The magnitude of the cathode current is sometimes controlled by a control grid or focusing ring. The entire resonator is operated at the same potential above cathode, and this potential is that through which the electrons are

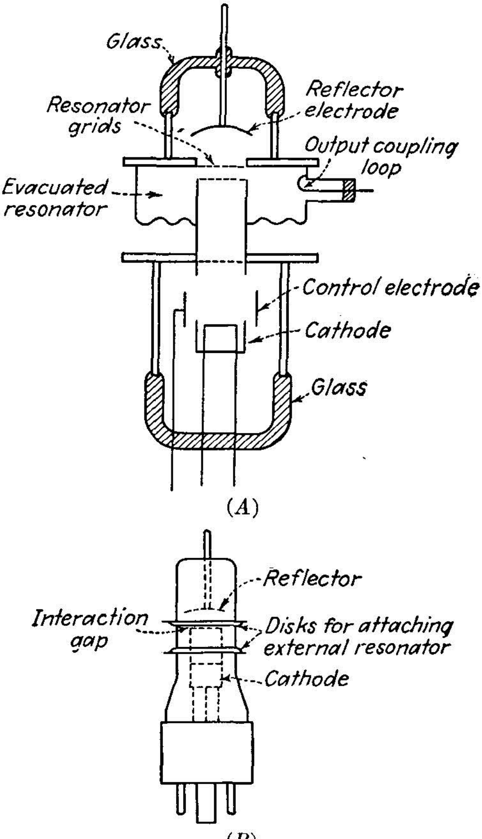
---

[1] HADLEY, *op. cit.*

[2] PIERCE, J. R., Reflex Oscillators, *Proc. I.R.E.*, vol. 33, pp. 112–118, February, 1945.

[3] GINZTON, E. L., and A. E. HARRISON, Reflex-klystron Oscillators, *Proc. I.R.E.*, vol 34, pp. 97–117, March, 1946.

accelerated. Since the reflex-klystron oscillator is seldom used to obtain appreciable power but rather finds its greatest application as a local oscillator tube, large currents are not needed and the resonator gap faces are made of a fine-mesh grid. The tube is usually constructed so that immediately beyond the second resonator grid there is a region of nearly constant potential gradient. The reflector electrode is ordinarily concave toward the resonator so that there is a focusing effect which directs the electrons back toward the center of the resonator grid.



(*A*)

(*B*)

FIG. 17.28.—Structure of the reflex klystron: (*A*) evacuated resonator, Sperry type; (*B*) external cavity required.

Behavior of Electrons in the Reflector Space. Electrons in the reflector space encounter a nearly constant gradient of potential opposing their motion. In all the subsequent analyses it will be assumed that the potential variation in the reflector space is linear with distance and that as a consequence the gradient of potential is indeed constant. As a result of this assumption, the laws of motion of electrons in the reflector space are identical with those of a ball thrown up into the air in the absence of friction. In this mechanical analogy the ball experiences a constant downward force due to gravity just as the electron experiences a constant force directed toward the resonator. The equations of motion for the acceleration, velocity, and position as a function of time of an electron injected into a region of retarding potential gradient of value $\frac{V_0 + V_r}{d}$ with an initial velocity v_0 are

$$a = \frac{dv}{dt} = -\frac{e(V_0 + V_r)}{md} \tag{17.68}$$

$$v = \frac{dx}{dt} = -\frac{et(V_0 + V_r)}{md} + v_0 \tag{17.69}$$

$$x = -\frac{et^2}{2md}(V_0 + V_r) + v_0 t \tag{17.70}$$

where a is acceleration, v is velocity, v_0 is initial velocity, t is time, V_0 is cathode-resonator potential difference, V_r is the potential difference existing between reflector and second resonator grid, which are separated a distance d, e and m are, respectively, the charge and mass of the electron, and x is distance measured from the second resonator grid. It is seen that the acceleration is constant with time, velocity decreases uniformly with time, and distance is a parabolic function of time. Hence, if a distance-time diagram be plotted for electrons in the reflector field, the curves will all be parabolas. The maximum distance to which an unmodulated electron will penetrate the reflector field will be given by

$$x_{\text{max}} = \frac{dV_0}{V_r + V_0} \tag{17.71}$$

since at this distance the potential has changed by an amount V_0, as is apparent from Fig. 17.29, which shows the potential profile of the tube. Hence the distance to which an electron penetrates the field is directly proportional to the initial potential through which the electron has been accelerated. The average velocity of the electron in the reflector field will be half of its initial velocity, and hence the time the electron spends in the reflector field will be

$$t_0 = \frac{4x_{\text{max}}}{v_0} \tag{17.72}$$

Since

$$\frac{v_0}{c} = \frac{\sqrt{V_0}}{506} \tag{17.73}$$

where c is the velocity of light, then

$$t_0 = \frac{2{,}024d\sqrt{V_0}}{c(V_0 + V_r)} \tag{17.74}$$

If both sides of this equation be multiplied by the angular frequency ω and use is made of the relation $c = \lambda f$, where λ is the wave length, then the resulting expression for the d-c *transit angle* in radians spent in the reflector space is

$$\tau_0 = 4{,}048\pi \frac{d}{\lambda} \frac{\sqrt{V_0}}{(V_0 + V_r)} \tag{17.75}$$

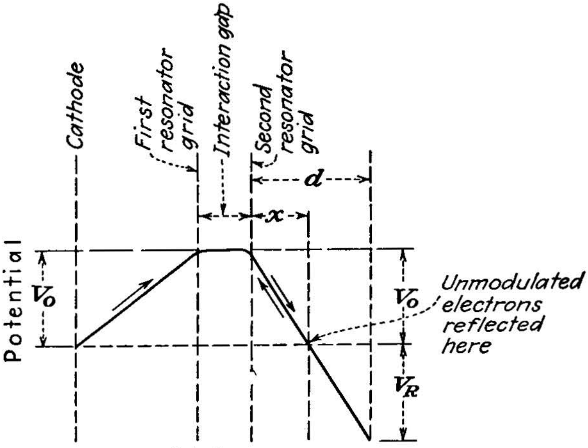


FIG. 17.29.—Potential profiles of a reflex-klystron oscillator.

Distance-time Diagram of a Reflex-klystron Oscillator. Since the distance to which an electron penetrates the reflector field against a constant gradient of potential is proportional to the initial energy and since the law of falling relative to the point at which the electron direction is reversed is the same for *all* electrons regardless of their initial energy, the distance-time curves of electrons entering the reflector space with different velocities will all be parts of the *same* parabola. This makes it relatively easy to construct a distance-time chart by means of a template since the energy of electrons leaving the resonator will be

$$V_a = V_0(1 + A\alpha \sin \omega t) \tag{17.76}$$

where V_a is the voltage equivalent of the electron energy associated with the first resonator transit (subscript a). Other symbols have their previous significance, that is, V_0 is beam potential, A is beam coupling coefficient, and α is excitation-voltage ratio. For the present, it is assumed that an r-f gap voltage

$$V_1(t) = V_1 \sin \omega t \tag{17.77}$$

exists, without saying how it is created. The distance to which any electron penetrates the field will be proportional to the value of V_a given by Eq. (17.76). The corresponding initial velocity of an electron entering the reflector space is

$$v_a = v_0 \left(1 + \frac{A\alpha}{2} \sin \omega t\right) \tag{17.78}$$

which is identical with the expression encountered in the klystron-amplifier bunching resonator. The value of v_a from Eq. (17.78) will determine the initial slope of the parabola associated with any electron. A sample distance-time diagram constructed by applying the above observations is shown in Fig. 17.30. The bunching action is quite evident and is slightly greater in this case than that required to produce a first infinite peak of current. If the bunch which is formed returns to the resonator at such a time that the electrons pass through the resonator gap *when they are opposed by the potential gradient* between the resonator grids, then energy will be extracted from the bunched current and the tube may oscillate if other conditions are suitable. Of great significance is the observation that the bunch forms about the electron which passes through the resonator when the modulating voltage is changing from *accelerating to retarding* in its action. (It will be remembered that in the klystron amplifier the bunch formed about the electron which passed through the bunching resonator when the modulating voltage was changing from retarding to accelerating.) This happens because those electrons which enter the reflector field with energies greater than the average will penetrate *farther* and take *longer* to return. Accordingly, electrons which have been slowed down will overtake those which have been speeded up, which is just the opposite to what happens in the klystron amplifier. A combination of this property and the requirement that the electrons return when the resonator voltage opposes their motion through the resonator indicates that oscillations can occur only when the d-c transit time is in the vicinity of $n + \frac{3}{4}$ cycles, where n is zero or any integer. The distance-time diagram of Fig. 17.30 shows a d-c transit time of $1\frac{3}{4}$ cycles, which admits of oscillation.

Bunching Theory of the Reflex-klystron Oscillator. There has already been given in Eq. (17.78) an expression for the electron velocity resulting from a first transit of the resonator. For the distance-time diagram of Fig. 17.30 the principle of conservation of charge will hold, just as it did for the klystron amplifier; *i.e.*, Eq. (17.37) will again apply,

$$I_b = I_a \left| \frac{dt_a}{dt_b} \right| \tag{17.37}$$

where I_a is again equal to I_0. Likewise it will be true that the relation between arrival and departure time of any electron will be the same as for the klystron amplifier; *i.e.*, Eq. (17.30) will hold,

$$t_b \cong t_a + t_0 \left(1 - \frac{A\alpha}{2} \sin \omega t_a \right) \tag{17.30}$$

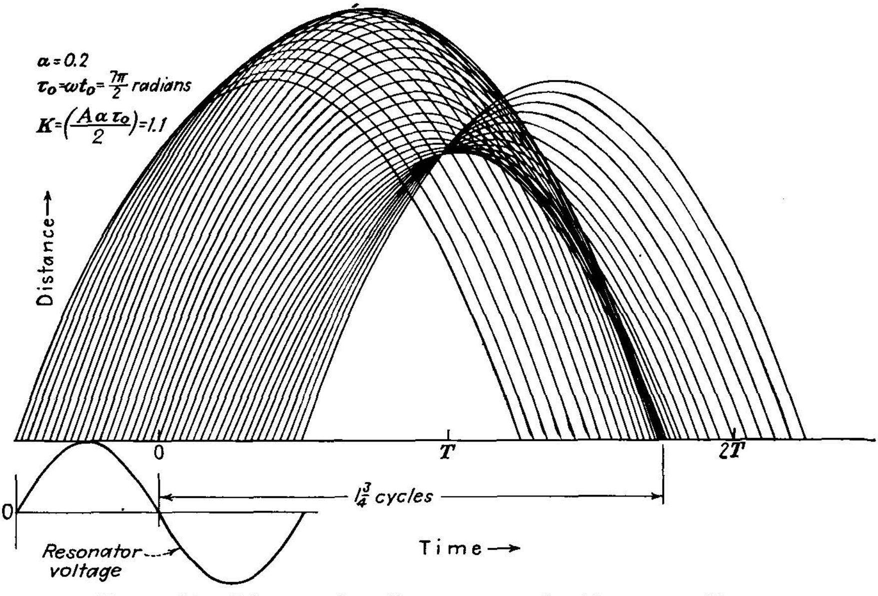


FIG. 17.30.—Distance-time diagram of a reflex-klystron oscillator.

Since the equations that determine the shape of the current pulse are the same as for the klystron amplifier, it is expected that the resultant current will be the same and it is. Thus

$$I_{b1} = 2I_0 J_1(k) \epsilon^{-j\left(\tau_0 - \frac{\pi}{2}\right)} \tag{17.50}$$

where all the symbols have their previous significance. It must be noted, however, that for the case of the reflex-klystron oscillator the current of Eq. (17.50) is taken with respect to the direction of the second electron transit and this defines the positive gap voltage. For some purposes it is more convenient to deal with the current associated with the direction of the first electron transit, which will be the negative of that given by Eq. (17.50) and will have the form

$$-I_{b1} = 2I_0J_1(k)\epsilon^{-j\left(\tau_0+\frac{\pi}{2}\right)} \tag{17.79}$$

In this form it is apparent that the klystron-amplifier and reflex-klystron-oscillator bunches form about zero-excitation-voltage points which are half a cycle apart.

From Eq. (17.50) it is apparent that whenever the d-c transit angle is $\frac{\pi}{2}$, $\frac{5\pi}{2}$, $\frac{9\pi}{2}$, etc., radians the current will be in phase with the gap voltage and the beam action will be equivalent to that of a positive resistance shunted across the gap. Whenever the d-c transit angle is $\frac{3\pi}{2}$, $\frac{7\pi}{2}$, $\frac{11\pi}{2}$, etc., the fundamental component of beam current will be 180 deg out of time phase with the gap voltage and the beam action will be equivalent to a negative resistance shunted across the gap. Under this last set of conditions the tube may oscillate if the magnitude of the negative beam resistance is less than the positive resonator resistance.

Self-admittance of the Beam. It is convenient for purposes of analysis to speak of the beam admittance, defined as the ratio of the fundamental component of induced resonator current to the gap voltage that produces it. From Eq. (17.50) this is

$$Y_e = \frac{AI_{b1}}{V_1} \qquad \text{17.80)}$$

$$Y_e = \frac{2AI_0}{V_1}J_1(k)\,\epsilon^{-j\left(\tau_0-\frac{\pi}{2}\right)} \tag{17.81}$$

$$Y_e = A^2G_0\tau_0\,\frac{J_1(k)}{k}\,\epsilon^{-j\left(\tau_0-\frac{\pi}{2}\right)} \tag{17.82}$$

where Y_e is the self-, or electronic, admittance of the beam, $G_0 = \frac{I_0}{V_0}$, and use has been made of the relation $k = \frac{A\alpha\tau_0}{2}$. The factor A that appears in Eq. (17.80) arises from the desirability of comparing beam and resonator admittances in terms of *induced* resonator currents.

The ratio of the electronic to the beam admittance is perhaps most conveniently written in component form by expanding the exponential into a complex quantity.

$$\frac{Y_e}{G_0} = A^2\tau_0\,\frac{J_1(k)}{k}\,(\sin\tau_0 + j\cos\tau_0) \tag{17.83}$$

The electronic admittance of the tube is seen to be a function of the d-c transit angle and the bunching parameter k, which also involves the transit angle. Likewise it is seen that the electronic admittance has both a conductive and a susceptive component depending upon the value of the d-c transit angle. Let it be assumed first that the value of k is zero, corresponding to zero r-f gap voltage. The factor $\frac{J_1(k)}{k}$ then has a value of $\frac{1}{2}$. Accordingly, the conductance and susceptance of the beam have the form shown in Fig. 17.31. The zero signal value of beam conductance is of the form $x \sin x$. It is first positive and then alternately negative and positive with increasing amplitude. Correspondingly, the zero signal value of beam susceptance is of the form $x \cos x$. It is first positive for a quarter cycle and then alternately negative and positive with increasing amplitude as d-c transit angle increases (reflector voltage decreases). The tube may oscillate whenever the beam conductance is negative and exceeds the magnitude of the positive resonator conductance. The negative of the resonator conductance is shown in Fig. 17.31. For the value shown, oscillations will not occur the first time the beam conductance is negative, for its magnitude is not large enough. Oscillations will occur the second time and subsequent times

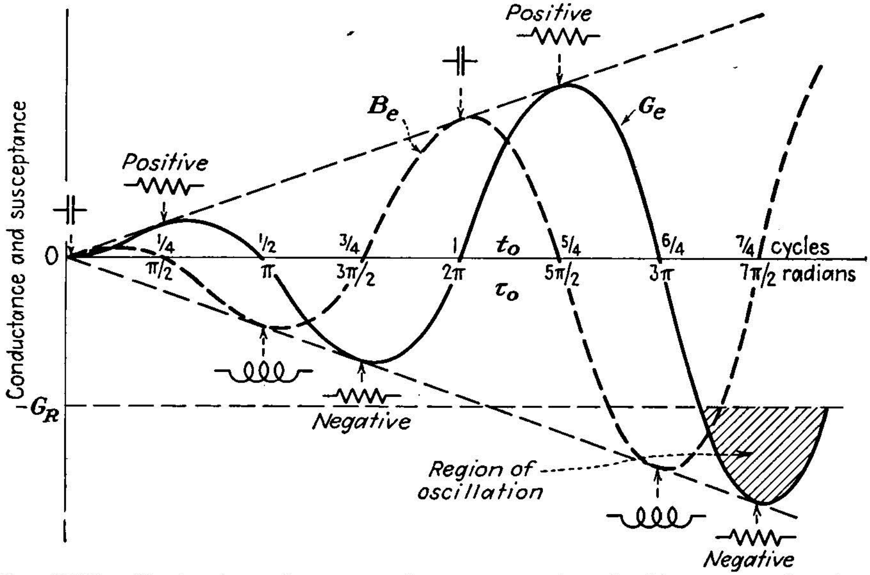


FIG. 17.31.—Electronic conductance and susceptance of a reflex klystron as a function of electron transit-angle.

the beam conductance is negative as d-c transit angle is increased, for the magnitude is then greater than the resonator conductance.

Mechanism by Which Oscillations Start. The above statements about conditions for oscillation are readily demonstrated by reference to the

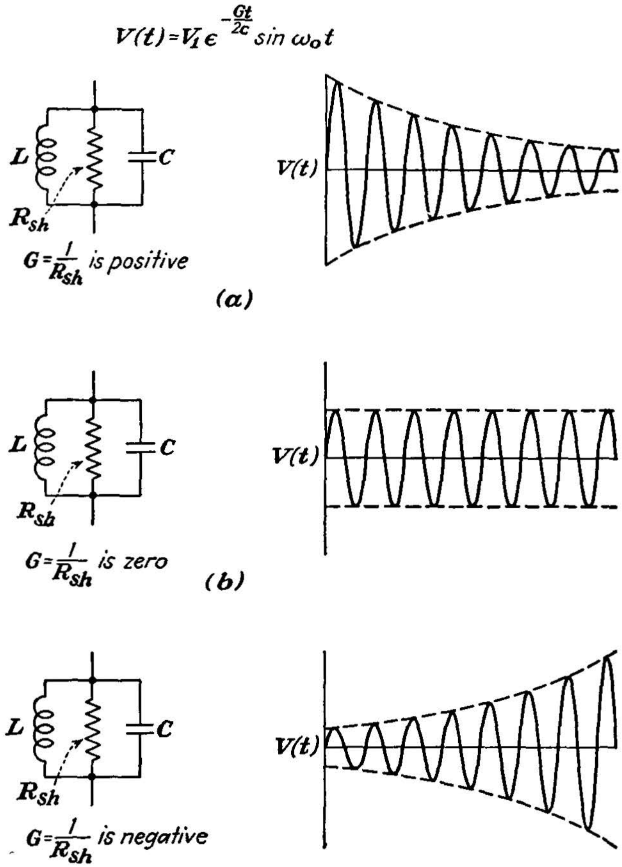


FIG. 17.32.—Transient response of an RLC circuit.

transient response of a parallel combination of a resistance, inductance, and a condenser. Let the circuit be as shown in Fig. 17.32. The voltage transient across such a circuit that has been shock-excited by some disturbance has the form

$$V(t) = V_1 \epsilon^{-\frac{Gt}{2C}} \sin \omega_0 t \tag{17.84}$$

where G is the value of the shunting conductance, C is the capacity, and

$\omega_0 = \frac{1}{\sqrt{LC}}$. This equation is quite a good approximation for circuits with a Q greater than 10. From the form of Eq. (17.84) it is seen that the transient response has the form of a damped sine wave, though the extent of the damping depends upon the nature of the conductance. If the conductance is *postive*, then the response is a damped sine wave, as shown in Fig. 17.32*a*. If the conductance is *zero*, the response will be an undamped sine wave, as shown in Fig. 17.32*b*. If the conductance is *negative*, then the response is an exponentially increasing sine wave, as shown in Fig. 17.32*c*. The circuit here discussed is the equivalent of that encountered in the reflex-klystron oscillator, the net conductance being the algebraic sum of the resonator and beam conductance. The net conductance can be positive, zero, or negative. If the beam conductance is smaller in magnitude than the resonator conductance, then the net conductance is positive. If the beam conductance is negative and equal in magnitude to the resonator conductance, then the net conductance is zero. If the beam conductance is negative and greater in magnitude than the resonator conductance, then the net conductance is negative.

The mechanism by which oscillations start in a reflex-klystron oscillator is evident from the above. Suppose that there is initially no r-f voltage but that the beam conductance is negative and greater in magnitude than the resonator conductance, as at the second negative-conductance peak shown in Fig. 17.31. The net conductance will be negative, and hence any small disturbance will start a transient response like that shown in Fig. 17.32*c*. As the transient gap voltage builds up, the beam conductance will decrease in magnitude in accordance with the factor $\frac{J_1(k)}{k}$, as shown in Fig. 17.33. The result will be that the transient voltage will build up less rapidly, but it will continue to increase in magnitude as long as the magnitude of the negative beam conductance exceeds the resonator conductance. As the transient gap voltage builds up, the beam conductance will continue to drop off until finally it is exactly equal to the resonator conductance. At this value of voltage the net conductance will be zero, and stable oscillations of constant magnitude as shown in Fig. 17.32*b* will result.

Variation of Beam Conductance with Amplitude of Oscillation. Examination of the real part of Eq. (17.83) shows that the beam conductance varies as the factor $\frac{J_1(k)}{k}$ with the degree of bunching. Of particular interest are the values of transit angle of $\frac{3\pi}{2}$, $\frac{7\pi}{2}$, $\frac{11\pi}{2}$, etc., for which the negative conductance has its greatest magnitude. Shown in Fig. 17.33

are relative values of beam conductance as a function of resonator gap voltage for different values of n, oscillations being considered possible for transit times of $n + \frac{3}{4}$ cycles. All these curves have the same form but differ in their initial magnitude, which is always $\frac{\tau_0}{2}$, and in their rate of decline, which increases as the value of n increases. Marked on the curves are the abscissas corresponding to maximum power. As will be shown, this occurs for a value of k equal to 2.405 and yields a conductance that is 43.1 per cent of the maximum value. The curves are extended only to a value of $\frac{AV_1}{V_0}$ of 0.5, which is well beyond the limit of accuracy of the first-order theory. From the slope of the curves it is expected that the oscillations for large values of n are more stable than for the low values.

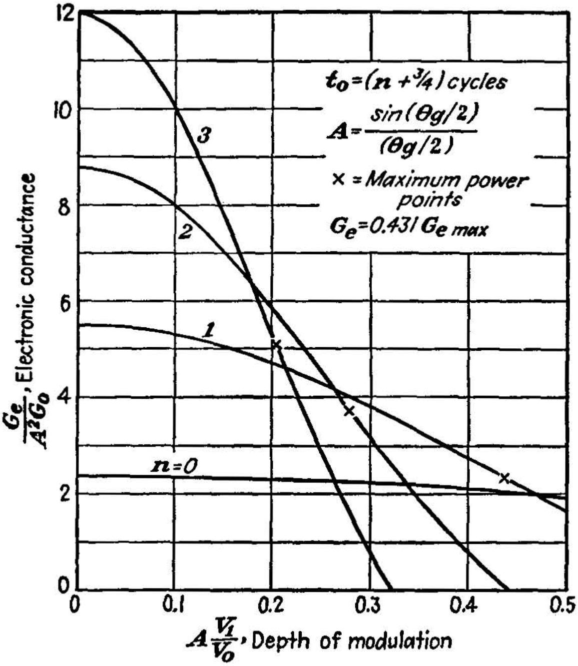


FIG. 17.33.—Electronic conductance of a reflex-klystron oscillator as a function of r-f voltage for transit angles admitting of oscillation.

The Electronic-admittance Spiral. The condition for oscillation of a reflex-klystron oscillator is that the electronic beam conductance be negative and equal in magnitude to the positive resonator conductance.

For the resultant circuit to be resonant it is also necessary that the net susceptance of the parallel resonator-beam combination be zero. All this can be stated by the single equation

$$Y_r = -Y_e \tag{17.85}$$

where Y_r is the resonator admittance and Y_e is the electronic admittance of the beam. For many purposes it is convenient to plot the locus

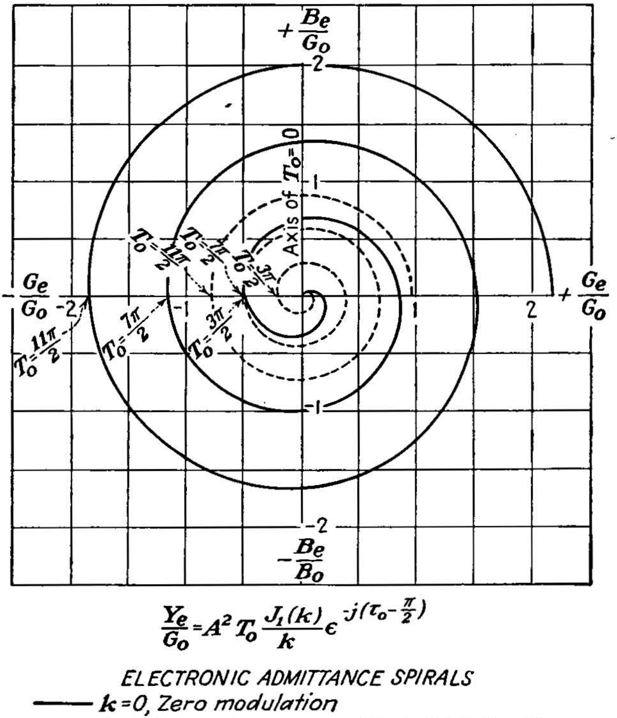
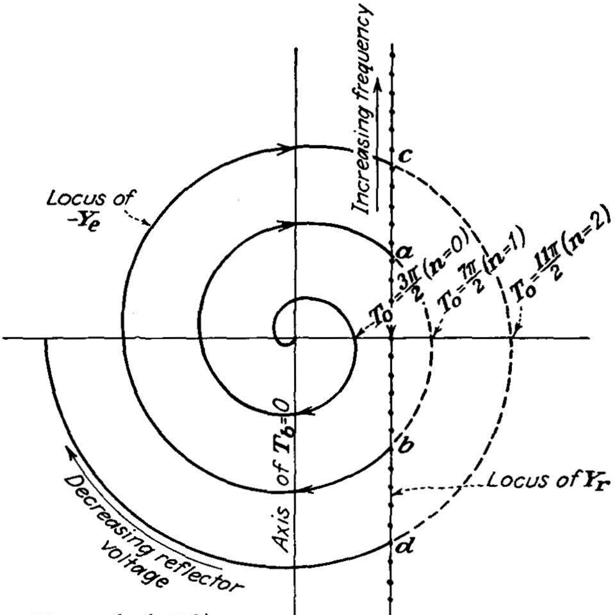


FIG. 17.34.—Electronic-admittance spirals of reflex-klystron oscillators.

of admittance as frequency is varied since for ordinary resonant circuits the loci will usually be of some simple geometrical form. If a locus of electron admittance also be plotted, then limits of oscillation can be determined by intersections of the resonator and negative-beam-admittance loci. Pierce has suggested the use of a plot of the locus corresponding to Eq. (17.83) for the beam admittance. For any fixed value of k the locus of the beam admittance is a spiral of Archimedes. Two such spirals are shown in Fig. 17.34. The solid curve is the spiral locus for

k equal to zero, or the limit-of-oscillation value. The dotted spiral locus is for a value of k of 2.405, or the maximum-power value. The spirals are geometrically similar except that the dotted spiral is only 43.1 per cent of the size of the solid spiral. Transit angle is measured clockwise from the positive susceptance axis, increasing transit angle corresponding to decreasing reflector voltage. The beam conductance is seen to be negative whenever the transit angle is within $\frac{\pi}{2}$ radians of $\frac{3\pi}{2}$, $\frac{7\pi}{2}$, $\frac{11\pi}{2}$, etc.



$Y_R = G_r(1+j2Q\delta)$

$\delta = \frac{\omega - \omega_0}{\omega_0}$, $Q = 100$

$\Delta\delta = 0.001$ (between points)

FIG. 17.35.—Analysis of oscillations of a reflex-klystron oscillator by means of admittance loci.

Reflex-klystron Oscillation with a Simple Resonant Circuit. When the resonant circuit is representable by a parallel combination of a resistance, inductance, and capacity, then the locus of the circuit admittance is a straight line parallel to the susceptance axis in the positive half of the admittance plane, as shown in Fig. 17.35.[1] On this line,

---

[1] Specifically, the approximate formula for admittance as a function of frequency is $Y_r(\omega) = G_r(1 + 2jQ\delta)$, where δ is the fractional frequency deviation from resonance.

frequency increases upward. Also plotted in Fig. 17.35 is the negative of the admittance spiral. Let it now be supposed that every parameter in the oscillating circuit and tube is kept constant except the reflector voltage, which is varied from some large negative value to zero. Then on the beam-admittance spiral this corresponds to a clockwise traversing of the spiral. As transit angle increases with reduction of the magnitude of repeller voltage, the beam admittance will spiral out from a point on the curve near the origin. When the beam admittance first has a negative-conductance component (right half of Fig. 17.35), oscillations will not occur for the case shown because the magnitude of the negative beam conductance is less than the positive resonator conductance. As transit angle is increased further, the beam conductance becomes positive (left half of Fig. 17.35) and then negative again. When the transit angle has increased to the point *a* on the spiral, the negative beam conductance equals the resonator conductance in magnitude for the first time and oscillations will start. As transit angle is increased still more, the beam admittance will now follow a segment of the resonator-admittance line from point *a* to *b*, the beam-admittance spiral shrinking with increased gap voltage and output until it is equal to the resonator

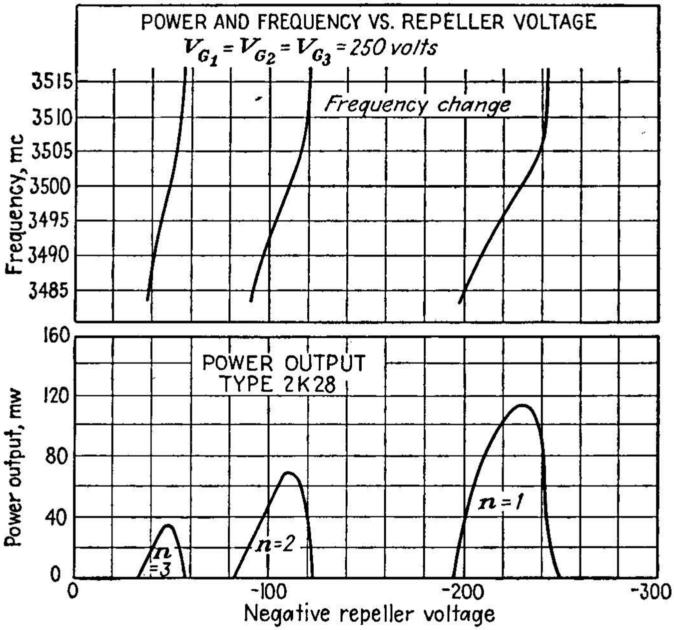


FIG. 17.36.—Power output of a reflex-klystron oscillator as a function of reflector voltage.

admittance at every point. Maximum gap voltage and output will occur when the beam admittance is a pure negative conductance on the real axis of Fig. 17.35. As transit angle is further increased, the tube will drop out of oscillation at *b*. The admittance will now trace the spiral, and oscillations will start again when the point *c* has been reached. As transit angle is increased still further, the beam admittance will now trace the segment of the straight-line resonator-admittance locus between *c* and *d*, where the tube will again drop out of oscillation, and so on.

If the power output of a reflex klystron oscillator be observed with an oscilloscope connected to a crystal output as the reflector voltage is swept sinusoidally, the trace shown in the lower part of Fig. 17.36 results. Each output pulse shown here corresponds to a segment of the resonator admittance between intersections with the beam admittance. The frequency corresponding to the center of the different output pulses is the same but changes through the pulses as shown. The frequency deviation for each mode of oscillation is about the same, but it will occur for a smaller change in voltage and be more linear for large values of n.

Power Relations in the Reflex-klystron Oscillator. The power transferred from the resonator to the bunched electron beam is the product of the resonator voltage by the in-phase component of induced resonator current. This has the value

$$P_e = \frac{V_1 A I_{b1}}{2} \tag{17.86}$$

$$P_e = V_1 I_0 A J_1(k) \sin \tau_0 \tag{17.87}$$

or

$$\frac{P_e}{P_0} = \frac{2}{\tau_0} k J_1(k) \sin \tau_0 \tag{17.88}$$

by application of the definition of k given in Eq. (17.33). The power transferred from the resonator to the beam will be negative whenever $\sin \tau_0$ is negative, which is to say that the power is actually transferred from the beam to the resonator under this condition. The power delivered to the resonator will be a function of the bunching parameter k, which is in turn determined by the requirement that the negative of the beam admittance equal the resonator admittance. A set of curves showing how power transferred from beam to resonator varies with r-f gap voltage is given in Fig. 17.37. The peak power transferred for each mode of operation shown occurs for a value of k of 2.405 and is lower for successively higher values of the transit angle. This might at first thought seem to indicate that the maximum power would be obtained with the lowest transit angle of the possible oscillation values, $\frac{3\pi}{2}$, $\frac{7\pi}{2}$,

$\frac{11\pi}{2}$, etc. However, the maximum power for the lowest transit angle occurs at such large gap voltages that the resonator and load demands may exceed the power which can be delivered and oscillations will not occur at all.

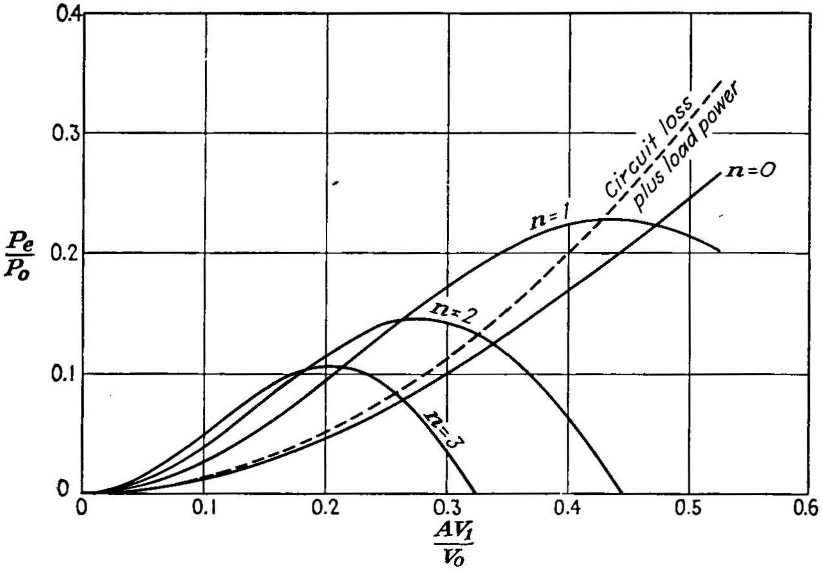


FIG. 17.37.—Power output of a reflex-klystron oscillator as a function of bunching parameter (theoretical).

Equation (17.88) is really an expression for efficiency. The maximum theoretical efficiencies apparent here for different values of n where oscillations are presumed to occur in $n + \frac{3}{4}$ cycles transit time are listed below:

n	Efficiency	$\frac{AV_1}{V_0}$
0	0.531	1.018 (not valid)
1	0.227	0.436
2	0.145	0.278
3	0.106	0.204
n	$\frac{0.398}{n + \frac{3}{4}}$	$\frac{0.767}{n + \frac{3}{4}}$ (for $n > 3$)

In the above tabulation the d-c transit angle has the value of $(n + \frac{3}{4})2\pi$. The values for n equal to zero are obviously not valid, for the excitation-voltage ratio greatly exceeds the limit of accuracy of the theory. Values for higher orders are progressively more accurate. Actual measured efficiencies are of the order of one-fourth of the theoretical values given above.

In the previous discussion there have been given formulas for the fundamental component of beam current [Eq. (17.50)], for electronic admittance [Eq. (17.83)], for electronic power [Eq. (17.88)]. Each o these quantities depends upon the bunching parameter k in some combination of the first-order Bessel function of the factor k. Thus for a

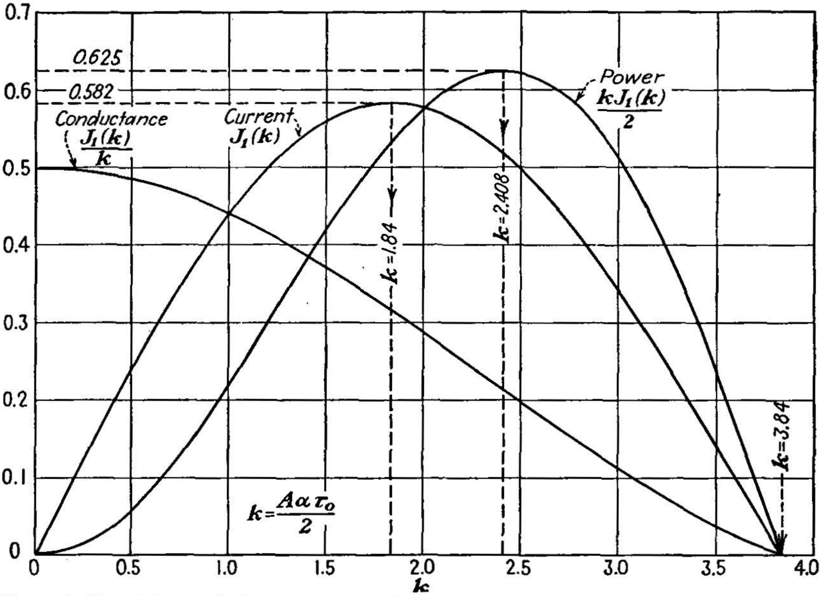
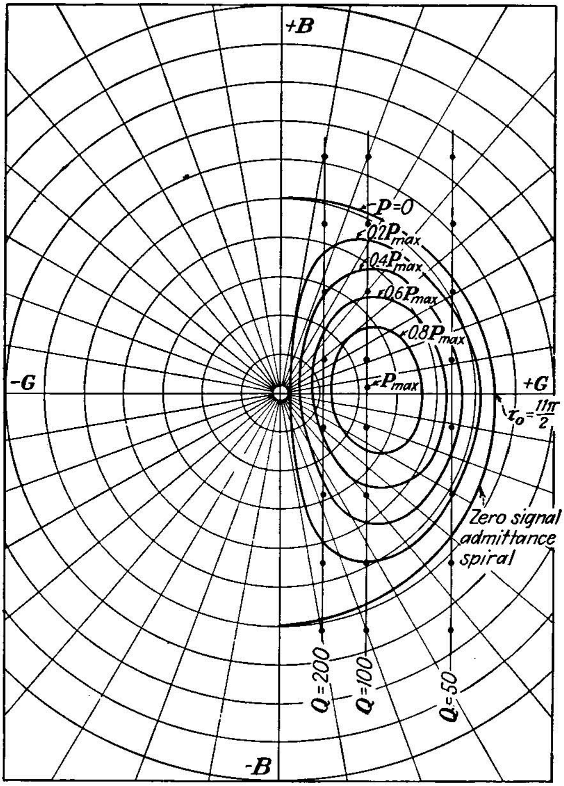


FIG. 17.38.—Theoretical current, conductance, and power function in a reflex-klystron oscillator.

fixed electron transit time the fundamental component of current is proportional to $J_1(k)$, the electronic conductance is proportional to $\frac{J_1(k)}{k}$, and the electronic power is proportional to $kJ_1(k)$. It is of interest to plot these functions side by side in order to compare their properties. This is done in Fig. 17.38. The three functions have in common a zero value for a value of k equal to 3.84. The maximum value of current occurs for a value of k equal to 1.84. The maximum value of power occurs for a value of k of 2.408, corresponding to 0.431 of the maximum value of conductance.

The magnitude of the actual power delivered to the resonator is best obtained by plotting contours of equal power output on an admittance diagram. Such contours may be calculated from Eq. (17.88) and are shown in Fig. 17.39 for transit angles in the vicinity of $\frac{11\pi}{2}$ radians ($n = 2$). Contours for other values of n will be similar. The power transferred from beam to resonator for any given load admittance is immediately evident from these contours. Of particular interest are the variations of power output with reflector voltage for different degrees of resonator loading. As the resonant circuit is loaded, its straight-line admittance locus assumes a larger conductance component, while at the same time equal frequency increments along the line become smaller relative to the conductance. This is due to the fact that the conductance



Frequency interval between points = 1/5 per cent

FIG. 17.39.—Contours of electronic power output of a reflex-klystron oscillator on an admittance plane. This representation shows the power supplied to both resonator and load.

is given by $\frac{B}{Q}$, where B is the magnitude of either the circuit inductive or capacitive susceptance. At the same time the frequency increment between values of admittance having angles of plus and minus 45 deg is $\frac{f_0}{Q}$. Hence, as the resonant circuit is loaded the Q decreases and the conductance component increases, as does also the frequency increment between the 45-deg values of admittance. Accordingly, the change in susceptance for a given frequency increment decreases. Figure 17.39 shows admittance loci of a resonant circuit for three degrees of loading. The dots on the straight-line admittance loci of this figure show ⅕ per cent frequency variations. The corresponding curves of power output versus reflector voltage are shown in Fig. 17.40.

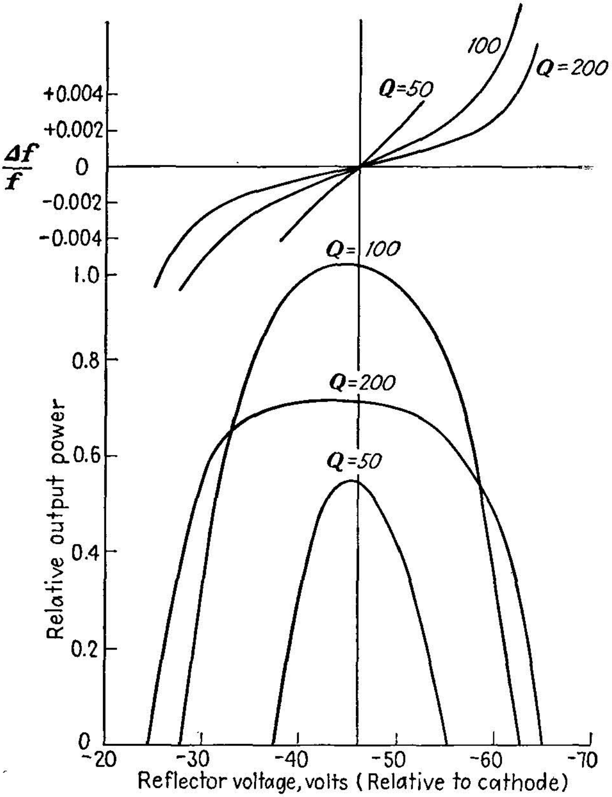


FIG. 17.40.—Power output of a reflex-klystron oscillator as a function of loading for variable reflector voltage.

Of interest is the amount of frequency variation between the limits of oscillation for a given resonator loading as the reflector voltage is varied. This may be calculated exactly from determining the frequencies corresponding to intersections of the zero-signal beam-admittance spiral and the straight-line resonator locus. This determination requires numerical or graphical computation. A good approximation for the amount of frequency variation between limits of oscillation may be obtained by assuming that the beam-admittance locus is a circle whose radius is that corresponding to a transit angle of $2\pi(n + \frac{3}{4})$ radians instead of the actual spiral. With this assumption the limiting intersections of the beam and resonator admittance loci can be calculated. The errors appearing at the two intersections as a result of the assumption of a circular locus cancel as far as the *difference* in frequency for the two intersections is concerned and the resulting expression is accurate within a few per cent. It is

$$\text{Half band width between oscillation limits} \quad \frac{\Delta f}{f_0} = \frac{1}{2Q}\sqrt{\left(\frac{\tau_0 G_0 A^2}{2G_r}\right)^2 - 1} \qquad (17.89)$$

where the Q is that of the resonator, G_0 and G_r are beam and resonator conductance at mid-mode, respectively, τ_0 is the transit angle at mid-mode of value $2\pi(n + \frac{3}{4})$.[1] It will be noted that this expression properly reduces to zero when the resonator-admittance line is tangent to the beam-admittance spiral, *i.e.*, when

$$G_r = \frac{\tau_0 G_0 A^2}{2} \qquad (17.90)$$

This equation represents a limit of oscillation. It may also be used to determine the minimum beam current required to start the oscillation for a given resonator and a given reflector mode of operation. Since the beam admittance as given by Eq. (17.83) is proportional to the d-c beam admittance, which in turn is proportional to the d-c beam current, then the *starting current* for any resonator and reflector mode is

$$I_{\min} = \frac{2G_r V_0}{\tau_0 A^2} \quad \text{amperes} \qquad (17.91)$$

$$I_{\min} = \frac{2V_0}{\tau_0 A^2 R_{sh}} \quad \text{amperes} \qquad (17.92)$$

where $R_{sh} = \frac{1}{G_r}$ is the shunt resistance of the resonator.

---

[1] This is arrived at as follows: At the intersection of the assumed circular beam-admittance locus and the straight-line resonator locus the equality of the conductance component gives $G_r = \frac{1}{2}G_0\tau_0A^2 \cos \Delta\tau_0$, whereas the equality of the susceptance components gives $QG_r\frac{\Delta f}{f_0} = \frac{1}{2}G_0\tau_0A^2 \sin \Delta\tau_0$, where $\Delta\tau_0 = \tau_0\frac{\Delta f}{f_0}$. Squaring and adding these relations and then solving for the band width yield Eq. (17.89).

Voltage Stability of Reflex-klystron Oscillators. Since the d-c transit angle of the electrons in the reflector space depends upon the electrode voltages, the frequency of oscillation changes with supply voltage, as has already been shown. This may be a serious limitation in application, for a small change in electrode voltage will cause a relatively large change in frequency of oscillation. If it is assumed that the cathode and reflector voltage both change by the same percentage with a given change in line voltage, then the frequency stability at mid-mode with respect to voltage is[1]

$$\frac{df}{f_0} = \frac{\tau_0}{4Q_0}\frac{dV_0}{V_0} \qquad \text{cycles per volt} \tag{17.93}$$

For the 2K28 with the characteristics shown in Fig. 17.36, this variation is of the order of ½ mc per volt at mid-mode.

If the cathode-reflector voltage instead of varying proportionally to cathode-resonator voltage is kept constant, then the frequency change with voltage is zero when the cathode-resonator and cathode-reflector voltages are equal. This occurs because under this condition the gradient of potential in the resonator-reflector space and the distance to the point of electron reversal change by the same percentage with a change of cathode-resonator voltage, thus keeping the transit angle constant. The frequency change with voltage for other conditions can be made low by giving the cathode-resonator and cathode-reflector voltages different degrees of regulation of the proper value.

The relatively large frequency change with voltage observed in Eq. (17.93) may actually be of use in some applications, for it indicates that frequency modulation is easily achieved. Even in cases where stability is desired, automatic frequency control is easy to apply.

17.10. Broad-band Operation of Reflex-klystron Oscillators. Tubes of the type shown in Fig. 17.28*b* are frequently used with an external resonant cavity in the form of a concentric line. Such an arrangement permits of an extremely wide band of frequency operation. The low-frequency limit of oscillation will be governed by the resonator-reflector distance, which in turn determines the largest transit time that can be used without the reflector drawing current. The high-frequency limit will be determined either by the gap transit angle, which reduces the beam

---

[1] Arrived at by applying the relation $\frac{df}{dV_0} = \frac{df}{dt_0}\frac{dt_0}{dV_0}$ to the formulas

$$\frac{2Q(f - f_0)}{f_0} = \cos \tau_0$$

and Eq. (17.74) on the assumption that $\frac{V_r}{V_0}$ is constant.

coupling coefficient A and thus reduces the beam conductance, or by the resonator shunt resistance, which may drop off with increasing frequency to the point where oscillations cannot be maintained. Operation over a frequency band of as great as 2 to 1 may be had with this type of tube and resonator.

Equivalent Circuit of Concentric-line Resonator. In the ordinary reflex-klystron-oscillator tube built to use an external resonator there will be an appreciable capacity across the electron bunching gap. As a result of this capacity, the tube will oscillate when the capacitive reactance of the gap is equal in magnitude to the inductive reactance of the shorted concentric line attached to it. Accordingly, the equivalent circuit is that shown in Fig. 17.41. This is simply a capacity in parallel with a shorted section of concentric line. Resonance will occur whenever the gap reactance equals the inductive line reactance in magnitude, *i.e.*, when

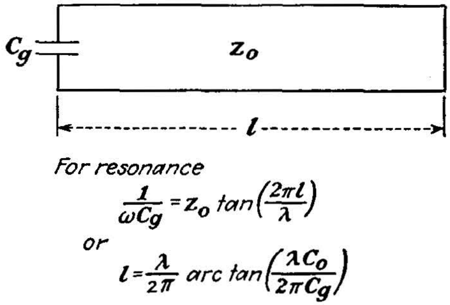


FIG. 17.41.—Equivalent circuit of a concentric-line resonator and reflex-klystron oscillator.

$$\frac{1}{\omega C_g} = Z_0 \tan\left(\frac{2\pi l}{\lambda}\right) \tag{17.94}$$

$$\frac{1}{\omega C_g} = Z_0 \tan\left(\frac{\omega l}{c}\right) \tag{17.95}$$

where ω is the angular frequency, C_g is the gap capacity, Z_0 is the characteristic impedance of the line given by $138 \log_{10}\left(\frac{r_2}{r_1}\right)$, l is the equivalent line length, and λ is the wave length. Curves of magnitude of capacitive reactance and inductive line reactance are shown in Fig. 17.42. Resonances will occur at the intersections of the two reactance curves shown. The resonances are multiple, which means that, for a given line length, resonance can occur at a number of different frequencies. The frequencies of resonance will occur at frequencies somewhat less than those for which the line is $\frac{1}{4}$, $\frac{3}{4}$, $\frac{5}{4}$, etc., wave lengths long.

For convenience in subsequent analysis, Eq. (17.94) may be solved for l and written in the form

$$l = \frac{\lambda}{2\pi} \arctan \left(\frac{\lambda C_0}{2\pi C_g} \right) \tag{17.96}$$

as a function of wave length, where C_0 is the distributed capacity of the concentric line per unit of length.[1] The general form of curves of l as a function of wave length is shown in Fig. 17.43. The various branches of this curve correspond to the different possible line lengths. The lowest branch of the curve gives resonant lengths slightly less than $\frac{1}{4}$ wave length, the next gives lengths slightly shorter than $\frac{3}{4}$ wave

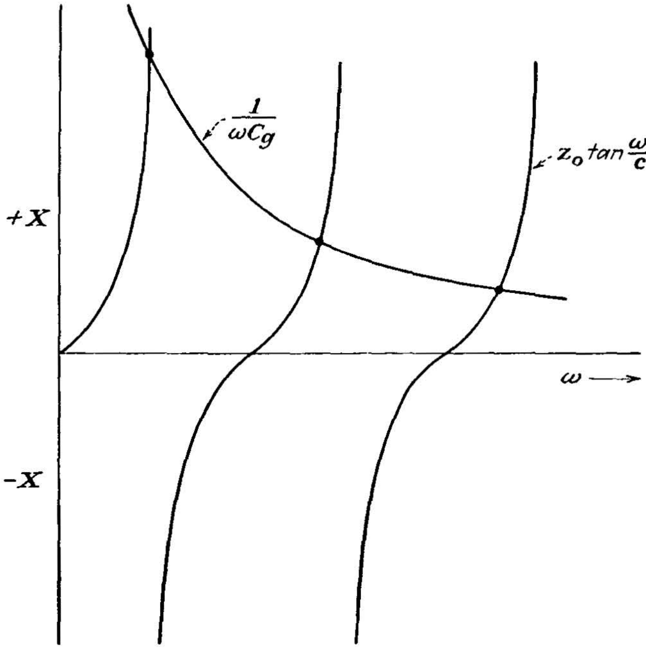


FIG. 17.42.—Curves of reflex-klystron gap reactance and inductive line reactance as a function of frequency.

---

[1] This follows from the fact that $Z_0 = \sqrt{\frac{L_0}{C_0}}$ and $c = \frac{1}{\sqrt{L_0 C_0}}$, as a result of which $cZ_0 = \frac{1}{C_0}$. Utilization of this last relation along with $c = \lambda f$ leads to Eq. (17.96). Distributed inductance of the line per unit length is L_0; velocity of propagation is $c = 3 \times 10^{10}$ cm per sec.

length, and so on. All curves start out like parabolas from the origin and then become straight lines parallel to a line whose slope is ¼, ¾, 5/4, etc. respectively for the different curves starting from the bottom.

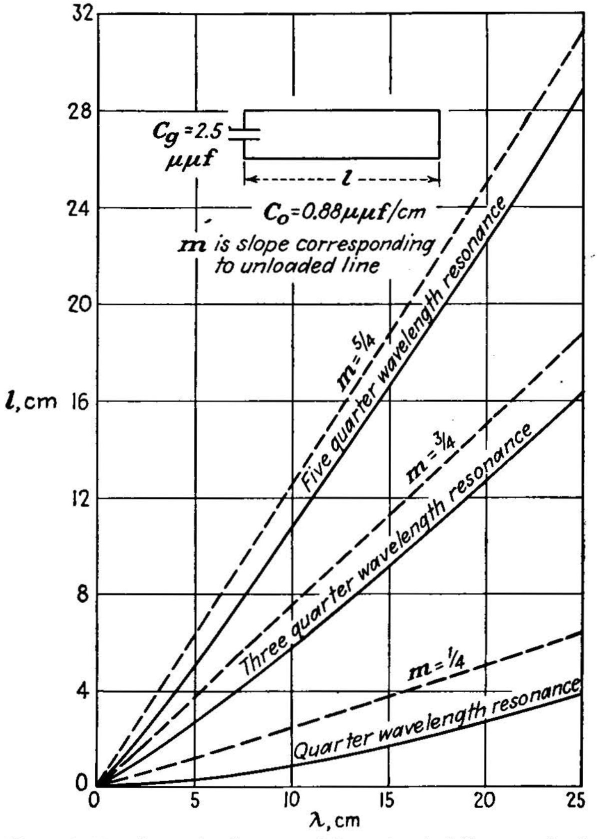


FIG. 17.43.—Length of a capacitively loaded line required to give resonance at various wave lengths.

Possible Modes of Oscillation. A reflex-klystron oscillator with an external concentric-line resonator has numerous possible modes of oscillation. Oscillations can occur whenever the transit time of electrons in the reflector space is $n + \frac{3}{4}$ cycles, where n is zero or any positive integer, and the cavity length is effectively an odd quarter of wave lengths long. The tube gap capacity always loads the line resonator so that its actual length is less than an odd quarter of wave lengths long, but it is convenient to speak of the *effective length* as that corresponding to the nearest odd number of quarter waves. Accordingly, if the cavity length and electron transit time in the reflector space are admitted as variables, a considerable number of oscillation modes are possible. Each oscillation mode needs to be designated in terms of both the electron transit time and the effective cavity length.

The possible oscillation modes of the reflex-klystron oscillator with a concentric-line resonator may be investigated conveniently by applying an alternating voltage to the reflector and observing by means of a cathode-ray oscilloscope the reflector voltages at which oscillations occur. This is done by connecting the output of a crystal detector to the vertical plates of the oscilloscope while at the same time applying the alternating reflector voltage to the horizontal plates. The resultant screen representation on the cathode-ray tube will be like that shown in the lower half of Fig. 17.36. If now the cavity length is varied, the modes will change position progressively as the voltage required to give different transit times changes. A plot can be made showing reflector voltage ranges that maintain oscillation as a function of cavity length, which is usually nearly linear with wave length. It is even possible to record such an oscillation mode plot photographically by intensity-modulating the oscilloscope with the crystal output, applying the alternating reflector voltage to the horizontal plates, and obtaining a vertical deflection proportional to cavity length by means of a potentiometer connected to the cavity-plunger drive. The cavity plunger is then moved uniformly throughout its entire travel while a camera integrates the line indication of the oscilloscope. Such a photographically recorded mode plot is shown in Fig. 17.44. The numerous possible modes of oscillation are labeled in terms of their corresponding electron transit time in cycles and their equivalent cavity length in wave lengths. For the type 707B tube shown, oscillations occur for transit times ranging from 1.75 to 3.75 cycles in integral steps and for equivalent cavity lengths ranging from 0.25 to 1.25 wave lengths in half-wave-length steps. Maximum output is obtained for an electron transit time of 2.75 cycles and an equivalent cavity length of 0.75 wave length.

Method of Calculating Oscillation Mode Plot. It is possible to determine graphically the form of a mode plot such as was obtained photographically in Fig. 17.44 on the basis of the simple theory proposed in the previous sections. Such a determination serves as a check upon the validity of the assumptions made in deriving the above theory.

There is desired a relation between reflector voltage and cavity length for the different cavity-length and electron-transit-time modes. A relation between cavity length and wave length has already been given in Eq. (17.96) and shown in Fig. 17.43. A relation between resonant wave length and reflector voltage has been given in Eq. (17.75). From this relation it is seen that for a given transit angle the reflector voltage is linear with the reciprocal of the resonant wave length. Plots of Eqs. (17.75) and (17.96) can be combined to give the desired relations between reflector voltage and cavity length by means of the construction shown in Fig. 17.45. In the second quadrant of this chart is given the linear relation between reflector voltage and reciprocal wave length. In the fourth quadrant is given the relation between resonant wave length and cavity length as shown in Fig. 17.43. In the third quadrant is given the curve relating wave length and its reciprocal. A set of rectangular construction lines tying together points in the second, third, and fourth quadrants for a given set of electron-transit-time and cavity-length modes yields an intersection in the first quadrant which is a point on one of the reflector voltage–cavity-length modes desired.

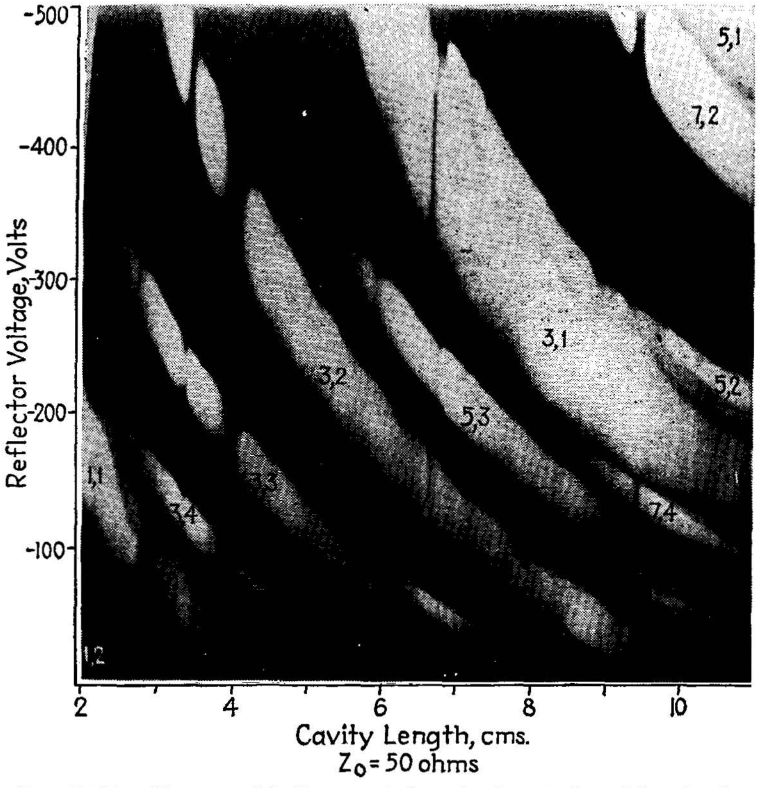


FIG. 17.44.—Photographically recorded mode plot of a broad-band reflex-klystron oscillator. The number pairs give the number of quarter waves in the resonator and the value of n respectively.

Differences between the theoretical and actual reflector-voltage–

cavity-length curves result from numerous limitations to the simple theory. Prominent among these is the inadequacy of the assumption of the equivalent circuit shown in Fig. 17.41. This equivalent circuit is probably adequate at long wave lengths, but not at short ones. At short wave lengths the effect of the corner of the concentric-line cavity adjacent to the tube needs to be considered. The corner has the effect of an impedance-transforming network in the form of a π section with a series inductive reactance and shunting capacitive reactances. Another

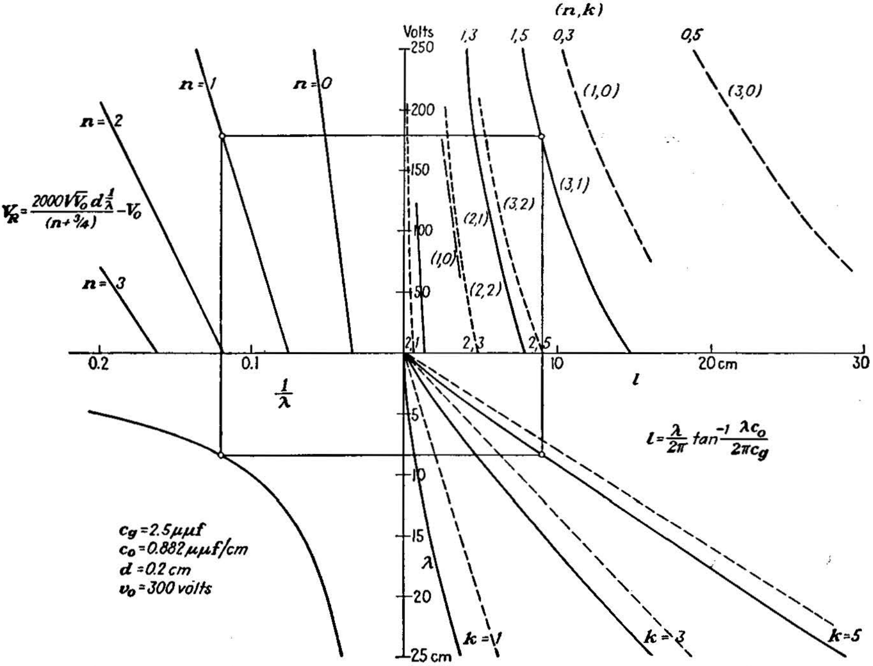


FIG. 17.45.—Construction of a reflector-voltage–cavity-length plot.

limitation to the simple theory is the assumption that the reflector field is linear. Actually, the curvature given the reflector electrode to focus the electrons on the resonator grids may give an appreciable departure from linearity.

Mode Interference. Mode interference may exist in reflex-klystron oscillators with concentric-line resonators. This results from the simultaneous resonance on several modes and may be a serious limitation in oscillator design. The mode interferences may be said to arise primarily from the external resonator characteristics in that they do not exist in an ideal line resonator which has no capacity loading and in that they are independent of the reflector geometry. For an unloaded line the length will be directly proportional to the wave length,

$$l = \frac{p\lambda}{4} \qquad \text{unloaded line length, cm} \tag{17.97}$$

where $p = 1,3,5$, etc., is the number of quarter wave lengths of field variation along the line. The transit time of an electron in the reflector space will be

$$t_0 = \frac{n + 3/4}{f} \qquad \text{sec} \tag{17.98}$$

$$t_0 = \frac{(n + 3/4)\lambda}{c} \qquad \text{sec} \tag{17.99}$$

Equations (17.97) and (17.99) can be combined to give the transit time in the reflector space in terms of idealized cavity length for the various reflector-transit-time and cavity-length modes.

$$t_0 = \frac{(n + 3/4)4l}{pc} \qquad \text{sec} \tag{17.100}$$

$$t_0 = \frac{133.3(n + 3/4)l}{p} \qquad \text{micromicroseconds} \tag{17.101}$$

A plot of electron transit time in the reflector space against ideal cavity length is shown in Fig. 17.46.[1] Examination of this chart shows that ideally there will be no tendency for the tube to oscillate simultaneously on two frequencies. (An exception is the coincidence of the 1.25-wave-length–3.75-cycle mode with the 0.25-wave-length–0.75-cycle mode.) However, many of the prominent modes are very close together, and a small change in the resonator tuning curve may bring them into coincidence.

In reflex-klystron oscillators with concentric-line resonators greatest dependence is placed at present upon the 0.75-wave-length-resonator modes. This is because the line loading by the cold-tube capacitance is usually so high that only very low frequencies can be obtained with quarter-wave resonance and interest is invariably centered about the high frequencies. Undoubtedly, the tubes of the future will be made smaller for a given wave length of operation so that greater use will be made of quarter-wave-length resonances. With 0.75-wave-length-resonator modes and the usual dimensions, oscillations will ordinarily not occur for 0.75-cycle electron transit times, for the beam conductance will be lower than the resonator conductance. Oscillations will not

---

[1] An alternative form of this type of chart was first proposed by W. Huggins and H. Zeidler.

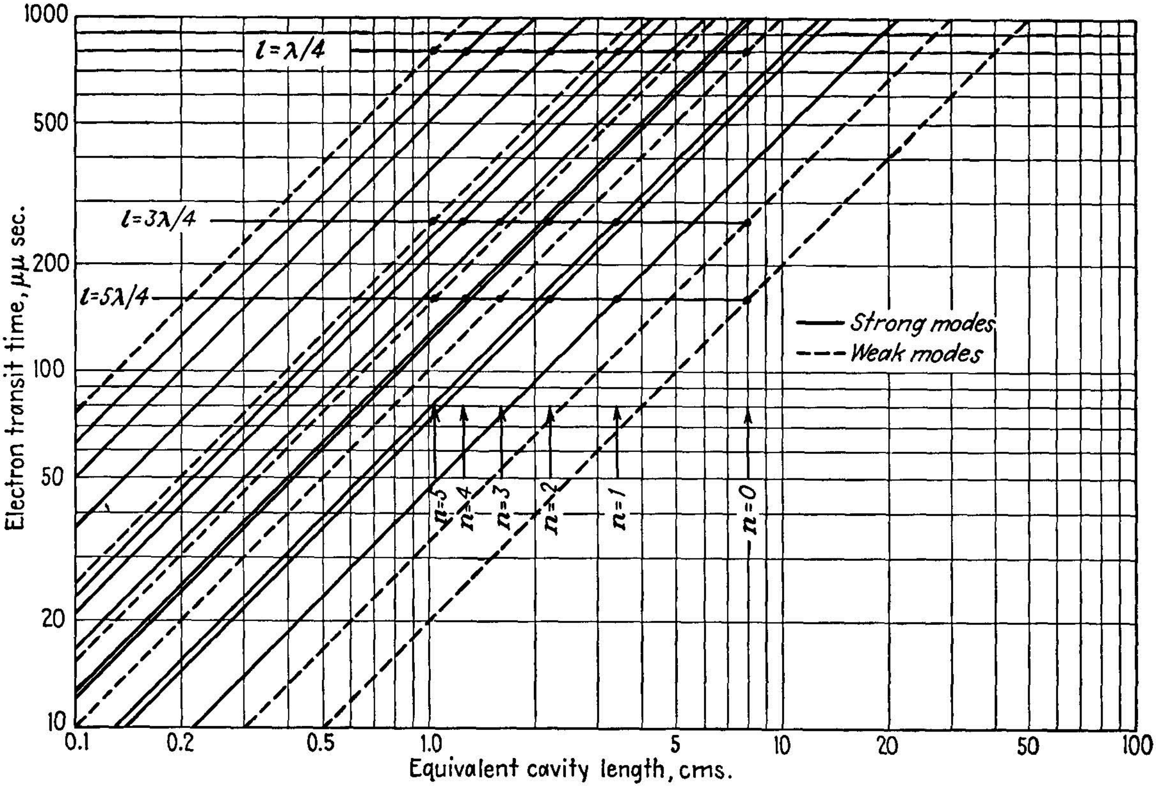


FIG. 17.46.—Electron transit time as a function of unloaded cavity length for a broad-band reflex-klystron oscillator showing the numerous possible modes of oscillation.

ordinarily occur for electron transit times greater than 4.75 cycles, for the electrons will eventually strike the reflector as its voltage is reduced to lengthen the transit time. The resultant range of frequencies that can be obtained with 0.75-wave-length resonators will bracket a 2-to-1 range of frequencies, though use will ordinarily have to be made of several transit-time modes.

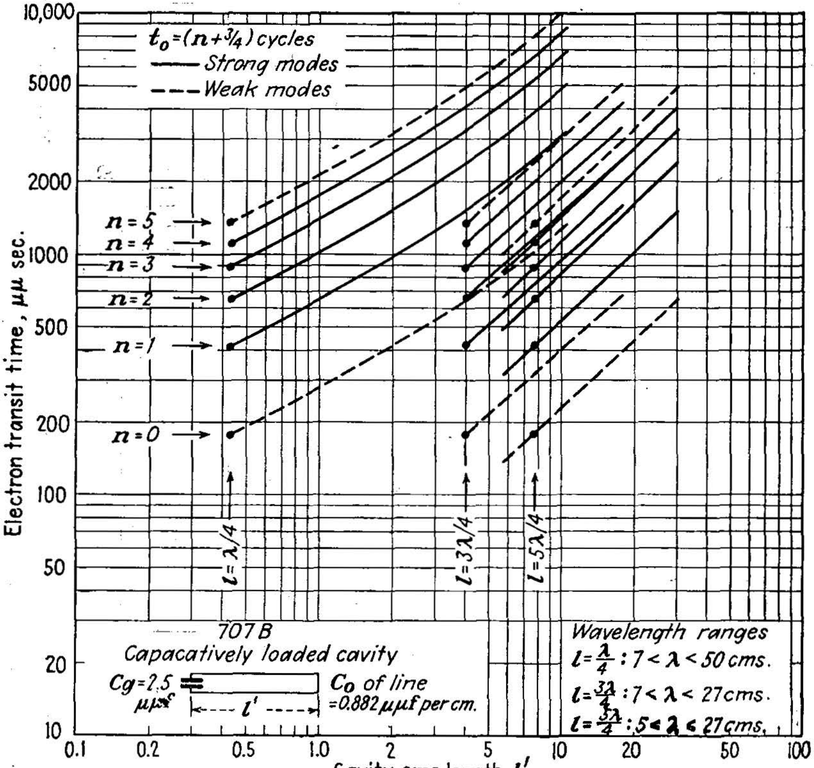


FIG. 17.47.—Electron transit time as a function of length of a capacitively loaded line used with a broad-band reflex-klystron oscillator.

The chart of Fig. 17.46 serves only to give a general idea of the relation between the modes. The effect of tube capacity is to put a bend in the quarter-wave mode line. From Eq. (17.96) it is seen that for very short wave lengths the cavity length varies as the square of the wave length and hence the electron transit time. All curves of the form given by Eq. (17.96) have the same form for any one cavity length when plotted on log-log paper, regardless of the ratio of C_0 to C_g. Some sample curves of this form are shown in Fig. 17.47. Various mode interferences are possible because of the bend in the curves introduced by the tube capacity. Unfortunately, the bend frequently occurs right in the region of desired operation. Examination of Fig. 17.47 shows that the mode interferences are independent of the reflector-field characteristics of the tube. This is because the possible crossovers in a chart such as that of Fig. 17.47 are entirely determined by the geometry of the resonant cavity. Since the reflector voltage required to give a certain transit time is always a single-valued function of the transit time regardless of the shape of the reflector field, it will be true that the mode interferences will not be altered by changing the reflector-field characteristics. This is to say that, if a given gap capacity and resonator dimensions result in mode interferences, merely changing the resonator-reflector distance or the shape of the reflector electrode will not eliminate these interferences. All that can be done is to change the reflector voltage at which they occur.

The above serves only to introduce the problem of mode interference. The actual problem is vastly more complicated than indicated above. This is because a simple capacity loading of the line is not an adequate equivalent circuit. Actual circuits may have impedance-transforming circuits associated with the corner connection. Some tubes will even have an equivalent shunting inductance at the resonator gap. In addition, the reflector field is seldom exactly linear. It may be expected that in the near future much will be added to the present knowledge of mode interferences.

Blind Spots. In addition to mode interferences the broad-band reflex-klystron oscillator will frequently exhibit blind spots, *i.e.*, regions of no oscillation. In general, such spots will occur when the cavity impedance is reduced by the effect of a coupled resonant circuit. The possibility of such coupled resonances are numerous. Most of them can be eliminated by proper design, but some cannot be eliminated by resonator design alone since they are inherent in the higher-order oscillations of the resonator.

The nature of the change of resonator admittance caused by coupled resonances is of considerable interest. The admittance locus of a single high-Q parallel resonant circuit is a straight line parallel to the susceptance axis in the admittance plane, with frequency increasing uniformly upward along the line in the vicinity of resonance. Mathematically, this may be represented by

$$Y_r(\omega) = \frac{B_1}{Q_1}(1 + 2j\,\delta_1 Q_1) \qquad (17.102)$$

where B_1 is the magnitude of either susceptance at resonance and δ_1 is the fractional deviation of frequency from resonance, $\frac{f - f_0}{f_0}$. If now another resonant circuit is coupled to the above circuit, a bump in the form of a protrusion to the right will appear in the line locus of resultant admittance as the coupled circuit passes through resonance. If the coupling to the second circuit be increased, the size of the bump will increase up to a critical value of coupling, $K_1 = \frac{1}{Q_2}$, at which value the bump will have the form of a cusp. If the coupling be increased still further, the cusp transforms into a loop. This action is shown in Fig. 17.48. It will be apparent that if the bump, cusp, or loop extends sufficiently far to the right the resonator admittance may exceed the electronic beam admittance of the oscillator tube and oscillations will cease. Even if the beam admittance is not exceeded, if the resonator admittance has a loop there will be a frequency discontinuity in the oscillations as reflector voltage or cavity length is changed because of the inability of the beam admittance to follow the loop.

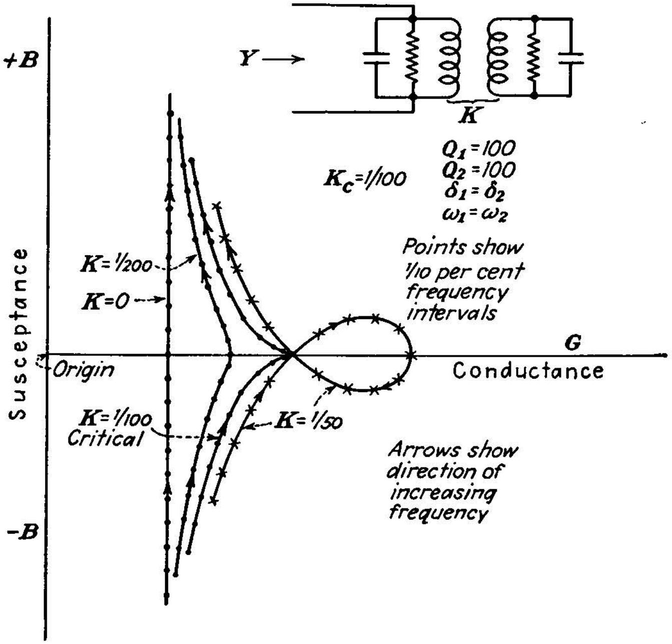


Fig. 17.48.—Admittance loci of a resonant circuit with a coupled resonant secondary.

The quantitative details of the above phenomena will be indicated a little further. The impedance of the secondary circuit in the vicinity of resonance is approximately

$$Z_2 = \frac{X_2}{Q_2}(1 + 2j\,\delta_2 Q_2) \tag{17.103}$$

where δ_2 is the fractional deviation of frequency from secondary resonance and X_2 is the magnitude of either reactance at resonance. The impedance coupled into the primary circuit will be $\frac{\omega^2 M^2}{Z_2}$, or, in component form,

$$\text{Coupled resistance} = \frac{\omega^2 M^2 Q_2}{X_2} \frac{1}{1 + (2\,\delta_2 Q_2)^2} \tag{17.104}$$

$$\text{Coupled reactance} = -\frac{2\,\delta_2 Q_2{}^2 \omega^2 M^2}{X_2} \frac{1}{1 + (2\,\delta_2 Q_2)^2} \tag{17.105}$$

Accordingly, the *series* impedance of the primary is

$$Z_1 = R_1(1 + a) + j\omega L_1(1 - b) + \frac{1}{j\omega C_1} \tag{17.106}$$

where

$$a = \frac{K^2 Q_1 Q_2}{1 + (2\,\delta_2 Q_2)^2} \tag{17.107}$$

and

$$b = \frac{2\,\delta_2 Q_2{}^2 K^2}{1 + (2\,\delta_2 Q_2)^2} \tag{17.108}$$

where use has been made of the relation $K^2 = \frac{M^2}{L_1 L_2}$. Since $Y_r = \frac{C_1 Z_1}{L_1}$, the corresponding resonator admittance across the capacity junction is approximately

$$Y_r = \frac{B_1}{Q_1}[(1 + a) + 2j\,\delta_1 Q_1 - jQ_1 b] \tag{17.109}$$

where all the symbols have their previous significance. It is seen that the effect of a coupled resonant secondary circuit is to increase the normal conductive component of the primary admittance by a fractional amount a and to reduce the normal susceptive component of the primary admittance by a fractional amount $b/2\delta_1$. The reduction in the susceptance component may be so large that the net susceptance actually decreases with frequency. This happens when the coupling exceeds the critical value

$$K = \frac{1}{Q_2} \tag{17.110}$$

for the case of equal primary and secondary resonant frequencies, as may be seen by setting the derivative of the susceptance term of Eq.

(17.109) with respect to δ equal to zero. The values of zero susceptance for equal primary and secondary resonant frequencies occur for

$$\delta_m = 0 \qquad \text{at maximum conductance} \tag{17.111}$$

$$\delta_x = \pm \frac{1}{2}\sqrt{K^2 - \frac{1}{Q_2{}^2}} \qquad \text{at loop crossover} \tag{17.112}$$

The corresponding values of conductance are

$$G_m = B_1\left(\frac{1}{Q_1} + K^2Q_2\right) \qquad \text{maximum value} \tag{17.113}$$

and

$$G_x = B_1\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \qquad \text{at crossover when crossover exists} \tag{17.114}$$

Because of the factor K^2 in the expression for maximum conductance and its nonappearance in the crossover value it is expected that the loop size is a sensitive function of the degree of coupling.

Some typical admittance curves in the presence of a coupled secondary resonant circuit are shown in Fig. 17.48 for fixed primary and secondary Q's, equal resonant frequencies, and different degrees of coupling. The transition from bulge to cusp to loop as the coefficient of coupling is increased is evident. It is seen that the loop can be eliminated by reducing the coupling. It can also be eliminated by decreasing the Q of the secondary circuit. Examination of Eq. (17.109) shows that the change in admittance introduced by the presence of the resonant secondary circuit depends only on the frequency parameter δ_2 and hence the shape of the resultant admittance does not depend upon the primary resonant frequency. The primary resonant frequency will determine only the position of the bulge, cusp, or loop and not its shape. In the usual case, where the secondary resonant frequency is fixed, the resultant bump on the resonator-admittance curve will move up as the resonator length is increased.

Resonances may be coupled into the line resonator in many ways. The resonator load may be resonant. Under certain conditions a resonant load may be connected to the cavity coupling loop through an unmatched line, in which case numerous loops may be induced in the resonator admittance. This is known as the "long-line effect." Sometimes the line plunger is not a perfect short, in which case back-cavity resonances may couple in admittance loops, which will cause blind spots or frequency jumps. When the resonant cavity is operating on a 0.75-wave-length resonance, there may occur a higher-order cavity resonance that will be coupled into the principal cavity resonance through virtually imperceptible asymmetries in the structure. The commonest higher order mode that may occur in cylindrical resonators is shown in Fig. 17.49 in both the actual and the developed form. This is a transverse electric mode for which the tube gap is a virtual short circuit, for voltage drops resulting from equal currents flowing in opposite directions cancel. Its resonant wave length is given approximately from the developed rectangular form by the formula

$$\left(\frac{2}{\lambda}\right)^2 = \frac{4}{\pi^2(r_2 + r_1)^2} + \frac{1}{l^2} \tag{17.115}$$

where r_2 and r_1 are the outer and inner radii of the concentric-line resonator, respectively, and l is its length. Resonance in this manner cannot exist until the length exceeds a half wave length and the mean circumference of the line exceeds a wave length. Tuning curves for both the desired cavity resonance and the undesired higher mode are shown in Fig. 17.50. For an equivalent cavity length of $\frac{3\lambda}{4}$ the desired line mode will have the shape shown and previously discussed. Its slope will be slightly greater than $4/3$ for the axes of Fig. 17.50. The first higher-order mode will have a curve that starts out with a slope of 2 and then becomes asymptotic to a wave length equal to the mean circumference of the line. It is inevitable that the two curves shown should intersect. For usual tube dimensions the intersection occurs at about 70 per cent of the maximum wave length of resonance of the higher-order mode. When the cavity is simultaneously resonant in both modes, a slight coupling between them through some asymmetry in construction will cause a loop to be induced in the resonator impedance with a resultant blind spot or frequency jump as line length or reflector voltage is varied. Such a blind spot is very difficult to eliminate. It should be pointed out

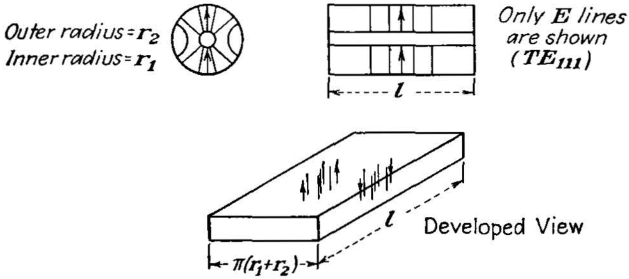


FIG. 17.49.—Commonest higher order resonance field of a concentric-line resonator.

that this type of blind spot *cannot* occur when the cavity is operated on a 0.25-wave-length resonance; for then the resonator tuning curve has a slope of about 4 on a plot like that of Fig. 17.50, and it is impossible for an intersection with the higher-order resonance to occur. This fact gives a great incentive for developing tubes which are small enough so that they can be operated on a 0.25-wave-length line resonance.[1]

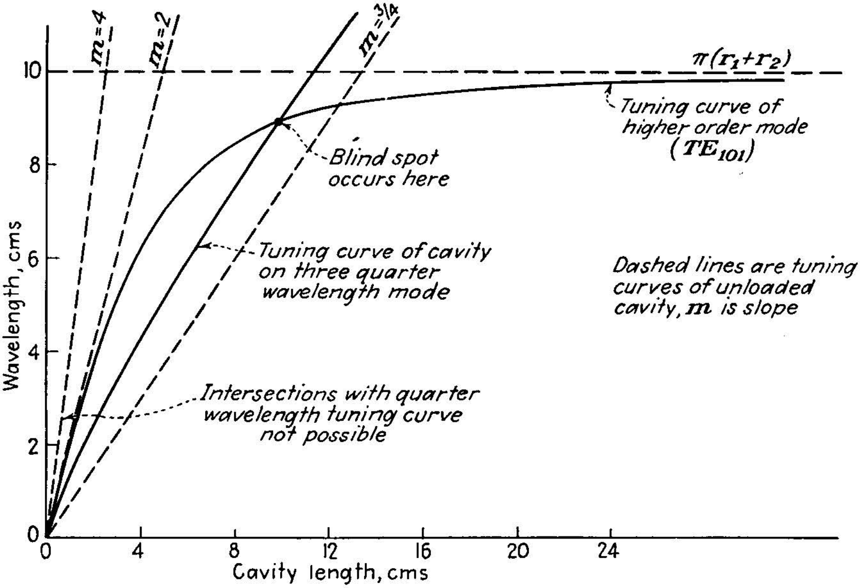


FIG. 17.50.—Tuning curves of a concentric-line cavity for the desired mode and the first higher order mode.

17.11. The Two-resonator Klystron Oscillator. A picture of an early type of two-resonator klystron has already been shown in Fig. 2.9. Modern tubes are quite similar except that the resonators are back to back so that coupling may be introduced through a set of coupling loops instead of by means of loops at the end of a transmission line of appreciable length. A schematic drawing of a two-resonator klystron oscillator and its equivalent circuit is shown in Fig. 17.51. The only difference between the amplifier and oscillator is that the oscillator has coupling between secondary and primary.

In the equivalent circuit shown in Fig. 17.51 several assumptions have been made in the interests of simplicity, all of which are justifiable.

---

[1] For further information on the subject of broad-band reflex-klystron oscillators see Chaps. 31 and 32, Vol. II of "Very High-Frequency Techniques," report of the wartime researches of the Radio Research Laboratory, Harvard University, McGraw-Hill, New York, 1947.

It has been assumed that the coupling between buncher and catcher resonators is purely inductive. This it is in many tubes, though in some there is coupling through a transmission line, which merely changes the phase of the mutual impedance from 90 deg to an arbitrary value. It has been assumed that all losses may be inserted in series with the resonator reactances. This is quite satisfactory if the proper conversion factors are always used and if the circuit Q's are greater than 20. Final answers will involve the Q's of the resonators and will be independent of whether the loss resistances are in series or in parallel with the resonator. The resistance in series with the buncher resonator includes the effect of the ohmic resonator losses and also the power required to bunch the beam. The resistance in series with the catcher resonator includes the ohmic losses of the resonator and also the load. The catcher resonator will ordinarily be more heavily loaded than the buncher resonator so that its Q will be lower.

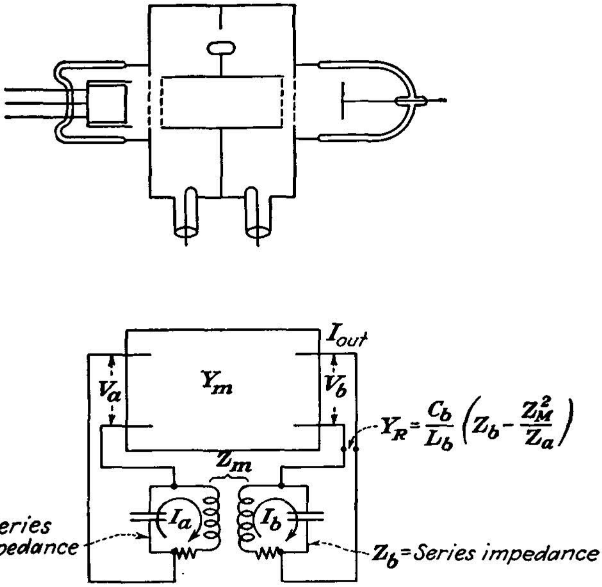


FIG. 17.51.—Schematic drawing of a two-resonator klystron oscillator and the equivalent circuit.

The analysis of the two-resonator klystron oscillator will proceed along the lines used for the reflex-klystron oscillator, though equivalent methods are just as satisfactory.[1,2] The fundamental requirement for

---

[1] WEBSTER, D. L., Theory of Klystron Oscillations, *Jour. Appl. Phys.*, vol. 10, pp. 864–872, December, 1939.

[2] HARRISON, A. E., Klystron Oscillators, *Electronics*, vol. 17, pp. 100–107, November, 1944.

oscillation in a klystron is the same as for any oscillator. This is to say that the transadmittance of the tube must equal the current-voltage reduction factor (transfer admittance) of the network

$$Y_m = Y_{ba} \tag{17.116}$$

where the transadmittance of the tube

$$Y_m = \frac{I_{\text{out}}}{V_a} \tag{17.117a}$$

has been given in Eq. (17.62) and the current-voltage step-down factor of the coupling circuit is

$$\frac{1}{Y_{ba}} = \frac{V_a}{I_{\text{out}}} \tag{17.1176}$$

under the condition that all circuit meshes are *closed*.

The beam transadmittance Y_m has a phase angle that depends only upon the electron transit angle between resonators, *i.e.*, only upon beam voltage. The magnitude, however, has a maximum value that is directly proportional to the transit angle, while the fraction of this value that is realized depends upon the buncher-resonator voltage V_a. Thus Y_m is a nonlinear admittance, decreasing with amplitude of exciting voltage. The circuit transfer admittance Y_{ba}, on the other hand, does not vary with electron transit angle, beam voltage, or amplitude of r-f voltage. It is a quantity that for a given adjustment of the resonators, coupling, and loading varies only with frequency. Hence for a fixed adjustment of the circuit the tube may be expected to go in and out of oscillation as beam voltage is varied, for this changes the phase of the beam transadmittance progressively. The resultant selective oscillation with respect to beam voltage has already been shown in Fig. 2.10. It may be expected that the two-resonator klystron can be analyzed by a method similar to that used for the reflex-klystron oscillator, the difference being that instead of equating resonator admittance and negative beam admittance for the condition of oscillation we now equate beam transadmittance and the circuit transfer admittance Y_{ba}.

In order to specify conditions of oscillation it will be necessary to know the form of the factor Y_{ba}. This is readily obtained by direct application of standard circuit theory. Let the series self-impedances of the input and output resonant circuits be Z_a and Z_b, these circuits including all the loss effects in the form of series resistance. The input and output impedances of the box representing the tube in Fig. 17.51 will be assumed infinite since the beam effects have been incorporated into the resistance of the input and output circuits. The output voltage is developed across an admittance Y_r, which includes the effect of the coupled input resonator and has the form

$$Y_r = \frac{C_b}{L_b}\left(Z_b - \frac{Z_m^2}{Z_a}\right) \tag{17.118}$$

where Z_m is the mutual impedance between the input and output circuits. In order to establish the relation between output current and input voltage let the sequence of current-voltage relations be traced backward from the input. The input voltage is related to the current flowing in the input resonant circuit by

$$V_a = j\omega L_a I_a \tag{17.119}$$

to a good degree of approximation for a high-Q circuit. The circulating current in the input resonator is related to that in the output resonator by

$$I_a = \frac{Z_m I_b}{Z_a} \tag{17.120}$$

The circulating current in the output resonator is related to the output current from the tube by

$$I_b = \frac{I_{\text{out}}}{j\omega L_b Y_r} \tag{17.121}$$

with sufficient accuracy for high-Q circuits. Putting these last four equations together to obtain the ratio of tube output current to the input voltage that it produces,

$$\frac{I_{\text{out}}}{V_a} = \frac{C_b}{L_a Z_m}(Z_a Z_b - Z_m^2) \tag{17.122}$$

Since $\omega L_b = \dfrac{1}{\omega C_b}$ in the vicinity of resonance and since the coefficient of coupling between buncher and catcher resonators is given by

$$K^2 = \frac{Z_m^2}{\omega^2 L_a L_b} \tag{17.123}$$

then

$$\frac{I_{\text{out}}}{V_a} = \frac{K^2}{Z_m^3}(Z_a Z_b - Z_m^2) \tag{17.124}$$

This ratio obtained from the circuit action must equal the ratio of output current to input voltage produced by the beam, *i.e.*,

$$Y_m = \frac{K^2}{Z_m^3}(Z_a Z_b - Z_m^2) \tag{17.125}$$

This is the important defining relation for oscillation, similar to the relation

$$Y_r = -Y_e \qquad (17.85)$$

that holds for reflex-klystron oscillators.

Conditions for oscillation can now be studied by examining the locus of beam transadmittance in the admittance plane and comparing with the locus of circuit transfer admittance. The beam-transadmittance locus will be exactly the same as the self-admittance spiral of the reflex-klystron oscillator *if the beam current does not change* as the beam voltage is varied. Under these conditions the locus of the beam transadmittance will be a spiral as the beam voltage is varied.

Another type of beam-transadmittance locus commonly occurs. If the cathode is *space-charge-limited*, the current will increase as the three-halves power of the beam voltage. For this condition the d-c beam conductance will be proportional to the square root of the beam voltage. Since the beam conductance is also directly proportional to the d-c transit angle, which depends upon the square root of the beam voltage, this means that the transadmittance magnitude will be independent of d-c transit angle for any given value of the bunching parameter. As a result, the locus of the beam transadmittance will be a *circle* as the beam voltage is varied. The power that the beam can deliver to the circuit will, however, vary as the beam voltage, and the maximum power will be proportional to the *cube* of the beam voltage.

The locus of the transadmittance through the circuit given by the right side of Eq. (17.125) is not so well known. It will be recognized, however, that the factor $Z_aZ_b - Z_m^2$ is one which appears frequently in coupled-circuit theory and is the one that contributes virtually all the variation of the transfer admittance in the vicinity of resonance. In the vicinity of resonance the factor Z_m is relatively constant compared with the factor $Z_aZ_b - Z_m^2$ and will be so considered. In particular, the factor $Z_aZ_b - Z_m^2$ appears in the denominator for the expression relating secondary current to series primary voltage in coupled-resonant-circuit theory.[1] It is this factor that gives rise to the well-known double-peaked response curves for couplings greater than critical. It would be expected therefore that the locus of $Z_aZ_b - Z_m^2$ itself would be such as to exhibit two minima for couplings greater than critical. This is readily shown to be the case.

To examine the defining relation for oscillation given by (Eq. 17.125) let the mutual impedance be assumed to be inductive of the form

$$Z_m = j\omega M \qquad (17.126)$$

---

[1] TERMAN, F. E., "Radio Engineering," 2d ed., p. 74, Eq. (42), p. 82, Eq. (45), McGraw-Hill, New York, 1937.

For small coupling factors, coupling of any other form will lead to the same result except that the phase of the coupling impedance may be different from 90 deg. With the above assumed form of the coupling impedance, Eq. (17.125) reduces to

$$Y_m = \frac{jK^2}{\omega^3 M^3} (Z_a Z_b + \omega^2 M^2) \tag{17.127}$$

If the input and output circuits have relatively high Q's then their series impedances can be represented by

$$Z_a = \frac{X_a}{Q_a} (1 + 2j\ \delta Q_a) \tag{17.128}$$

and

$$Z_b = \frac{X_b}{Q_b} (1 + 2j\ \delta Q_b) \tag{17.129}$$

where δ is the fractional deviation from resonance and it has been assumed that the resonant frequencies of the input and output resonators are the *same*. The above are simple linear approximations for the actual expressions but hold well enough in the vicinity of resonance. The effect of different input and output resonant frequencies will be discussed later. Using Eq. (17.125) and substituting Eqs. (17.128) and (17.129) into Eq. (17.127),

$$Y_m = \frac{j}{\omega M Q_a Q_b} [(1 + K^2 Q_a Q_b - 4 Q_a Q_b\ \delta^2) + j2\ \delta(Q_a + Q_b)] \tag{17.130}$$

The locus of the circuit transfer admittance is a simple parabola. The size of the parabola depends upon the Q factors and upon the coupling factor. The parabola is symmetrically disposed about the imaginary axis when the primary and secondary resonant frequencies are the same. For a given set of Q values the parabola will merely be shifted upward and have its curvature increased by an increase in coupling. For critical coupling

$$K_c^2 = \frac{1}{Q_a Q_b} \tag{17.131}$$

and smaller values of coupling, the vertex of the parabola will be the closest point to the origin. For values greater than critical coupling there will be two points symmetrically disposed about the vertex that are closest to the origin, while the vertex itself will be slightly farther away.

Some typical parabolic loci are shown in Fig. 17.52. These reveal all the well-known characteristics of tuned coupled circuits and some of the less well-known. Below critical coupling the transfer admittance is *large*, which means that there is a *small* buncher voltage for a large output current. It is convenient to talk in terms of the reciprocal magnitude of the transfer admittance since this gives the ratio of buncher voltage to catcher current. Assuming a constant catcher current, the buncher voltage increases as the coupling increases up to the critical value and is single-peaked as frequency is varied. As critical coupling is approached, the single resonant peak becomes broader, indicated by the decreased size of a constant-frequency interval on the admittance locus. Up to and including critical coupling the vertex of the parabola is the point on the parabola closest to the origin, which means that the frequency response will be single-peaked and that maximum buncher voltage will occur at the frequency of resonance of the resonant circuits. For couplings greater than critical, the vertex of the admittance locus recedes from the origin, but the parabolas become curved strongly enough so that two points symmetrically disposed on either side of the vertex are closest to the origin. This means that the input voltage is double-peaked as frequency is varied and that the response at the peaks is less than for critical coupling. This is a proper characteristic of coupled circuits. The peak response for couplings greater than critical will be equal to that at critical coupling only if the circuits are the same,

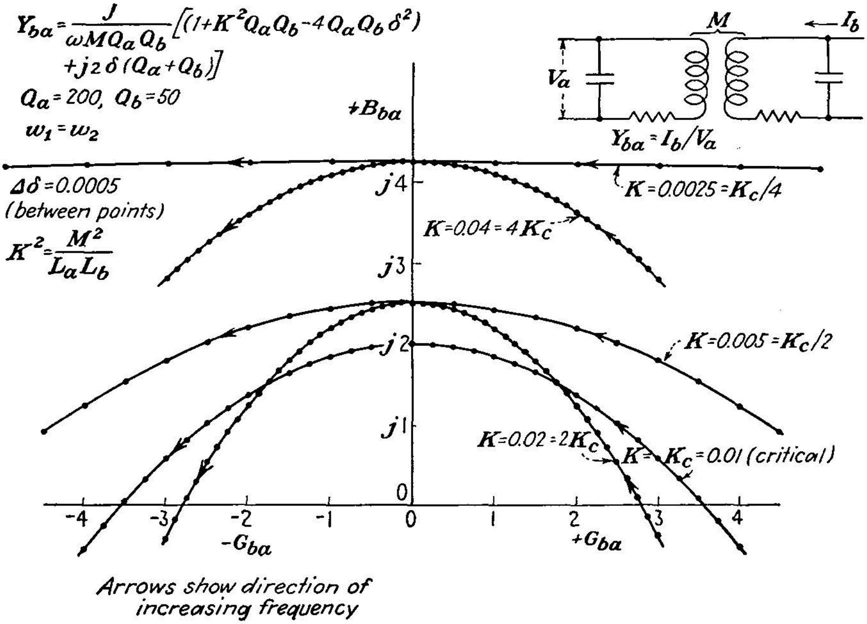


FIG. 17.52.—Parabolic transfer-admittance loci of a two-mesh coupled resonant circuit.

i.e., if the Q's are identical.[1] The peak response with greater than critical coupling drops off more rapidly with coupling as the dissimilarity of the two circuits is increased (same resonant frequency but different Q's). In addition to becoming double-peaked and smaller in peak amplitude as coupling increases, the response curves become broader with frequency, as may be seen from the decreased size of the constant-frequency interval. For klystron oscillators with the Q ratio given, it is seen that coupling will probably have to lie within a factor of two of the critical value in order to give appreciable output. The error made in using the linear approximations to the correct expressions is quite small. It is independent of the Q values and depends only upon the value of δ. The error is equal to three-halves of the δ value. Thus the maximum error for any of the loci shown in Fig. 17.52 is about two per cent.

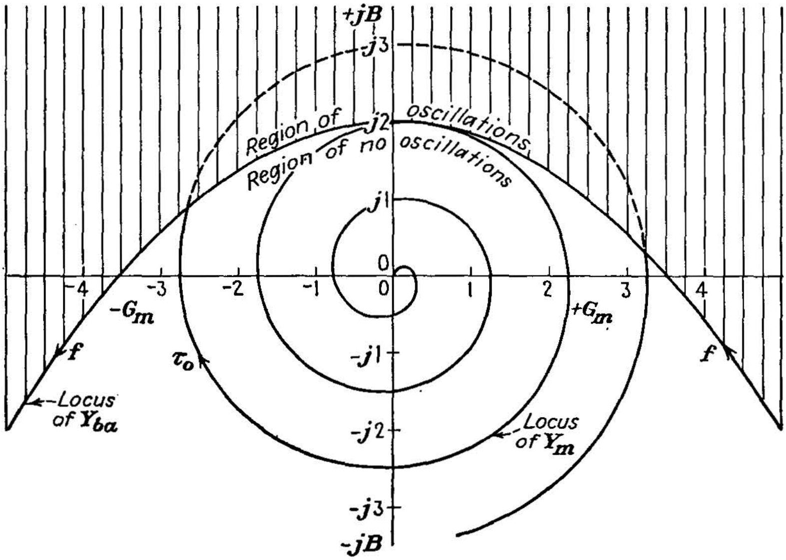


FIG. 17.53.—Two-resonator klystron-oscillator operation in terms of admittance loci.

The two-resonator klystron oscillator will oscillate whenever the beam-transadmittance locus enters the area outside the circuit-transfer-admittance parabola (the area outside of the parabola is that in which all the possible tangents to the parabola lie). This requires that the coupling between buncher and catcher circuits be great enough, that the beam transadmittance be large enough in magnitude, and that the phase angle be correct. Shown in Fig. 17.53 is the intersection of a beam-

---

[1] AIKEN, C. B., Two Mesh Tuned Coupled Circuit Filters, *Proc. I.R.E.*, vol. 25, pp. 230–272, February, 1937.

transadmittance spiral and the circuit-transfer-admittance parabola. Whenever the zero-signal beam transadmittance exceeds the circuit transadmittance in magnitude, oscillations will build up that will shrink the beam-transadmittance locus. This has the result that as beam voltage is varied the beam transadmittance will trace the circuit transfer admittance whenever the latter is smaller than the zero-voltage beam transadmittance. Oscillations will in general occur whenever the electron transit angle is in near equality with the phase of the circuit transfer admittance, *i.e.*, whenever

$$-\left(\tau_0 - \frac{\pi}{2}\right) = \phi + 2\pi n \qquad (17.132)$$

where ϕ is the angle of the circuit transfer and $n = 1, 2, 3$, etc. For the particular case shown in Fig. 17.53 this will give oscillations in the vicinity of

$$\tau_0 = 2\pi n \qquad (17.133)$$

In general, oscillations will occur for values of the transit angle differing by integral multiples of 2π.

For any one loop of the beam-transadmittance locus, relative power contours can be drawn, as for the reflex oscillator in Fig. 17.39. Positive-power-output contours will occur only in the upper half plane of Fig. 17.53, because as may be seen by comparing Eqs. (17.118) and (17.127) the circuit transfer admittance equals the output admittance multiplied by some numerical factors and rotated 90 deg in the counterclockwise direction. This means that the positive-conductance region of the output admittance appears in the upper half plane of the circuit-transfer-admittance plot. The relative power contours will have the same general shape as those of the reflex tube.

The beam-transadmittance locus is traced in a *clockwise* direction as the beam voltage is *decreased*. Hence, for a constant current, the magnitude of the beam transadmittance will decrease as the beam voltage is increased. This means that there is a highest voltage at which the tube will oscillate which occurs when the transit angle is so small that the magnitude of the beam transadmittance is reduced to the point where it does not intersect the circuit-transfer-admittance parabola. With actual tubes it may not be possible to reach this voltage without exceeding some design limitation of the tube.

The selective oscillation with voltage is shown in Fig. 2.10. Output-power loops as shown in Fig. 2.10 may be single- or double-peaked. If the resonator coupling is much below critical, the mode loops will generally be single-peaked. However, with critical coupling the output-power pulses may be either single- or double-peaked. In general, if the vertex of the circuit-transfer-admittance parabola is exceeded only slightly in magnitude by the zero-signal beam transadmittance, the power loops will be single peaked. If the vertex of the circuit-transfer-admittance parabola is greatly exceeded in magnitude by the zero-signal beam transadmittance, the output loops will generally be strongly double-peaked. When the resonator coupling is considerably above critical, the power-output loops will generally be double-peaked.

As the beam voltage is changed during a condition of oscillation, the frequency of oscillation will change slightly. The curves of frequency deviation will resemble the curves of phase shift through the coupled resonant circuits as a function of frequency. When the resonator coupling is critical or less, the frequency deviation will be nearly linear with transit angle or beam voltage everywhere but at the edges of the oscillation modes. The frequency will increase as the beam voltage increases. When the resonator coupling is in excess of the critical value, there will be a strong kink in the frequency-deviation curve near the middle of the mode due to the fact that the frequency changes more rapidly there. Of interest is the rate of frequency change at mid-mode with beam voltage. By a method similar to that used in obtaining Eq. (17.93) it is readily shown that

$$\frac{df}{f_0} = \left(\frac{1 + K^2 Q_a Q_b}{Q_a + Q_b}\right) \frac{\tau_0}{4} \frac{dV_0}{V_0} \tag{17.134}$$

at the mid-mode. The frequency stability will ordinarily be of the order of tens of kilocycles per volt.

If all operating conditions of the two-resonator klystron oscillator but the magnitude of the beam current are kept constant, it will be found that there is a minimum beam current which will sustain oscillations. The limiting condition for oscillation is that at which the beam-transadmittance spiral is just tangent to the circuit-transadmittance parabola. This point of tangency will be near the vertex of the parabola for all cases except coupling greatly in excess of critical. For the usual conditions the limiting condition of oscillation, from Eqs. (17.60) and (17.130),

$$\frac{G_0 A^2 \tau_0}{2} = \frac{1 + K^2 Q_a Q_b}{\omega M Q_a Q_b} \tag{17.135}$$

from which the minimum current that will sustain oscillations is

$$I_{\min} = \frac{2}{A^2 \tau_0} \frac{1 + K^2 Q_a Q_b}{\omega M Q_a Q_b} V_0 \tag{17.136}$$

Two-resonator klystron oscillators have been built for frequencies ranging from 600 to 4,000 mc. In power output they have ranged from a fraction of a watt to one kilowatt. In the United States the development of these tubes has been pioneered by the Sperry Gyroscope Company, which has specialized in tubes with evacuated resonators. European tube developers have favored tubes with cavities external to the evacuated parts of the tube.

17.12. The Heil Tube. Historically, the first tube to use the velocity-modulation principle was one proposed by the Heil brothers.[1] A diagram of such a tube is shown in Fig. 17.54. The tube makes use of a beam of electrons, which are shot across the ends of a concentric-line resonator. Each electron thereby follows a path along which r-f voltage appears twice. The voltage across the second gap is instantaneously 180 deg out of phase with the voltage across the first gap. After crossing both gaps the beam electrons are taken out of action by a collector electrode. The tube operates by bunching the electron beam in the first gap and extracting energy from the bunched electrons in the second gap. As with the reflex-klystron oscillator the tube will resonate when the transit angle between gap crossings is $n + \frac{3}{4}$ cycles and if the resonator admittance is smaller in magnitude than the beam conductance. The analysis of the Heil tube proceeds along exactly the same lines as does that of the reflex tube. Some excellent Heil tubes have been made, but they seem to have lost the application race in competition with the reflex-klystron oscillator. This is due to a number of reasons prominent among which are the following: The transmission-line type of resonator is not quite as efficient at superhigh frequencies as is the reentrant-cavity type of resonator; the tube is not so well adapted to an external resonator. Against these disadvantages, the Heil tube is superior to the reflex-klystron oscillator in that multiple electron transits are avoided and that tubes can be built to which magnetic beam focusing is easily applied.

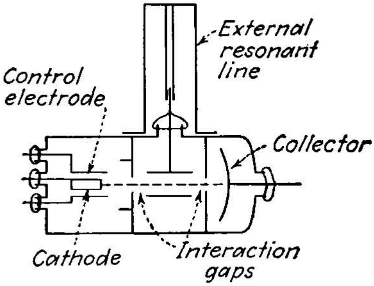


FIG. 17.54.—Diagram of the Heil tube.

17.13. Bunching Effects in Negative-grid Tubes. The analysis of klystron tubes by means of simple bunching theory is so enlightening that the question is raised whether the action of negative-grid tubes cannot be explained in similar terms. Certainly, when transit times are large and voltages are not excessive, there will be a bunching action occurring in the electron stream of negative-grid tubes. An examination of this

---

[1] HEIL and HEIL, *op. cit.*; see also HAHN, W. C., and G. F. METCALF, Velocity Modulated Tubes, *Proc. I.R.E.*, vol. 27, pp. 106–116, February, 1939.

bunching action should reveal some relations that will tie in with the observed action of such tubes at ultra-high frequencies.

In attempting to analyze the action of negative-grid tubes in terms of velocity modulation one immediately encounters the difficulty of accounting properly for space-charge effects. However, even though the neglecting of space-charge effects is bound to lead to large errors, it is instructive to examine tube behavior under the assumption of their absence. Let it be assumed that the potential variation from cathode to grid plane and from grid plane to plate is linear. Let it further be assumed that bunching action occurs only in the grid-cathode region, a reasonable assumption since the a-c components of field will be larger here and since the time spent in this region will be larger too. Let it further be assumed that alternating components of voltage appear only on the grid and that these are small compared with the direct potentials involved. The final assumption is that the emission is directly proportional to the equivalent grid voltage $V_g + \frac{V_p}{\mu}$. In addition to space charge, the displacement components of current are neglected.

If with the above assumptions an analysis of the electron action is made, it is possible to solve for electron arrival time at the plate in terms of its departure time from the cathode. Plate current can then be solved for by allowing for variation of emission over the cycle. The resulting expression for plate current is expected to be a function of the amplitude of grid voltage and of the transit angles involved. It should reduce properly to approximate expressions for current flow for negligible transit angles and should exhibit some bunching effects.

The resulting expression for plate current obtained by neglecting all but first- and second-order terms in grid voltage is

$$I_{\text{in phase}} = U \cos \tau_{cp} + V \sin \tau_{cp} \qquad (17.137a)$$

$$I_{\text{out of phase}} = U \sin \tau_{cp} - V \cos \tau_{cp} \qquad (17.137b)$$

where the phase is taken relative to the grid voltage and τ_{cp} is the cathode-plate transit angle and where

$$U = -2G\left(\frac{V_b}{\mu} + V_c\right) J_1\left(\frac{\alpha^2 \tau_{cg}{}^2}{6}\right) J_0\left(\frac{1}{2}\alpha\tau_{cg}\right) + GV_g J_0\left(\frac{\alpha^2 \tau_{cg}{}^2}{6}\right) J_0\left(\frac{1}{2}\alpha\tau_{cg}\right) \qquad (17.138)$$

and

$$V = 2G\left(\frac{V_b}{\mu} + V_c\right) J_0\left(\frac{\alpha^2 \tau_{cg}{}^2}{6}\right) J_1\left(\frac{1}{2}\alpha\tau_{cg}\right) - GV_g J_1\left(\frac{\alpha^2 \tau_{cg}{}^2}{6}\right) J_1\left(\frac{1}{2}\alpha\tau_{cg}\right) \qquad (17.139)$$

where G is mutual conductance, V_b and V_c are d-c components of plate and grid-plane potentials, α is the ratio of a-c to d-c grid-plane potential, τ_{cg} is cathode-grid transit angle, V_g is a-c component of grid-plane potential, and J_0 and J_1 are the zero- and first-order Bessel functions of the first kind. From the above the expression for magnitude of plate current can be written

$$|I| = \sqrt{U^2 + V^2} \tag{17.140}$$

and the corresponding phase by which plate current lags grid voltage is

$$\phi_p = \tau_{cp} - \arctan\left(\frac{V}{U}\right) \tag{17.141}$$

The above expressions reveal much about the nature of the plate current as affected by the transit time and the bunching action. In the first place the magnitude of plate current as given by Eq. (17.140) is independent of the cathode-plate transit angle and depends only upon the magnitude of the a-c component of grid voltage and the cathode-grid transit angle. The dependence is partly in terms of the well-known first-order bunching parameter $\frac{1}{2}\alpha\tau_{cg}$ and also in terms of the second-order bunching parameter $\frac{\alpha^2\tau_{cg}^2}{6}$. The components of plate current given by Eqs. (17.137*a*) and (17.137*b*) each contain two types of terms. The first term gives the a-c components of current that would result if the emission were constant over the cycle and the plate current were produced only by bunching action. These terms properly reduce to zero for zero transit angle. The second terms give the current components resulting from the normal grid action but reduced by the dispersing effect of the bunching action. The normal-current terms are correctly maximum for zero transit angle and drop off in magnitude as the transit angle increases. The bunching terms initially are zero, increase, and then decrease again.

The nature of the plate current is best illustrated by some specific examples. In Fig. 17.55 are shown the magnitude and phase of plate current for a typical tube under the conditions that $\alpha = 0.2$ and $\tau_{cp} = 2\tau_{cg}$. The magnitude is seen to drop off more or less gradually. Undoubtedly this should be a smooth curve, but it exhibits some small ripples because higher-order effects have been neglected. The phase of the current initially is -180 deg and then drops back progressively as the cathode-plate transit angle is increased. The change of phase is always less than the change of cathode-plate transit angle but has its principal dependence upon this angle.

In Fig. 17.56 is given another set of curves showing magnitude and phase of plate current for the conditions that $\alpha = 0.5$ and $\tau_{cp} = \frac{4\tau_{cg}}{3}$.

This is a rather large value of α for bunching relations but is not too large in this case because second-order effects have been included. The magnitude in this case exhibits a striking decrease with transit angle and then a slight increase. The locus of plate current in amplitude and phase is again a form of spiral, but with the second turn not enclosing the origin. As before, the magnitude depends only on the cathode-grid transit angle, and the phase has its principal dependence upon the cathode-plate transit angle, though the actual phase is always somewhat less than this value. The plate-current spirals, which are really transadmittance spirals, bear a striking resemblance to the transadmittance spirals for klystron tubes.

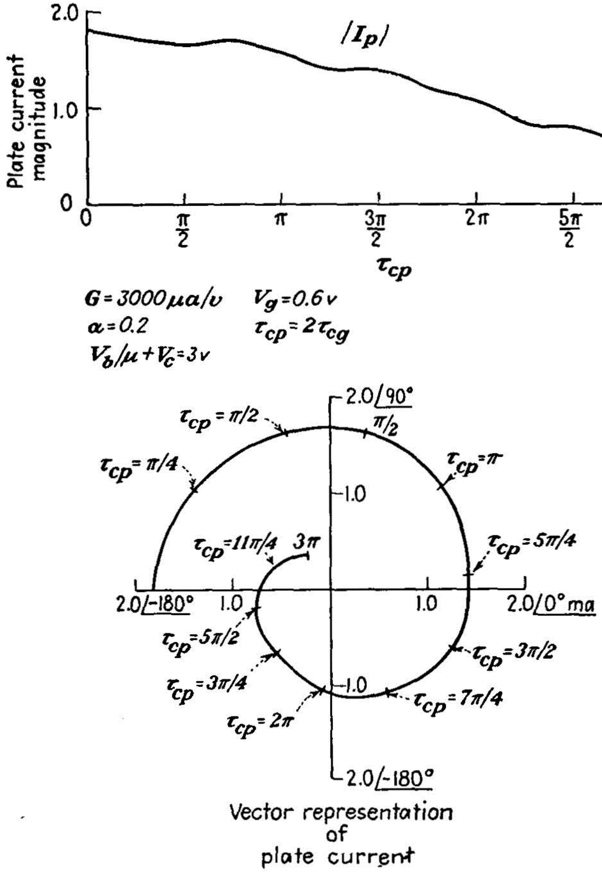


FIG. 17.55.—Triode plate current as a function of transit angle for relatively small excitation.

The above analysis must not be taken too seriously because it has neglected space-charge and displacement currents. However, the *nature* of the variations in plate current is quite possibly not too different from that shown in Figs. 17.55 and 17.56. Correlations with reduction in amplifier output with increasing frequency and reduction of oscillator output as a result of both reduction in magnitude of current and change in phase are evident.

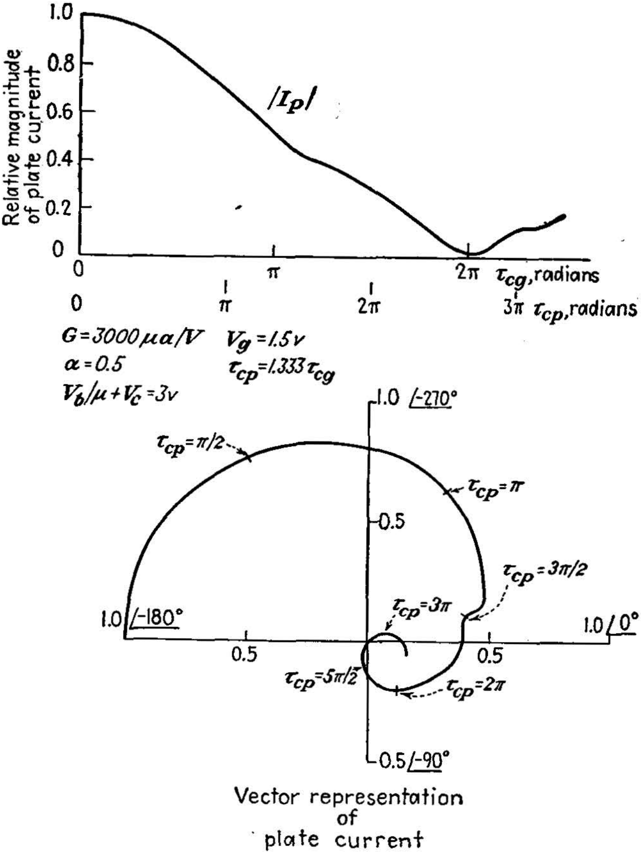


FIG. 17.56.—Triode plate current as a function of transit angle for moderate excitation.

CHAPTER 18

MAGNETRON OSCILLATORS

18.1. Introduction. A brief description of the magnetron has already been given in the chapter on Basic Tube Types. Basically, the magnetron is a tube containing a cathode and what is usually a symmetrical distribution of anodes in which electrons move under the influence of an internal electric field and a crossed externally supplied static magnetic field. The electrons move in complicated curved paths, and under certain conditions powerful oscillations will be sustained.

The magnetron underwent a tremendous development during the Second World War. Its development made possible the numerous microwave radars, which used it as a source of extremely high power pulses in the frequency range of 700 to 24,000 mc. It has amazed everyone by its efficiency, relatively high for an electronic device. Efficiencies are of the order of 50 to 80 per cent, and these are obtained at reasonable values of current, voltage, and magnetic field. Furthermore, the physical dimensions of magnetrons are of the order of the wave length, so that even the highest frequency magnetrons are not too hard to build.

A brief study of electronic motion of the type encountered in magnetrons has already been made in Chap. 6. Here it was found that electrons in combined electric and magnetic fields will move in strongly curved paths with periods of rotation corresponding to microwave frequencies. Thus, an electron in a uniform magnetic field of 1,070 gausses rotates in a circular path at a frequency of 3,000 mc per sec. By the use of multisegmented anodes, oscillations at frequencies higher than that corresponding to the simple rotation can be obtained. Inherently, then, electron motion in magnetic fields is of the right nature to produce microwave oscillations.

Many kinds of magnetron tubes can be built, ranging from a single-anode tube to multisegment-anode tubes. Many kinds of oscillations are also possible. The nature of the electron paths is such that a negative resistance can be obtained from the division of current between anode segments. This leads to negative-resistance oscillations. This type of oscillation is effective only at low frequencies and is no longer considered of great importance. Early work in the field was largely confined to oscillations of two- and four-segment-anode tubes involving electron-transit-time characteristics. Later development showed that such tubes were the least efficient of the entire class of magnetrons, and they are now not of much interest except for special applications.[1] Present interest is concentrated mainly in electronic oscillations of multisegment cavity magnetrons, and most of the comments in this chapter will be restricted to this case.

The complete theory of magnetron operation is not known at present. Wartime developments were largely of an empirical sort. Even a large fraction of the present information available on the subject is based upon specific calculations and tests.[2] It will probably be some time before the great mass of information on this subject is reduced to a simple organized treatment. The most complete organization of this sort appears in the report of the wartime researches of the Radiation Laboratory.[3] The complete story will, of course, have to be written by the men who are responsible for most of the recent development and by their successors who will carry on this work. There will be given in this chapter only those fundamental relations which are fairly well established. This will obviously be insufficient as a basis for the design of tubes but will serve as an introduction to the subject, which will furnish a basis for understanding the detailed reports on this subject.

18.2. Structural Form of Magnetrons. All magnetrons have in common a cathode, an anode, and an output-coupling device. In addition, magnetrons may have tuning mechanisms, mode suppressors, and end plates. Early two- and four-segment magnetrons were housed in glass envelopes with the cathode in the form of a tungsten filament and the anode segments supported from a two-wire transmission line brought out through the end of the tube opposite to that at which the filament leads were brought out. In some tubes, special end plates supported from leads brought out at the filament end of the tube were used to remove out-of-phase electrons from the cathode-anode region.[4,5]

Modern multicavity magnetrons are housed in metal and use glass

---

[1] Bibliographies of magnetron articles prior to 1941 are given in "High Frequency Thermionic Tubes" by A. F. Harvey, Wiley, New York, 1943, and "Einführung in der Theorie und Technik der Dezimeterwellen" by O. Groos, Herzel, Leipzig, 1937.

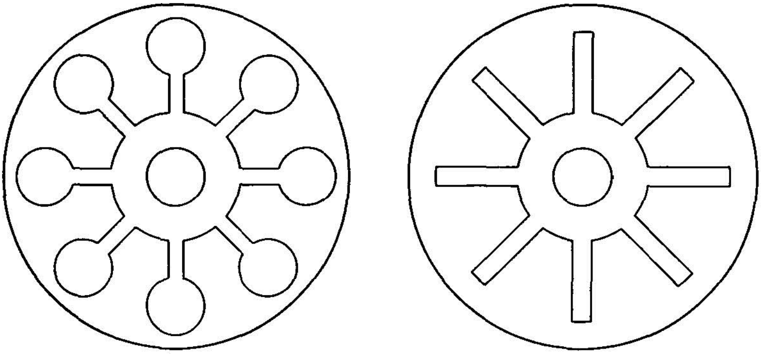
[2] Fisk, J. B., H. D. Hagstram, and P. L. Hartman, The Magnetron as a Generator of Centimeter Waves, *Bell Sys. Tech. Jour.*, vol. 25, pp. 1–188, April, 1946.

[3] *Radiation Laboratory Series*, 28 Volumes, McGraw-Hill, New York, 1947–1948.

[4] Linder, E. G., Description and Characteristics of End Plate Magnetrons, *Proc. I.R.E.*, vol. 24, pp. 633–653, April, 1936.

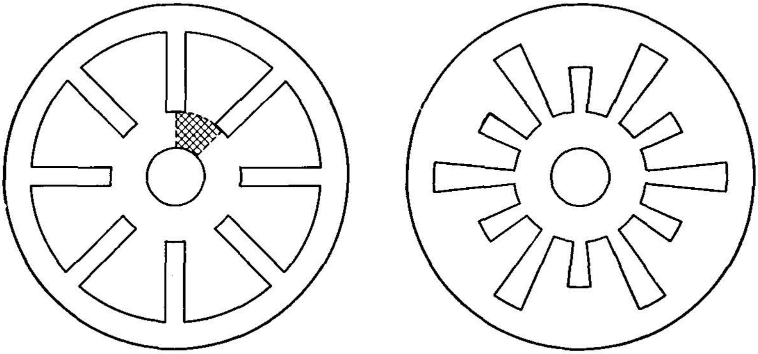
[5] Linder, E. G., Anode Tank Circuit Magnetron, *Proc. I.R.E.*, vol. 27, pp. 732–738, November, 1939.

only around the high-voltage filament leads and at the point where the output power is taken from the tube. The multisegmented anode is commonly formed of laminations having one of the forms shown in Fig. 18.1. Each of these consists in effect of a number of parallel resonant circuits in series around the inner circumference of the anode. In form *a* the individual resonant circuits are nearly lumped, *i.e.*, there is a capacity across each gap in parallel with the inductance formed by the inner surface of the circular hole. Actually, such a circuit is not truly lumped, for the dimensions of the various parts may be an appreciable fraction of a wave length long. In other forms of the anode the resonant circuit consists of a shorted section of strip transmission line.



(*a*)
Hole and slot type

(*b*)
Slot type



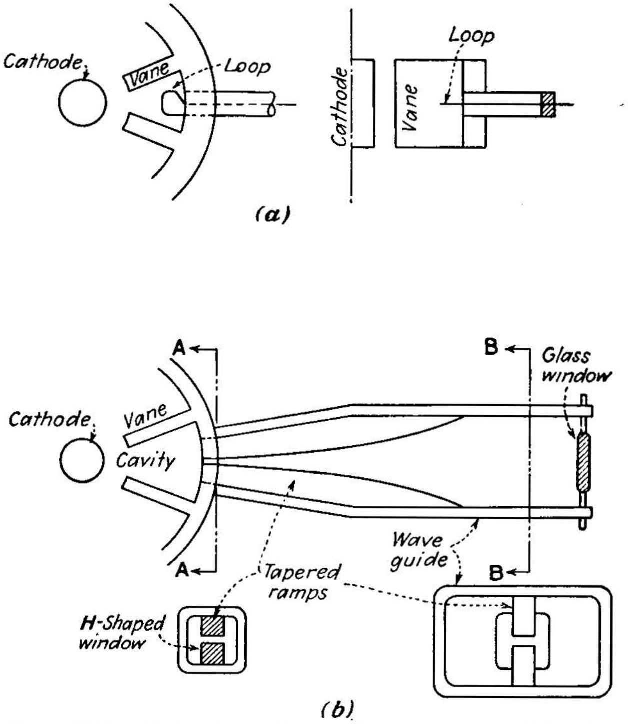
(*c*)
Vane type

(*d*)
Rising sun type

Fig. 18.1.—Various forms of multisegment anodes.

The cathodes of multicavity magnetrons are usually of appreciable diameter and in tubes for pulsed operation are indirectly heated and make use of oxide emitters. The cathode is usually supported by the filament leads, which are brought out at right angles to the axis of the tube. The cathode usually extends on each end about 25 per cent in length beyond the stack of anode laminations. Great precautions are taken to insulate the cathode leads for the high voltage that the tube must withstand; often the cathode lead insulator takes up about one-third the volume of the tube.

The output-coupling device in a multicavity magnetron is commonly a loop located at the base of one of the resonant radial anode spaces and



(a)

(b)

FIG. 18.2.—Output-coupling schemes for multicavity magnetrons.

leading out of the tube through a concentric line with a vacuum glass seal. Such an arrangement is shown in Fig. 18.2*a*. In some tubes the output coupling is accomplished by means of a tapered transmission line feeding from a narrow slot at the base of one of the radial resonant spaces and leading to a wave-guide section, with the vacuum seal effected by a window at the end of the guide section as shown in Fig. 18.2*b*. Numerous variations of these two basic schemes, including aperture coupling to a wave guide, form the bulk of the output-coupling arrangements.

Multicavity magnetrons will frequently have iron pole pieces built into them, with the iron brought close to the cathode and arranged so that a magnetic field parallel to the cathode is created. The pole pieces are brought to the external surface of the tube so that a magnet can be attached external to the vacuum. In some tubes the cathode leads are brought out axially through a hole in the iron pole pieces.

In addition, there are frequently straps interconnected between the anode pole pieces in order to separate the various natural resonant frequencies of the resonant circuit. Various tuning devices are also used. These will be described in connection with the resonant properties of the multicavity circuit.

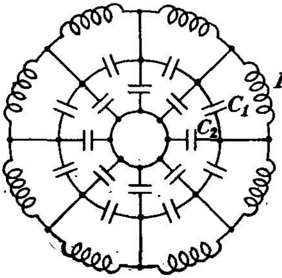


FIG. 18.3. — Approximate equivalent circuit of the magnetron of Fig. 18.1a.

18.3. Resonant Properties of Multicavity Magnetrons. The resonant systems shown in Fig. 18.1 will have a series of natural resonant frequencies. These frequencies are most properly determined by an analysis of the electromagnetic fields of the system. However, since most engineers are more familiar with circuits than with fields, a partial approximate analysis will be made in terms of some equivalent circuits. It must be recognized, however, that the determination of suitable equivalent circuits depends originally upon a knowledge of the fields.

The anode-cathode arrangement of Fig. 18.1a is, at first glance, expected to have the equivalent circuit shown in Fig. 18.3. The capacity C_1 represents the capacity between a pole face and the anode. The capacity C_2 represents the capacity between two adjacent pole faces. The inductance L_1 is the inductance of the inner surface of the circular hole. Actually, this is a poor equivalent circuit, for it neglects transmission-line effects and the large mutual inductance between adjacent anode spaces. It will, however, serve as a basis for an initial discussion. Let the circuit be developed by unwrapping the structure to give the arrangement of Fig. 18.4. This is seen to be a low-pass filter. As such, it will have a pass band in which the attenuation is substantially zero and in which there is a phase shift per section which increases uniformly from zero at zero frequency to π radians per section at cutoff. When the total phase shift along the series of N pole faces and hence N

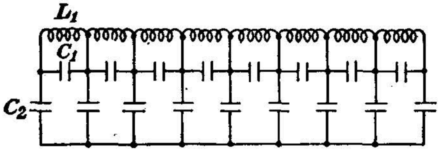


FIG. 18.4.—Developed form of the equivalent circuit of the magnetron of Fig. 18.1a.

sections is any integral multiple of 2π, then standing waves can exist in the circular arrangement, which is merely the developed form connected onto itself. The actual resonant fields are formed by two waves of equal amplitude traveling in opposite directions around the cathode. Analytically, this occurs whenever

$$\beta = \frac{2n\pi}{N} \tag{18.1}$$

where β is the phase shift per section in radians, n is the number of cycles of a traveling wave around the cathode, commonly referred to as the *mode number*, and N is the number of gaps or pole faces. When n equals $\frac{N}{2}$, then the phase shift per section is π radians and the fields will reverse at adjacent gaps. This mode is called the π *mode* and is the one ordinarily used in multicavity magnetrons. In Fig. 18.5 are shown the fields in a multicavity magnetron of eight segments for n equal to 1 and 4, the latter being the π mode. The resultant field has the properties of a standing wave, *i.e.*, it remains stationary and only varies in magnitude periodically with time. Such a standing wave can, of course, be resolved into traveling waves, and it is the study of the interaction of the electrons with one of the traveling-wave components that leads to the best picture of magnetron operation.

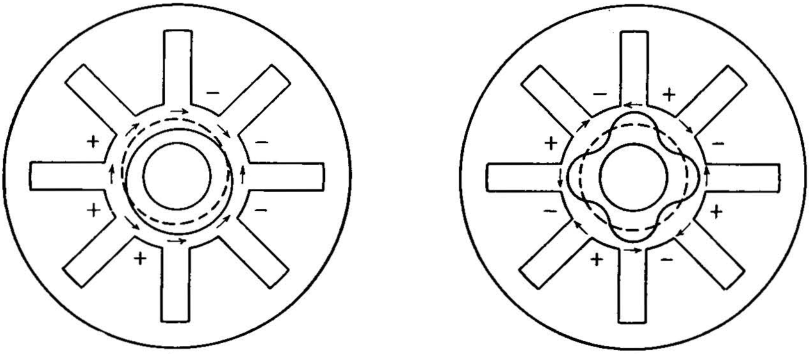

(a) (b)

FIG. 18.5.—Fields in multicavity magnetrons.

The phase-shift function of the circuit of Fig. 18.4 can be evaluated by applying Campbell's formula.[1] In the pass band this has the form

$$\cos\beta = 1 + \left(\frac{Z_1}{2Z_2}\right) \tag{18.2}$$

---

[1] EVERITT, W. L., "Communication Engineering," 2d ed., p. 173, McGraw-Hill, New York. 1937.

where β is the phase shift per section in radians, Z_1 is the total series impedance per section, and Z_2 is the total shunt impedance per section. In this case

$$Z_1 = \frac{\dfrac{L_1}{C_1}}{j\omega L_1 + \dfrac{1}{j\omega C_1}} \tag{18.3}$$

which reduces to

$$Z_1 = \frac{j\omega L_1}{1 - \dfrac{\omega^2}{\omega_1{}^2}} \tag{18.4}$$

where ω_1 is the angular resonant frequency of the parallel L_1C_1 combination equal to $\dfrac{1}{\sqrt{L_1C_1}}$ in magnitude. The shunt impedance is given by

$$Z_2 = \frac{1}{j\omega C_2} \tag{18.5}$$

Making the indicated substitutions into Eq. (18.2),

$$\cos \beta = 1 - \frac{\dfrac{1}{2}\dfrac{\omega^2 C_2}{\omega_1{}^2 C_1}}{1 - \dfrac{\omega^2}{\omega_1{}^2}} \tag{18.6}$$

If now Eq. (18.1) is invoked and the above equation solved for $\left(\dfrac{\omega}{\omega_1}\right)^2$, there is obtained

$$\left(\frac{\omega}{\omega_1}\right)^2 = \frac{1}{\dfrac{C_2}{2C_1\left[1 - \cos\left(\dfrac{2\pi n}{N}\right)\right]} + 1} \tag{18.7}$$

where ω is now the angular resonant frequency corresponding to a given value of n and N. The resonant frequencies for the assumed circuit will have the form shown in Fig. 18.6. The important observation about Fig. 18.6 is that the frequency of the π mode is not very different from the next resonant frequency. This is a bad situation and cannot be tolerated if the frequency separation is too small. A 1 per cent frequency separation is poor and will give trouble from the oscillation jumping to the adjacent frequency. A 3 per cent frequency separation is marginal. A 15 per cent frequency separation is good.

The above analysis is not very satisfactory, for it neglects the mutual

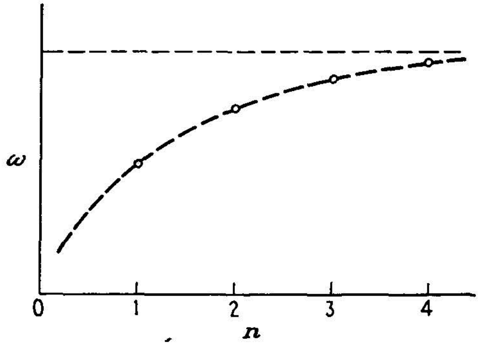


FIG. 18.6.—Resonant frequencies of the circuit of Fig. 18.3.

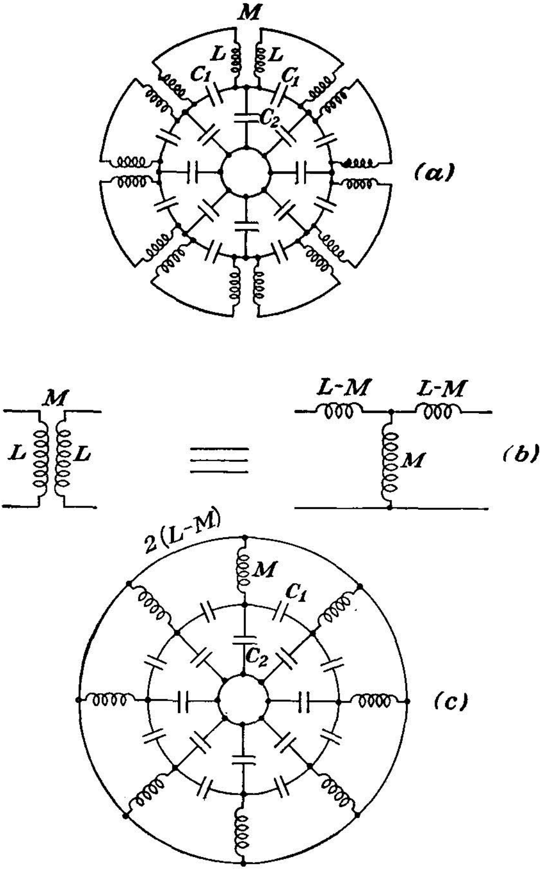


FIG. 18.7.—Equivalent circuits of multicavity magnetrons including the mutual inductance between slots.

inductance between adjacent slots, which is expected to be quite high. The magnetic flux lines are parallel to the axis of the tube in the slots. At the edges of the slots the magnetic flux lines divide and return through the adjacent slots. Accordingly, the equivalent circuit is expected to look like that shown in Fig. 18.7*a*. The ratio of the number of magnetic flux lines returning through adjacent slots to the total number will be nearly unity, which means that the coupling is nearly unity. The mutual inductances of Fig. 18.7*a* can be replaced by the T-section equivalent of Fig. 18.7*b*.[1] This allows the circuit of Fig. 18.7*a* to be represented as in Fig. 18.7*c*. This circuit has the characteristics of a band-pass filter, the phase shift at the low-frequency cutoff being π radians per section. For this circuit the resonant frequencies are given by

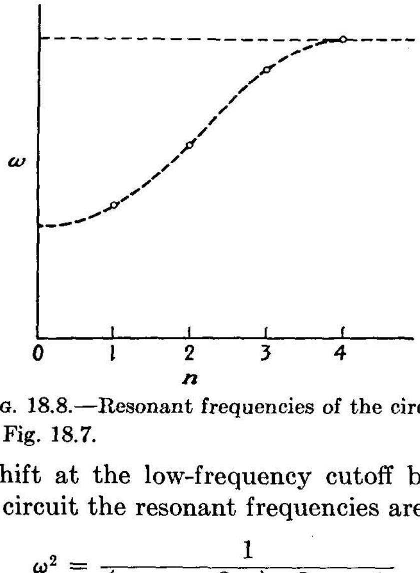


FIG. 18.8.—Resonant frequencies of the circuit of Fig. 18.7.

$$\omega^2 = \frac{1}{\left(1 - \cos\frac{2\pi n}{N}\right)\frac{2}{\omega_1^2} + \frac{1}{\omega_2^2}} \tag{18.8}$$

where $\omega_1^2 = \frac{1}{MC_1}$ and $\omega_2^2 = \frac{1}{MC_2}$. The relative disposition of the resonant frequencies for the different mode numbers is shown in Fig. 18.8. Again the frequency separation between the π mode and its neighbor is very small.

Actual magnetrons will have characteristics between those corresponding to the two cases discussed, the behavior more frequently corresponding to a band-pass filter and the resonant frequency of the π mode occurring slightly above the low-frequency cutoff.

The difficulties associated with closely spaced resonant frequencies

---

[1] EVERITT, *op. cit.*, p. 232.

can be greatly reduced by strapping alternate pole tips together. The commonest form of strapping, known as "double-ring strapping," is shown in Fig. 18.9. In this arrangement two rings are run around the pole tips. Each ring is connected to alternate pole tips, one ring being connected to all the odd-numbered pole tips and the other to all the even-numbered pole tips. In the π mode of resonance, alternate poles are 180 deg out of time phase with each other. As a result, the straps will be 180 deg out of phase with each other, and thus the capacity between the straps is added in parallel with the capacity C_1 in the equivalent circuits. From Eq. (18.8) this is seen to *lower* the resonant frequency of the π mode. Because of the symmetry and phasing no current will flow in the straps at the resonance frequency of the π mode. For other modes, the phase shift between adjacent poles is not 180 deg, and so currents will flow in the straps. This effectively puts more inductance in parallel with the inductance of the slots and so *raises* the frequency of the adjacent resonances. Both the capacity and inductances thus combine to increase the frequency separation between the π-mode resonance frequency and the adjacent resonant frequency. The shorter and heavier the strap segments, the more heavily strapped the magnetron is said to be. Another device sometimes used to increase the frequency separation between the desired resonance and its neighbor is that of making alternate anode slots of different length, as shown in Fig. 18.1*d*. Magnetrons using this arrangement are designated as being of the *rising-sun* type. By proper proportioning of the lengths a very good frequency separation can be achieved.

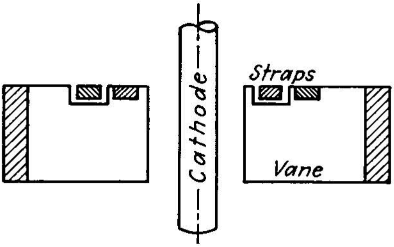


FIG. 18.9.—Double-ring strapping.

All the multicavity-magnetron resonances correspond to standing waves formed by two traveling waves of equal magnitude moving in opposite directions. These traveling waves will have radial components of electric field that are strongest at the plate and drop off somewhat toward the cathode. They will also have tangential or angular components of electric field that are very strong at the anode gaps and drop off very rapidly toward the cathode. For the plane-electrode case of Fig. 18.15 the tangential component of electric field will drop off exponentially from anode to cathode. In the cylindrical case the tangential component will vary approximately as the *n*th-order Bessel function of the radius, where *n* is the mode number.

Multicavity magnetrons may be tuned over an appreciable range by changing either the slot capacity or the inductance. One way of doing this is to use a set of plugs, in a crown-shaped arrangement, that are dropped into the slots at the appropriate point. If the plugs are inserted near the interaction gap of the anode slots, the capacity of the resonant slots will be increased and the frequency decreased. If the plugs are inserted near the base of the anode slots, the inductance will be decreased and the frequency increased. An arrangement using both an L ring and a C ring for tuning is shown in Fig. 18.10. Where a very broad range is desired, both an L and a C ring are used. In this case the rings are ganged so that the L ring enters as the C ring emerges. For a narrow tuning range a capacity tuning alone suffices. This tuning arrangement admits of a great variety of forms. Numerous other methods of changing the slot capacity or inductance are also used.

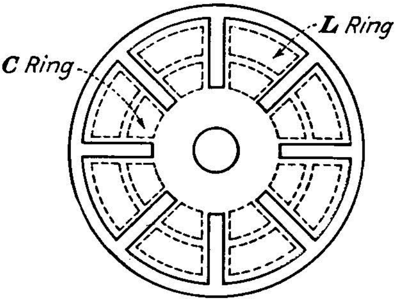


FIG. 18.10.—Inductive and capacitive tuning rings.

18.4. Electron Behavior in Crossed Static Magnetic and Static Electric Fields: Plane Case. The behavior of electrons in combined electric and magnetic fields has already received a brief treatment in the chapter on Laws of Electron Motion. This subject will be reviewed and extended here.

First review the behavior of an electron moving at right angles to a uniform magnetic field with a flux density B in the absence of an electric field. The electron in this case will describe a circular path whose radius will be

$$R = \frac{3.372 \times 10^{-6} V^{1/2}}{B} \qquad \text{meters} \tag{18.9}$$

where V is the potential in volts through which the electron has been accelerated and B is the magnetic-flux density in webers per square meter (10^4 gausses). If this relation is converted to practical cgs units, it becomes

$$R' = \frac{3.37 V^{1/2}}{B'} \qquad \text{cm} \tag{18.10}$$

where R' is in centimeters, V in volts, and B' in gausses [see Eq. (6.62) for the development of this relation].

The frequency of rotation of an electron under the above conditions depends only upon the magnetic-field strength. This is because the electron velocity and radius are in direct proportion. The frequency of rotation is given by

$$\omega_0 = \frac{e}{m} B \qquad \text{radians per sec} \tag{18.11}$$

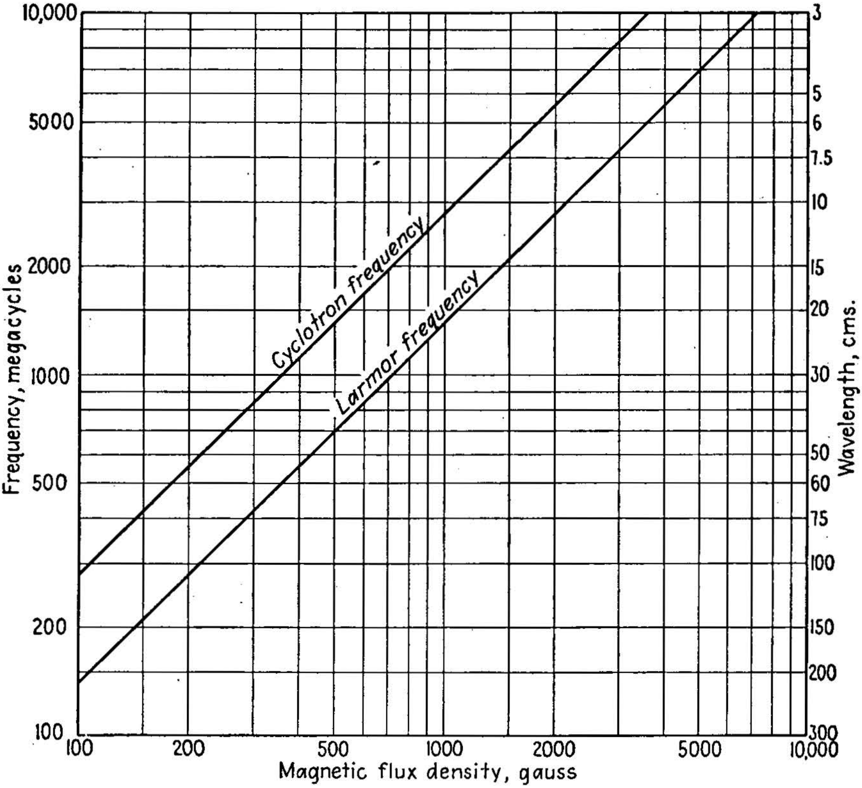


FIG. 18.11.—Dependence of the cyclotron and Larmor frequencies upon the magnetic-flux density.

where B is in webers per square meter. In practical cgs units the frequency of rotation is

$$f_0 = 2.800B' \qquad \text{mc} \tag{18.12}$$

where B' is magnetic-flux density in gausses. For obvious reasons, this frequency will hereafter be referred to as the *cyclotron frequency*. Note that electron rotations in a magnetic field are inherently of the right frequency for microwave operation. Thus a magnetic-flux density of

1,000 gausses gives rise to a cyclotron frequency of 2,800 mc or a wave length of 10.82 cm. A curve of the cyclotron frequency as a function of the magnetic-flux density is given in Fig. 18.11.

Another frequency that appears in magnetron electron orbits is the *Larmor frequency, which is just half the cyclotron frequency.* The Larmor frequency is the frequency with which atoms will precess about lines of magnetic flux.[1] The atoms may be thought of as little gyroscopes having the property of magnetic dipoles because of the fact that each rotating electron is equivalent to a small current loop. If an external magnetic field is applied, the magnetic dipole of the atom has a torque applied to it, which causes the atom to precess like a top. The Larmor precession frequency is

$$\omega_L = \frac{eB}{2m} \qquad \text{radians per sec} \tag{18.13}$$

where B is in webers per square meter and $\frac{e}{m}$ is in coulombs per kilogram. In practical cgs units this is

$$f_L = 1.400B' \qquad \text{mc}$$

where B' is in gausses. The Larmor frequency as a function of magnetic-flux density is also shown in Fig. 18.11. Something akin to atomic precession is encountered in magnetron orbits. If an electron is moving in a circular path under the influence of a radial electric and an axial magnetic field and is then disturbed, it will oscillate about the original circular path at the Larmor frequency. It is also found that electron rotations in the presence of space charge occur at the Larmor frequency.

An electron starting from rest in a region that has a uniform gradient of potential in the positive y direction and a uniform magnetic field in the negative z direction will move in a cycloidal path progressing in the positive x direction with components of displacement given by

$$x = \frac{at}{\omega_0} - \frac{a}{\omega_0{}^2} \sin \omega_0 t \tag{18.14}$$

and

$$y = \frac{a}{\omega_0{}^2} (1 - \cos \omega_0 t) \tag{18.15}$$

where $a = -\frac{eE_y}{m}$ and $\omega_0 = \frac{eB_z}{m}$. The corresponding velocity terms are given by

---

[1] See, for instance, HARNWELL, G. P., "Principles of Electricity and Electromagnetism," p. 336, McGraw-Hill, New York, 1938.

$$\dot{x} = \frac{a}{\omega_0}(1 - \cos \omega_0 t) \tag{18.16}$$

$$\dot{y} = \frac{a}{\omega_0} \sin \omega_0 t \tag{18.17}$$

where the dots over the letters indicate derivatives with respect to time. (See Sec. 6.8 for the development of these equations.) Examination of the above equations shows that the cycloidal motion is a combination of a circular motion at a frequency equal to the cyclotron frequency and a linear translational motion at a constant velocity of $\frac{a}{\omega_0}$ or $\frac{E_y}{B_z}$, the field-neutralizing ratio (see Sec. 6.8).

When there is again a y component of gradient of potential and a negative z component of magnetic field and the electrons have an initial velocity with components $\dot{x}_0$ and $\dot{y}_0$ at a point of zero potential, then the equations of motion are

$$x = \frac{at}{\omega_0} + (1 - \cos \omega_0 t)\frac{\dot{y}_0}{\omega_0} - \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0{}^2}\right) \sin \omega_0 t \tag{18.18}$$

$$y = \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0{}^2}\right)(1 - \cos \omega_0 t) + \frac{\dot{y}_0}{\omega_0} \sin \omega_0 t \tag{18.19}$$

(These were also developed in Sec. 6.8.) The corresponding velocity components are

$$\dot{x} = \frac{a}{\omega_0} + \dot{y}_0 \sin \omega_0 t - \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0}\right) \cos \omega_0 t \tag{18.20}$$

$$\dot{y} = \dot{y}_0 \cos \omega_0 t + \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0}\right) \sin \omega_0 t \tag{18.21}$$

Consider now only the periodic terms in the displacement.

$$x_1 = -\frac{\dot{y}_0}{\omega_0} \cos \omega_0 t - \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0{}^2}\right) \sin \omega_0 t \tag{18.22}$$

$$y_1 = \frac{\dot{y}_0}{\omega_0} \sin \omega_0 t - \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0{}^2}\right) \cos \omega_0 t \tag{18.23}$$

This is seen to be a circular motion with a radius given by

$$R_1{}^2 = x_1{}^2 + y_1{}^2 \tag{18.24}$$

$$R_1{}^2 = \left(\frac{\dot{y}_0}{\omega_0}\right)^2 + \left(\frac{a - \omega_0 \dot{x}_0}{\omega_0{}^2}\right)^2 \tag{18.25}$$

For zero initial velocity, R_1 reduces to $\frac{a}{\omega_0{}^2}$, which checks the cycloidal case, as may be seen from Eqs. (18.14) and (18.15). Note that the constant term in the x component of velocity is independent of the initial velocity and is equal to $\frac{a}{\omega_0}$ or $\frac{E_y}{B_z}$ as with the cycloidal case. Note also that the frequency of the rotational component of the motion is again the cycloidal frequency. Motion is again a combination of a circular motion and a translational one. The velocity of the circular motion is

$$v_1^2 = \dot{x}_1^2 + \dot{y}_1^2 \tag{18.26}$$

$$v_1^2 = \omega_0^2 R_1^2 \tag{18.27}$$

Thus the velocity is proportional to the radius and therefore corresponds to motion in a magnetic field alone. The resulting paths are those generated by a point on a projecting spoke of a rolling wheel and are known as *trochoidal paths*. The radius of the rolling wheel and its angular velocity are determined by the ratio of the fields and the magnetic field, respectively, but are independent of the initial velocity, which is to say the energy of the electrons. The square of the radius of the tracing circle is directly proportional to the energy of the system and may drop to zero if sufficient energy is extracted from the electron.

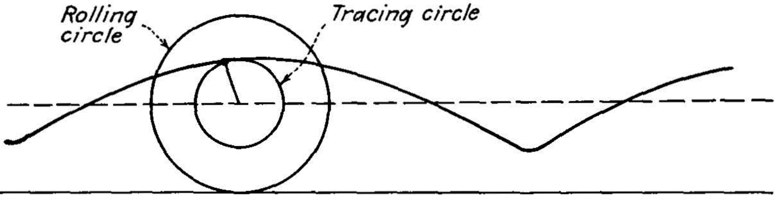


FIG. 18.12.—Modification of cycloidal path by subtraction of energy.

Because of the fact that the average translation velocity and the frequency of rotation do not change with instantaneous velocity of the electron, some observations can be made on the electron paths as energy is added or subtracted from the electron by any means. Assume an electron starting from rest at zero potential. Then the resulting path will be cycloidal. Suppose now that energy is gradually taken from the electron by some means. The path then becomes trochoidal, with the radius of the tracing circle smaller than the radius of the rolling circle but maintaining the same average translational velocity and the same cyclotron frequency of rotation. This situation is shown in Fig. 18.12. If energy were added to the electron in its original cycloidal path, the orbit would again become trochoidal, but with the radius of the tracing circle greater than the radius of the rolling circle. The resulting path is shown in Fig. 18.13. Such electrons would be removed from operation in an actual tube, for the electrons would strike the cathode as soon as any energy were added. The above conclusions on the effect of changing the electron energy will be verified quantitatively in the next section.

18.5. Electron Behavior in Crossed Magnetic and Alternating Electric Fields: Plane Case. *Alternating Transverse Electric Field.* Consider now the case of an electron starting from rest at a point of zero potential when the electric field is y-directed and consists of a constant component with a superimposed alternating component of a frequency different from the cyclotron frequency and when the magnetic field is directed in the negative z direction. To treat this problem it is best to go back to the original differential equations of motion

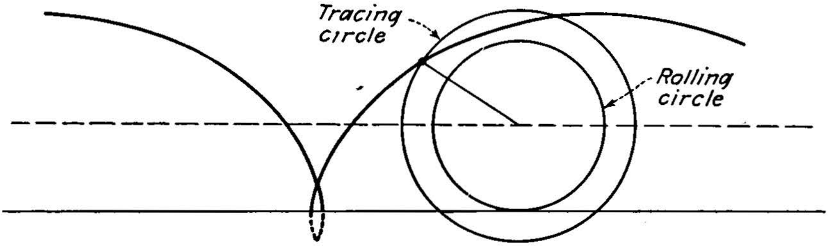


FIG. 18.13.—Modification of cycloidal path by addition of energy.

$$\ddot{x} = \omega_0 \dot{y} \tag{18.28}$$

and

$$\ddot{y} = a(1 - \alpha \cos \omega_1 t) - \omega_0 \dot{x} \tag{18.29}$$

where $E_y = -E_1(1 - \alpha \cos \omega_1 t)$, $B = -B_z$, $a = -\dfrac{eE_y}{m}$, and $\omega_0 = \dfrac{eB_z}{m}$. The starting conditions are that the initial velocity is zero, that is, $\dot{x} = 0$ and $\dot{y} = 0$ for $t = 0$. Equation (18.28) integrates to give

$$\dot{x} = \omega_0 y \tag{18.30}$$

Substituting this value into Eq. (18.29),

$$\ddot{y} + \omega_0^2 y = a(1 - \alpha \cos \omega_1 t) \tag{18.31}$$

This is analogous to the circuit problem of a series inductance and capacity with an impressed voltage consisting of a direct potential with a superimposed alternating potential of a frequency different from the resonant frequency. The solution will consist of two parts. The first part is the transient response known as the "complementary function" and is the same as that given in Eq. (18.15). The second part is known as the "particular integral" and corresponds to the steady-state solution in the equivalent electrical circuit. It is expected to be of the form

$$y_i = A \cos \omega_1 t + B \tag{18.32}$$

from inspection of Eq. (18.31). Substitution into Eq. (18.31) shows that the particular integral associated with the α term of Eq. (18.31) is of the form

$$y_i = \frac{a\alpha}{{\omega_0}^2 - {\omega_1}^2} (\cos \omega_0 t - \cos \omega_1 t) \tag{18.33}$$

The complete solution is therefore represented by the sum of Eqs. (18.15) and (18.33),

$$y = \frac{a}{{\omega_0}^2} (1 - \cos \omega_0 t) + \frac{a\alpha}{{\omega_0}^2 - {\omega_1}^2} (\cos \omega_0 t - \cos \omega_1 t) \tag{18.34}$$

Substituting this into Eq. (18.30) and integrating to obtain x,

$$x = \frac{a}{\omega_0}\left(t - \frac{\sin \omega_0 t}{\omega_0}\right) + \frac{a\alpha\omega_0}{{\omega_0}^2 - {\omega_1}^2}\left(\frac{\sin \omega_0 t}{\omega_0} - \frac{\sin \omega_1 t}{\omega_1}\right) \tag{18.35}$$

the constant of integration being zero because the initial velocity was taken as zero. The above equations reduce to the cycloidal form for $\alpha = 0$. Each of the coordinate displacements is seen to have alternating components with frequencies ω_0 and ω_1. Because of this we expect that the resultant path will display some beat phenomena at the difference frequency of $\omega_0 - \omega_1$. This occurs because the alternations at frequencies ω_0 and ω_1 are alternately in phase and out of phase. A plot of the resultant path is shown in Fig. 18.14. The amplitude is seen to be high initially, to decrease to a minimum, and then to build up again. As the amplitude decreases, average kinetic energy of the electron drops and then builds up again. Average translational velocity is $\frac{a}{\omega_0} + \frac{a\alpha\omega_0}{{\omega_0}^2 - {\omega_1}^2}$, a value that is maintained constant regardless of the amplitude of oscillation. It should be pointed out, however, that, although the translational velocity will be constant for an electron starting at any particular time, the magnitude of the translational velocity will vary for electrons starting at different points on the cycle. The value given above is the maximum translational velocity that will be encountered.

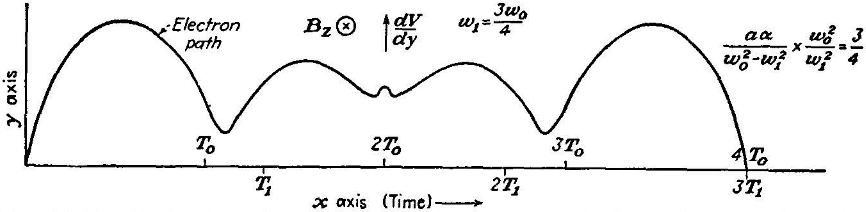


FIG. 18.14.—Path of an electron in crossed magnetic and alternating electric fields.

The minimum value will be $\frac{a}{\omega_0} - \frac{a\alpha\omega_0}{\omega_0^2 - \omega_1^2}$ and will occur for electrons that leave half a cycle later than for the case solved above. The ratio of the alternating components of displacement of frequency ω_0 and ω_1 will be

$$\frac{\text{Magnitude of } \omega_0 \text{ component}}{\text{Magnitude of } \omega_1 \text{ component}} = \frac{\omega_0^2 - \omega_1^2}{\alpha\omega_0^2} \tag{18.36}$$

When the frequency of the alternating component of electric field is the same as the cyclotron frequency, a resonance will occur that may build up the oscillations to infinite amplitude. For the off-resonance case discussed above the instantaneous radius of the rotational motion will be given approximately by

$$R_1^2 \cong \left(\frac{a}{\omega_0^2}\right)^2 + \frac{2a^2\alpha}{\omega_0^2(\omega_0^2 - \omega_1^2)} \cos{(\omega_0 - \omega_1)t} \tag{18.37}$$

on the assumption that ω_0 and ω_1 are not greatly different. This shows that the radius changes periodically at the difference-frequency rate, which means that the rotational kinetic energy changes periodically at the same rate. In an actual magnetron, use is made of electrons behaving somewhat like the one discussed above. Electrons liberated at the proper point on the cycle will have high initial kinetic energy, which they will lose at first through interaction with the alternating component of electric field. If such electrons can be removed from the field before they begin to take energy from the electric field and if electrons that initially take energy from the electric field can be removed quickly, there will be a selective mechanism by which the electrons will convert their kinetic energy derived from the static field to r-f energy, which is supplied to the alternating field. This naturally occurs in cylindrical magnetrons, for electrons that lose energy will move away from the cathode, and, with proper design, they will be taken out of action by striking the plate before they begin to absorb energy. Electrons that tend to take energy from the field will have the amplitude of their oscillations built up and will usually be removed from action by coming back and striking the cathode on the first loop of their orbit.

Effect of a Traveling Electric Field. In actual tubes the alternating components of field result from standing waves, which may be resolved into traveling waves of equal amplitude moving in opposite directions. Such waves will have both transverse and longitudinal components. In general, both the transverse component and the longitudinal component of electric field will increase in strength from cathode to plate. If an electron moves with a translational velocity corresponding closely to the velocity of the traveling waves, then the field components of the wave moving with the electron will have a considerable effect, while the field components of the wave traveling in the opposite direction will be going by the electron at twice the frequency of the alternating field and will merely introduce some perturbations, which will average out over short periods of time.

It is possible to make a reasonably exact analysis of such a case as is cited above, though by now enough properties of the electron orbits have been pointed out so that a qualitative discussion will reveal the outstanding characteristics of the resulting paths. Assume that the translational velocity of the electron is nearly equal to that of the traveling-wave components, and neglect the effect of the wave traveling in the direction opposite to that of the electron. Consider the effect of the longitudinal component of electric force, which increases toward the plate. An electron initially moving in the *same* direction as the longitudinal force will pick up energy on the portion of its loop closest to the plate and lose relatively less on the portion of its loop nearest to the cathode. There will thus be a net gain in energy, and the radius of the rotational part of the motion will increase, with the result that the electron will probably strike the cathode at the end of its first loop and be retired from action. Such electrons as tend to extract energy from the traveling wave will therefore in general be quickly removed. Those electrons which initially move against the longitudinal component of electric force will lose considerable energy on the portion of their loop closest to the plate, where the longitudinal field is strongest, and regain relatively less energy on the portion of the loop closest to the cathode. There is therefore a net loss of energy, which will cause the electron to have the radius of the rotational part of its motion decreased, indicating that the electron is giving up energy to the traveling wave. If now the associated transverse component of force is in the direction to attract the electron to the plate, the electron will drift toward the plate, where it will strike with an energy less than that corresponding to the direct potential of the plate. Electrons moving under these conditions constitute the useful, or *working, electrons* and serve to supply energy to the traveling wave. All other groups will be retired from action by striking one of the electrodes in a relatively short time, and the energy which they take from the traveling wave will be much less than that supplied by the working electrons. The mechanism by which the nonworking electrons are retired from action is highly selective and accounts for the high efficiencies obtainable with magnetron oscillators.

The diagram in the upper half of Fig. 18.15 is the approximate form of the electric field at the peak of a cycle. Both traveling-wave components will have the same shape of field, and the shape of this field will be preserved approximately as the waves move along. Shown in the figure are lines of electric *force* on electrons. The direction of the force on an electron will be opposite to the direction of the flux and field lines. In magnetron tubes designed so that the average translational velocity of the electrons is approximately equal to the velocity of the traveling waves the electrons will be subjected to a nearly constant electric force as they move along, except that the field strength increases as the electrons move from cathode to anode.

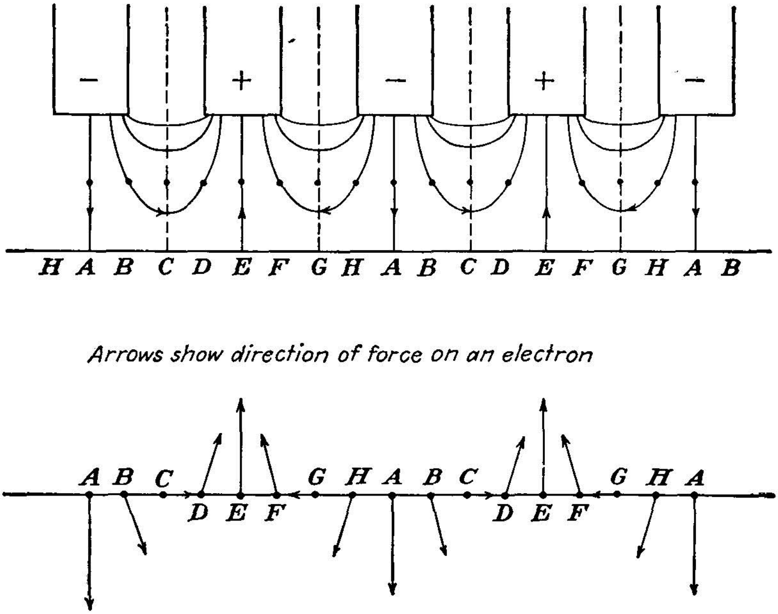


FIG. 18.15.—Force lines in a plane-electrode multianode magnetron.

The direction of the electric force at a point midway between cathode and anode is approximately as shown in the lower half of Fig. 18.15. The force is seen to rotate progressively along a line parallel to the electrodes.

Consider now the behavior of electrons emitted at different points along the cathode (or at different times on the cycle). An electron emitted in the region B will encounter a transverse force, which will tend to drive it back toward the cathode. It will also encounter a longitudinal force, which will tend to accelerate it. This means that the axis of the rotational component of its motion will be bowed, as shown by the dotted line of Fig. 18.16*a*. Also, because the electron is being accelerated in the longitudinal direction the amplitude of its rotational component of motion will increase, giving rise to a trochoidal orbit, with the result that the electron will strike the cathode after a half cycle of rotation. This is one of the nonworking electrons. It extracts a little energy from the traveling-wave component of the alternating field.

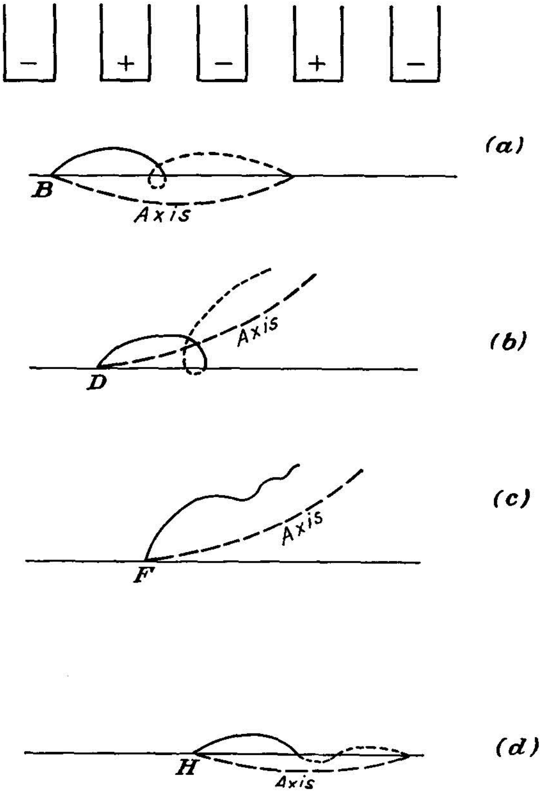


FIG. 18.16.—Approximate electron paths in a plane-electrode multianode magnetron.

An electron emitted in the region *D* will meet with accelerating components of both transverse and longitudinal forces. The axis of its rotational motion will be bowed toward the anode, as shown by the dotted line of Fig. 18.16*b*. Because this electron is accelerated, the amplitude of its rotation will increase and it will probably strike the cathode after a half cycle of its rotational motion. This is also a nonworking electron, and it extracts a little energy from the alternating field.

An electron emitted in the region F will meet with an accelerating component of transverse force and a retarding component of longitudinal force. As a result, the axis of its rotational motion will be bowed toward the anode, as shown in Fig. 18.16*c*. Because of the retarding component of longitudinal force the electron will give up energy to the alternating field and suffer a decrease in its rotational amplitude. It will strike the anode after about one cycle of its rotational motion, and during this time the rotational component of its kinetic energy will be greatly reduced. This is one of the *working electrons*, and it is electrons in this group that convert the energy of the direct component of electric field into r-f energy.

An electron emitted in the region H encounters retarding components of both transverse and longitudinal force. As a result, the axis of its rotational motion will be bowed, as shown by the dotted line in Fig. 18.16*d*. Because of the retarding component of longitudinal force the amplitude of the rotational component of its motion will decrease somewhat, though not very much, for it is forced back toward the cathode, where the longitudinal component of force is very weak. Such electrons will probably strike the cathode after the first half cycle of rotation, but some may drift along the cathode, where they will form a space-charge cloud, which will act as a source of electrons at different parts of the cycle. These electrons are low-grade working electrons in that they will contribute a little to the energy of the alternating field.

There is a bunching action associated with electrons in the F group. Those electrons in the F group emitted near the point G will meet with a larger retarding component of longitudinal force than those emitted near the center of the group. Accordingly, they will be retarded more, will move more slowly, and will fall back on those emitted near the center of the region. Those electrons which are emitted in the F group near the point E will meet with a smaller retarding longitudinal component of force and hence will not be retarded so much, will move faster, and so will catch up with those electrons emitted near the center of the group. Calculations of electron paths show that this bunching action is very strong and undoubtedly contributes to the efficiency with which energy is transferred to the alternating field.

18.6. Electron Behavior in Crossed Magnetic and Radial Electric Fields. The motions of electrons in crossed magnetic and radial electric fields are somewhat similar to those for the plane case. The similarity is close in the limiting case of very large radii but disappears as the radii become small. The equations of motion for such fields are best expressed in polar coordinates of radius and angle. The differential equations of motion may be obtained by transforming the well-known rectangular-coordinate equations to polar coordinates.[1] In polar coordinates the equations for the case of an axial magnetic field and a radial electric field are

$$\ddot{r} - r\dot{\theta}^2 = -\frac{e}{m} r\dot{\theta} B_z + \frac{e}{m}\frac{dV}{dr} \tag{18.38}$$

and

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = r\ddot{\theta} + 2\dot{r}\dot{\theta} = \frac{e}{m}\dot{r}B_z \tag{18.39}$$

where the dots over the coordinates indicate derivatives with respect to time and the other symbols have their usual significance in mks units. The equation for the radial component of motion is seen to consist of two acceleration terms and two force terms. The first acceleration term is the simple radial acceleration. The second radial-acceleration term represents the acceleration associated with circular motion. The difference in signs is due to the fact that a positive radial force is required to sustain positive radial acceleration, while a negative radial force is required to overcome the acceleration due to circular motion. The first force term in the radial equation is the radial force caused by the reaction of the angular component of velocity with the axial component of magnetic field. The second radial-force term is that due to the radial electric field. The equation for the angular component of motion involves two angular-acceleration terms and one angular-force term. The first angular acceleration results from the change of angular velocity with time. The second angular-acceleration term corresponds to the force required to maintain a constant angular velocity as the radial distance changes. The angular component of force is entirely derived from the magnetic field and results from the reaction of the radial component of velocity with the axial magnetic field.

The above equations of motion are more simply written if the cyclotron angular frequency $\omega_0 = \dfrac{eB_z}{m}$ is introduced. The equations in terms of this frequency are

$$\ddot{r} - r\dot{\theta}^2 = -\omega_0 r\dot{\theta} + \frac{e}{m}\frac{dV}{dr} \tag{18.40}$$

and

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = \omega_0\dot{r} \tag{18.41}$$

These equations are amenable to a little simplification if attention is initially restricted to cases in which both the radial and the angular

---

[1] See, for instance, MacMillan, W. D., "Statics and the Dynamics of a Particle," p. 238, McGraw-Hill, New York, 1927.

component of velocity are zero at a cathode, $r = r_c$. Equation (18.41) can be integrated to give

$$r^2\dot{\theta} = \frac{\omega_0}{2}(r^2 - r_c{}^2) \tag{18.42}$$

or, solving for $\dot{\theta}$,

$$\dot{\theta} = \frac{\omega_0}{2}\left(1 - \frac{r_c{}^2}{r^2}\right) \tag{18.43}$$

This equation is subject only to the restriction that the velocity be zero when the radius is equal to the cathode radius. It shows that the angular velocity depends only on the radius and the magnetic-field strength. The angular velocity is seen to rise from a value of zero at the cathode to a limiting value of half the cyclotron angular frequency, *i.e.*, the Larmor angular frequency, at very large radii. A curve giving the relation between the angular velocity and the cyclotron angular frequency as a function of radius is shown in Fig. 18.17. Equation (18.43) does not apply if the electron gains or loses energy after its departure from the cathode.

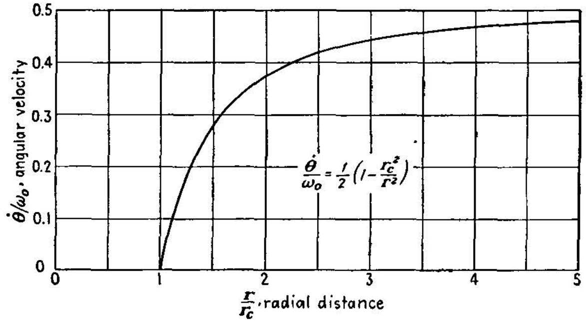


FIG. 18.17.—Angular velocity of an electron moving under the influence of an axial magnetic field and a radial electric field as a function of radius.

It is possible to get a differential equation for the radial component of motion alone by substituting the value of the angular velocity as given in Eq. (18.43) into Eq. (18.40). The resulting equation is

$$\ddot{r} + \frac{r\omega_0{}^2}{4}\left(1 - \frac{r_c{}^4}{r^4}\right) = \frac{e}{m}\frac{dV}{dr} \tag{18.44}$$

This equation is rather difficult to solve in general because it is nonhomogeneous and because V is a function of r (usually logarithmic). However, many useful deductions about orbits in limiting cases can be made from this equation.

The energy equation for polar coordinates is like that in any set of coordinates except that the velocity is expressed in terms of radial and angular components. It is

$$\frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) = eV(r) \tag{18.45}$$

on the assumption that the velocity and potential are zero at the cathode. This simply states that the kinetic energy gained is equal to the potential through which the electron has fallen. The magnitude of the velocity is seen to depend only upon the potential and to be independent of the magnetic field and the direction of the velocity. The direction of the velocity will, however, depend upon the magnetic field. With the above assumptions it is possible to specify conditions under which an electron will just graze the plate of a cylindrical magnetron. Substitute the value of the angular velocity from Eq. (18.43) into Eq. (18.45) to eliminate this factor. There results

$$\dot{r}^2 + \frac{r^2\omega_0^2}{4}\left(1 - \frac{r_c^2}{r^2}\right)^2 = \frac{2e}{m}V \tag{18.46}$$

For an electron grazing the plate, *i.e.*, for cutoff, the condition that $\dot{r} = 0$ for $r = r_p$ is imposed, where the subscript p refers to the plate. This gives

$$\frac{r_p^2\omega_0^2}{4}\left(1 - \frac{r_c^2}{r_p^2}\right)^2 = \frac{2e}{m}V_{pc} \tag{18.47}$$

or, solving for V_{pc} in terms of the other factors,

$$V_{pc} = \frac{e}{m}\frac{r_p^2}{8}B_z^2\left(1 - \frac{r_c^2}{r_p^2}\right)^2 \tag{18.48}$$

where V_{pc} is the voltage below which no electrons emitted with zero velocity will reach the plate. This equation shows that the voltage required to give cutoff in a magnetron increases as the square of the magnetic field for a given tube geometry. It is often referred to as the "cutoff parabola" and was originally derived by Hull.[1] For convenience in calculation let the magnetic field be expressed in gausses as B_z', let distance be measured in cms, and let the constant be numerically evaluated. Then

$$V_{pc} = \frac{r_p^2 B_z'^2}{45.48}\left(1 - \frac{r_c^2}{r_p^2}\right)^2 \quad \text{volts} \tag{18.49}$$

The cutoff equation given above is exact whether there is space charge present or not, for it is derived from the energy relation. The shape

---

[1] HULL, A. W., Effect of a Uniform Magnetic Field on the Motion of Electrons between Coaxial Cylinders, *Phys. Rev.*, vol. 18, pp. 31–61, July, 1921.

of the potential field between cathode and plate will be influenced by the presence of space charge, but the grazing relation will not. A nomographic chart of the cutoff relation of Eq. (18.49) is given in Fig. 18.18.

In the absence of oscillations and space charge, electrons will move out from the cathode in cardioid-like orbits, returning again to the cathode with zero velocity provided that they are not intercepted by the plate en route. When there is no energy added or subtracted en route, the orbits will always consist of single loops between the contacts with the cathode. For the limiting case of a very small cathode the orbits are represented approximately by[1]

$$r \doteq r_{\max}\left(\sin\frac{2\theta}{3}\right)^{3/2} \tag{18.50}$$

Another static orbit of interest is that in which the electron simply rotates in a circular orbit around the cathode at a constant radius. In the absence of space charge, it is a little difficult for an electron to get into such an orbit, but such an orbit is possible. The equation for this case is obtained from Eq. (18.40) by setting the radial acceleration equal to zero. The resulting equation may then be written

$$\dot{\theta}^2 - \omega_0\dot{\theta} - \frac{\omega_0 E}{rB_z} = 0 \tag{18.51}$$

where $E = -\dfrac{dV}{dr}$ is directed inward. This may be solved for the angular velocity to give

$$\dot{\theta} = \frac{\omega_0}{2} + \frac{1}{2}\sqrt{\omega_0{}^2 + \frac{4\omega_0 E}{rB}} \tag{18.52}$$

Numerical substitution shows that the second term in the radical is invariably much smaller than the first, and thus the first two terms of the binomial expansion may be used to give

$$\dot{\theta} \cong \omega_0 + \frac{E}{rB} \tag{18.53a}$$

or

$$\dot{\theta} \cong \omega_0 - \frac{1}{rB}\frac{dV}{dr} \tag{18.53b}$$

where B is in webers per square meter (10^4 gausses). This shows that the angular velocity is a little less than the cyclotron angular frequency. Numerically, the second term seldom exceeds 10 per cent of the cyclotron angular frequency. This means that the inward-directed radial magnetic force is much greater than the outward-directed radial electric force.

---

[1] *Ibid*

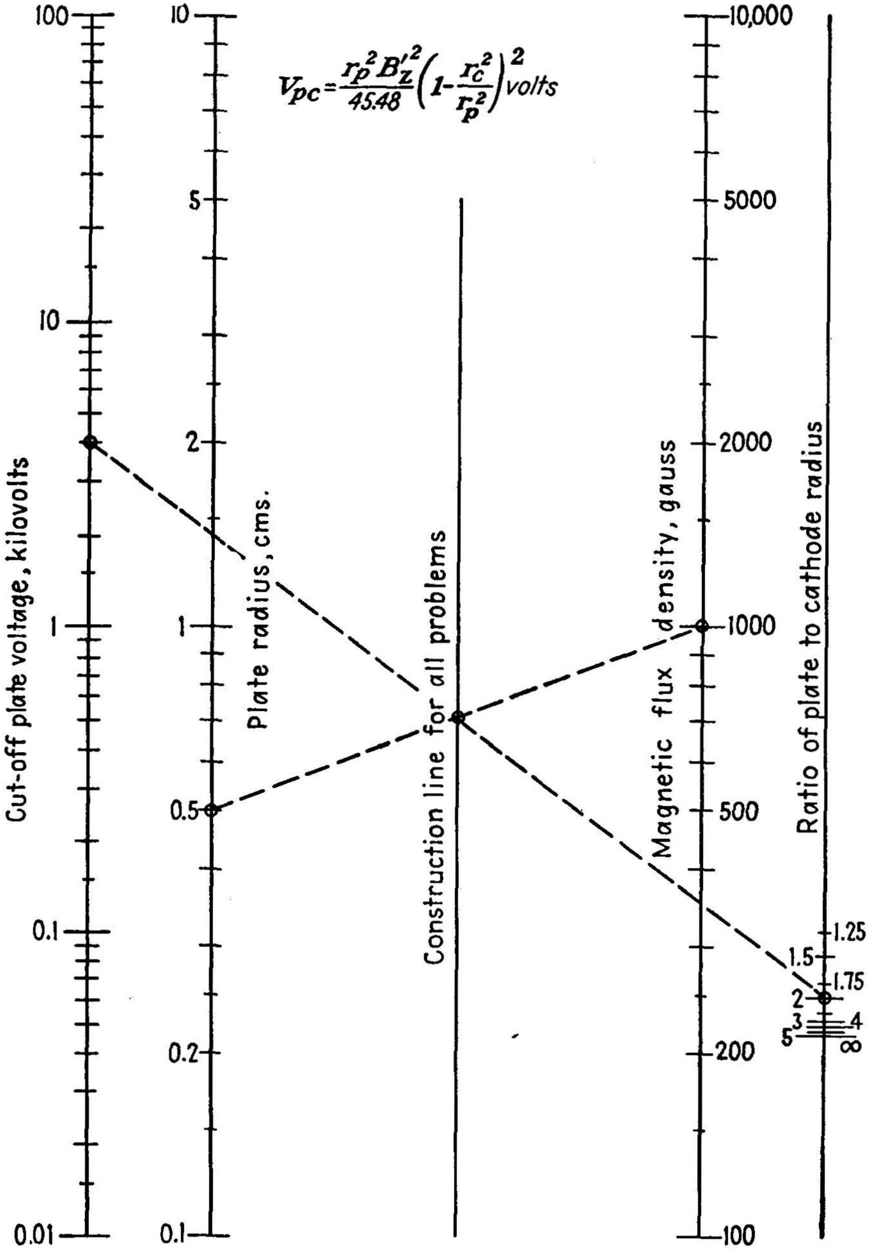


FIG. 18.18.—Nomographic chart of cutoff voltage in a cylindrical magnetron.

For ordinary purposes the angular velocity for a fixed radial distance can be taken as the cyclotron angular frequency.

18.7. The Effect of Space Charge. In an actual operating magnetron it is expected that space-charge effects cannot be neglected. Since the transit time associated with the curved paths is relatively large, the electrons will stay in the interelectrode space of a magnetron much longer than in that of a cylindrical diode without axial magnetic field. As a result, the space-charge effects should be much more pronounced and should exhibit a considerable smoothing effect upon the shape of the electron paths. The analytical treatment of space-charge effects is expected to be somewhat difficult; yet a considerable impression has been made on this subject.

The basic differential equations that have been given before are expected to apply to the space-charge case, with the difference that the potential distribution will be altered by the space charge. Specifically, the equations involving angular velocity but not the potential distribution [Eqs. (18.41) to (18.43)] will be unchanged. Likewise, the energy equation [Eq. (18.45)] and the corresponding differential equation for radial displacement [Eq. (18.46)] will apply, with the difference that the potential function is influenced by the space charge. The potential distribution will be given by Poisson's equation in polar coordinates for the single coordinate of radius,

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right) = \frac{-\rho}{\varepsilon_0} \tag{18.54}$$

where ρ is space-charge density in coulombs per cubic meter, negative for electrons, and ε_0 is the dielectric constant of free space in mks units. The radial current through any cylinder concentric with the axis of the tube is proportional to the radial velocity and the space-charge density,

$$J_r = 2\pi r\rho\dot{r} \tag{18.55}$$

With this substitution, Eq. (18.54) becomes

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right) = \frac{-J_r}{\varepsilon_0 2\pi r\dot{r}} \tag{18.56}$$

If now the value of $\dot{r}$ from Eq. (18.46) be substituted, there results

$$\frac{d}{dr}\left(r\frac{dV}{dr}\right) = \frac{-J_r}{\varepsilon_0 2\pi\sqrt{+\dfrac{2eV}{m} - \dfrac{\omega_0{}^2}{4}\left(r - \dfrac{r_c{}^2}{r}\right)^2}} \tag{18.57}$$

This is the differential equation for the potential as a function of radial distance, including the effect of space charge.

A detailed study of Eq. (18.57) shows that the radial acceleration of an electron is governed by an apparent potential which varies as the two-thirds power of the radius near the cathode and as the inverse square of the radius close to the plate as long as the current is not cut off. The apparent potential referred to is the real potential less the critical potential that would just prevent an electron from reaching a plate of radius r.[1-3] Between the cathode and plate the apparent potential is a continuously increasing function of radius. As a result, the radial velocity will *always be positive*, increasing rapidly at first and then more slowly. Since the corresponding angular velocity as given by Eq. (18.43) and Fig. 18.17 is a continuously increasing function of the radius, being small at first and then increasing with the radius, the resultant electron paths will be nearly radial at the cathode, and will then curve strongly into a spiral orbit out to the plate. In Fig. 18.19 are shown some

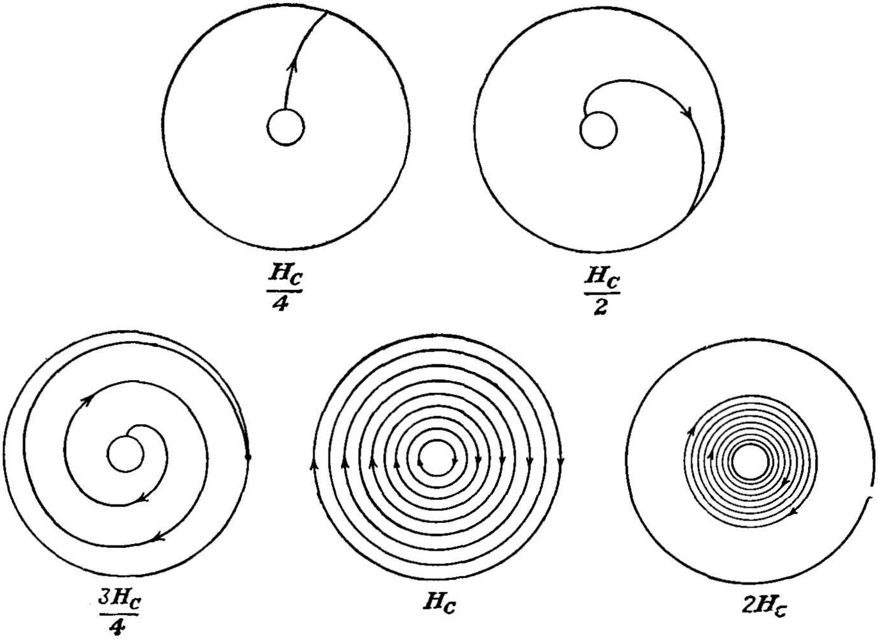


FIG. 18.19.—Spiral electron orbits in the cylindrical magnetron in the presence of space charge. (*After Brillouin.*)

---

[1] BRILLOUIN, L., Theory of the Magnetron, *Elec. Commun.*, vol. 20, pp. 112–121.

[2] BRILLOUIN, L., Theory of the Magnetron I. *Phys. Rev.*, vol. 60, pp. 385–396, Sept. 1, 1941.

[3] BRILLOUIN, L., Practical Results from Theoretical Studies of Magnetrons, *Proc. I.R.E.*, vol. 32, pp. 216–230, April, 1944.

electron orbits in the presence of space charge for a fixed plate potential as the magnetic-flux density is increased. For low magnetic fields, the paths are nearly radial, with only a slight curvature. As the magnetic field is increased, the spiral orbits evidence themselves and the total angular progression increases. At cutoff all the electrons move in circular paths, constituting a core of space charge that rotates about the cathode almost as a solid body. As the magnetic field is increased still further, the radius of the space-charge core decreases but still maintains its composition of electrons moving in circular paths with a nearly constant angular velocity.

The case of the electrons moving in circular orbits for voltages beyond cutoff is of considérable interest, for it is found that actual magnetrons operate most efficiently well beyond cutoff. The rotating core of space charge undoubtedly plays an important role in the operation. This case may be handled analytically. Setting the radial current in Eq. (18.57) equal to zero requires that the radical in the denominator of the right-hand term also be zero. Hence

$$V(r) = +\frac{m}{2e}\frac{\omega_0^2}{4}\left(r - \frac{r_c^2}{r}\right)^2 \qquad (18.58)$$

Upon differentiating this in accordance with Eq. (18.54) there is obtained an expression for the space-charge density as a function of the radial distance,

$$\rho = -\varepsilon_0\omega_0^2\left(\frac{m}{2e}\right)\left(1 + \frac{r_c^4}{r^4}\right) \qquad (18.59)$$

Each of the above expressions applies only out to the radius at which the cutoff relation of Eq. (18.49) holds, with a general radius substituted for plate radius. The potential is seen to increase nearly quadratically with radius out to the edge of the space-charge core. Beyond that it will follow the logarithmic function that applies for cylindrical electrodes in the absence of space charge. The space-charge density is seen to be nearly constant for large values of r but will rise to twice the large-radius value at the cathode. The angular velocity follows the law of Eq. (18.43) and Fig. 18.17. Accordingly, the core is one whose density is nearly constant except for an increased density near the cathode and whose outer portions rotate at half the cyclotron frequency and whose inner portions rotate at lower frequencies.

The above picture of a rotating core of space charge has been verified experimentally. In an experiment in which an indication of the current flowing is measured by the number of positive ions created by collision, the positive-ion current is found to increase sharply as the plate current is cut off, indicating that a greater current is flowing around the cathode than was flowing to the plate.[1] Further confirmation of this type of motion is obtained by considering the equivalent relative dielectric constant of an electron cloud, which in this case, by application of Maxwell's equations, is found to be

$$\varepsilon' = 1 - \frac{\omega_0^2}{2\omega^2} \tag{18.60}$$

If an experimental coaxial diode is made that can be inserted into a coaxial line, it is found that the equivalent dielectric constant of the tube section of the line follows very closely the relation given above.[2,3] The quantitative agreement with the simple theory in the above experiments, while not perfect, is very convincing, though the complete validity of the ideas involved is subject to some question.[4]

18.8. Electron Behavior in Crossed Magnetic and Alternating Radial Electric Fields. No complete analysis of electron motion in crossed magnetic and alternating radial electric fields is available although the relations seem to be reasonably well understood. Relations for small-amplitude oscillations with and without space charge can be given, though these obviously tell only part of the story since actual magnetron oscillations involve large amplitudes. Large-amplitude relations can be calculated numerically for specific tube dimensions and operating conditions, from which some general deductions can be made. It is worth considering the small-amplitude relations, however, in that they will contain some elements of truthful representation of the actual picture.

Consider first the small-amplitude oscillations without space charge, based upon Eq. (18.44). Let the gradient of potential at any radius r_0 be given by

$$\frac{e}{m}\frac{dV}{dr} = a_0 + a_1(r - r_0) \tag{18.61}$$

i.e., a constant plus a linear term, and let

$$\zeta = r - r_0 . \tag{18.62}$$

---

[1] HULL, A. W., The Paths of Electrons in the Magnetron (Abstract Only), *Phys. Rev.*, vol. 23, p. 112, January, 1924.

[2] BLEWETT, J. P., and S. RAMO, High Frequency Behavior of a Space Charge Rotating in a Magnetic Field, *Phys. Rev.*, vol. 57, pp. 635–641, April 1, 1940.

[3] BLEWETT, J. P., and S. RAMO, Propagation of Electromagnetic Waves in a Space Charge Rotating in a Magnetic Field, *Jour. Appl. Phys.*, vol. 12, pp. 856–859, December, 1941.

[4] GABOR, D., Stationary Electron Swarms in Electromagnetic Fields, *Proc. Roy. Soc.*, (*London*), Ser. A, vol. 183, pp. 436–453, 1945.

Upon making these substitutions in Eq. (18.44) and preserving only first-power terms in ζ there results the differential equation for the perturbed motion about any radius r_0,

$$\ddot{\zeta} + \frac{\zeta\omega_0^2}{4}\left[1 - 3\left(\frac{r_c}{r_0}\right)^4 + \frac{4a_1}{\omega_0^2}\right] = a_0 - r_0\left[1 - \left(\frac{r_c}{r_0}\right)^4\right]\frac{\omega_0^2}{4} \qquad (18.63)$$

Of principal interest is the periodic term in the solution of this equation. This will have the form

$$\zeta_1 = A \cos \frac{\omega_0}{2}\sqrt{1 - 3\left(\frac{r_c}{r_0}\right)^4 + \frac{4a_1}{\omega_0^2}}\,t \qquad (18.64)$$

The resulting path is like that generated by a point on a small circle rolling on a large circle of radius r_0 concentric with the cathode. The angular frequency of the perturbed motion is seen to differ from $\frac{\omega_0}{2}$ by a radical containing a distance ratio raised to the fourth power (generally small) and the coefficient of the gradient, a_1, which will be positive in the presence of space charge and negative in its absence. This means that the perturbation frequency will ordinarily be less than the Larmor frequency (half the cyclotron frequency) in the absence of space charge and greater than the Larmor frequency in its presence. Correspondingly, the *average* angular velocity will be

$$\dot{\theta} = \frac{\omega_0}{2}\left(1 - \frac{r_c^2}{r_0^2}\right) \qquad (18.65)$$

which is the same as previously given by Eq. (18.43). At large radii and in the presence of space charge the perturbation frequency can be many times the average angular velocity.

The conclusion that the perturbation frequency is more than the Larmor frequency (half the cyclotron frequency) is confirmed by examination of a simple oscillation mode in the presence of space charge. Let it be assumed that there is under consideration a rotating core of space charge. At the outer edge of the core, where $\left(\frac{r_c}{r}\right)^4$ is much less than unity, the differential equation of the radial component of motion will be

$$\frac{\ddot{r}}{r} = \frac{e}{m}\frac{1}{r}\frac{dV}{dr} - \frac{\omega_0^2}{4} \qquad (18.66)$$

from Eq. (18.44). Now let $r = r_0 + \zeta$ as before, and apply this to a uniform expansion through the core. Conservation of charge then requires that

$$\rho = \rho_0\left(1 - \frac{2\zeta}{r_0}\right) \tag{18.67}$$

Now since

$$\frac{1}{r}\frac{dV}{dr} = \frac{-\rho}{2\varepsilon_0} \tag{18.68}$$

from Eq. (18.54) and the large-radius value of ρ_0 is

$$\rho_0 = -\varepsilon_0\omega_0{}^2\left(\frac{m}{2e}\right) \tag{18.69}$$

from Eq. (18.59), then the differential equation [Eq. (18.66)] takes the form

$$\frac{\ddot{\zeta}}{r_0\left(1 + \dfrac{\zeta}{r_0}\right)} = \frac{\omega_0{}^2}{4}\left(1 - \frac{2\zeta}{r_0}\right) - \frac{\omega_0{}^2}{4} \tag{18.70}$$

which reduces to simply

$$\ddot{\zeta} = -\frac{\omega_0{}^2}{2}\zeta \tag{18.71}$$

if terms in $\dfrac{\zeta}{r_0}$ of powers higher than unity are disregarded. From this it is seen that the perturbation frequency of the electrons in the space-charge cloud is $\dfrac{\omega_0}{\sqrt{2}}$ for the simple mode of oscillation in which the whole cloud pulsates uniformly.[1,2]

Solutions other than the simple one indicated above can be obtained for the magnetron with space charge. These will not be discussed in detail, for their application is limited to small-amplitude oscillations. In addition to the pulsating core of space charge just referred to, solutions have been found in which the edge of the space-charge core has sinusoidal ripples appear on it in the form of standing waves, with an integral number of sine waves around a complete circumference. These standing waves can be resolved into traveling waves of equal amplitude traversing the circumference of the core with equal velocities in the two directions.[1,2]

---

[1] BRILLOUIN, L., Theory of the Magnetron II, *Phys. Rev.*, vol. 62, pp. 166–177, Aug. 1 and 15, 1942.

[2] BRILLOUIN, L., Theory of the Magnetron III, *Phys. Rev.*, vol. 63, pp. 127–136, Feb. 1 and 15, 1943.

Some of these modes exhibit an associated negative resistance and so may give rise to oscillations. Likewise, there have been found solutions in which there is a rotating cylinder of space charge with definite inner and outer edges not in contact with the electrodes. In such cylinders of charge it is possible to have clumps or spokes of increased space-charge density, which rotate at half the cyclotron angular frequency.[1] The appearance of spokes of space-charge density in an analytical solution is of great significance, for it confirms the existence of such spokes predicted from simple qualitative considerations.

So far the information obtained about electron behavior in the presence of alternating components of electric field has not been very enlightening with regard to efficiency of operation and other practical matters. It is perhaps too much to expect that an analysis of this complex problem will yield neat and simple engineering-design formulas. The best that can be done at present is to attempt to get a composite picture of the mechanism of operation by combining the impressions obtained by looking through the various windows corresponding to the different avenues of approach to the problem.

Considerable information is obtained from considering the reaction of electrons with rotating-field components. If an electron moves so that it is being continuously retarded by a tangential component of electric force, it will give up energy, which will allow it to move in a larger radius path. Since energy is being given up, it is possible for such an electron to have its angular velocity become progressively less than the value it would have at any radius if it had not lost energy. Accordingly, it is possible for electrons to spiral out to the plate with a constant or nearly constant angular frequency of rotation.[2]

The alternating components of the electric field of a cylindrical multicavity magnetron contain both radial and tangential components, which can be resolved into components traveling in the two directions. Let the radial component of the alternating gradient of potential in the direction of the electron travel be $R(r)\phi(n\theta + \omega t)$ and the tangential component be $T(r)\psi(n\theta + \omega t)$. The components rotating in the opposite direction will be neglected. The functions ψ and ϕ are periodic functions of the angle θ, with n an integer equal to the number of full-period variations of field around the magnetron. Near the cathode, ψ and ϕ will be simple cosine waves, but near the plate they will be nearly square waves. If these components of the gradient of potential are

---

[1] Blewett, J. P., and S. Ramo, High Frequency Behavior of a Space Charge Rotating in a Magnetic Field, *Phys. Rev.*, vol. 57, pp. 635–641, 1940.

[2] Application of the above ideas was first made by Posthumus, K., Oscillations in a Split Anode Magnetron, *Wireless Eng.*, vol. 12, pp. 126–132, March, 1935.

included in the basic differential equations of motion, then Eqs. (18.40) and (18.41) become

$$\ddot{r} - r\dot{\theta}^2 = \frac{e}{m}\left[\frac{dV}{dr} + R(r)\phi(n\theta + \omega t)\right] - \omega_0 r\dot{\theta} \tag{18.72}$$

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = \frac{e}{m}T(r)\psi(n\theta + \omega t) + \omega_0\dot{r} \tag{18.73}$$

Let it now be considered whether it is possible for an electron tc follow the field around in such a way that

$$n\theta = -\omega t + \alpha \tag{18.74}$$

where the negative sign goes with the counterclockwise rotation of an electron which occurs for a magnetic field in the negative z direction and it is assumed that α changes very little with time. The interpretation of the angle α is that it is the angle by which the electron lags some reference point on the rotating field, conveniently the maximum. Then since

$$\dot{\theta} \cong -\frac{\omega}{n} \tag{18.75}$$

the equations of motion above become

$$\ddot{r} - \frac{r\omega^2}{n^2} = \frac{e}{m}\left[\frac{dV}{dr} + R(r)\phi(\alpha)\right] + \omega\frac{\omega_0 r}{n} \tag{18.76}$$

$$-\frac{2\dot{r}\omega}{n} = \omega_0\dot{r} + \frac{e}{m}T(r)\psi(\alpha) \tag{18.77}$$

These equations can be partly solved without knowing the exact nature of the functions ψ and ϕ. Let Eq. (18.76) be integrated on the assumption that the radial velocity and potential at the cathode are zero. Then

$$\dot{r}^2 = r^2\left(1 - \frac{r_c^2}{r^2}\right)\frac{\omega}{n}\left(\frac{\omega}{n} + \omega_0\right) + \frac{2e}{m}V(r) + \frac{2e}{m}\int R(r)\phi(\alpha)\,dr \tag{18.78}$$

If now the value of $\dot{r}^2$ from Eq. (18.77) is substituted in the above, there results

$$\left[\frac{-\frac{e}{m}T(r)\psi(\alpha)}{\frac{2\omega}{n} + \omega_0}\right]^2 = r^2\left(1 - \frac{r_c^2}{r^2}\right)\frac{\omega}{n}\left(\frac{\omega}{n} + \omega_0\right) + \frac{2e}{m}V(r) + \frac{2e}{m}\int R(r)\phi(\alpha)\,dr \tag{18.79}$$

This equation tells how the angle α by which the electron lags some reference line on the rotating field varies with the radial distance. Simple physical reasoning indicates that electrons will be in equilibrium when they are slightly behind a radial line of maximum retarding force. Under this condition a momentary increase of angular velocity increases the radius of the orbit and brings the electron into a region of stronger retarding force that acts to decrease the angular velocity. The argument here is the same as that used in consideration of Figs. 18.15 and 18.16. The radial force is not necessary to the argument and will for the moment be considered negligible. A possible situation demonstrated by Eq. (18.79) is shown in Fig. 18.20. Shown here are tangential components of electron force rotating in the counterclockwise direction for a six-segment magnetron operating on its π mode. Nodal planes of force are shown by dashes. The position of electrons in equilibrium with the field is shown by the dotted curve lagging a plane of maximum retarding force. If the square of the radial function $T(r)$ increases less rapidly than the radial function of the right-hand side of Eq. (18.79), then the angle α by which the electron lags the line of maximum retarding force must increase as the radius increases. Note that, although the effects of space charge have not been specifically considered, this treatment admits of solution in cases with space charge, for then it is merely necessary to introduce the proper form of the potential, $V(r)$. Including the effect of the radial forces will only change the locus of the electrons in equilibrium with the field. The locus will always lie within the zone between a plane of maximum tangential force and the nodal plane of tangential force behind it.

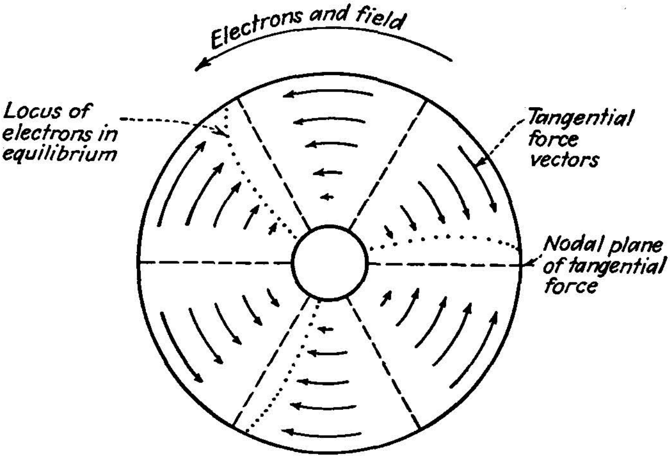


FIG. 18.20.—Electron orbits in a multicavity magnetron as determined by rotating tangential components of field.

For the case of negligible radial force the square of the radial velocity of the equilibrium electrons at the plate will be

$$\dot{r}_p^2 = r_p^2\left(1 - \frac{r_c^2}{r_p^2}\right)\left(\frac{\omega^2}{n^2} + \frac{\omega\omega_0}{n}\right) + \frac{2e}{m}V_p \tag{18.80}$$

Correspondingly, the square of the total velocity at the plate will be obtained by adding $r_p^2\dot{\theta}^2$, as obtained from Eq. (18.75), to the value of $\dot{r}_p^2$ above.

$$v_p^2 = r_p^2\left[\left(2 - \frac{r_c^2}{r_p^2}\right)\frac{\omega^2}{n^2} + \left(1 - \frac{r_c^2}{r_p^2}\right)\frac{\omega\omega_0}{n}\right] + \frac{2e}{m}V_p \tag{18.81}$$

The above two equations are those whose properties it is desired to study. For purposes of simplification let the angular velocity at the plate be written as

$$r_p\dot{\theta} = \frac{r_p\omega}{n} = xv_0 \tag{18.82}$$

where v_0 is the velocity corresponding to the plate potential and introduce the factor

$$z = \frac{B}{B_c} = \frac{\omega_0}{\omega_c} \tag{18.83}$$

where B_c is the cutoff value of magnetic-flux density corresponding to the plate potential V_p, as obtained from Eq. (18.48). Let the cutoff relation be written

$$r_p^2\omega_c^2\left(1 - \frac{r_c^2}{r_p^2}\right) = 4v_0^2 \tag{18.84}$$

With the above substitutions, Eq. (18.80) becomes

$$\frac{\dot{r}_p^2}{v_0^2} = \left(1 - \frac{r_c^2}{r_p^2}\right)\left(x^2 + \frac{2xz}{\sqrt{1 - \frac{r_c^2}{r_p^2}}}\right) + 1 \tag{18.85}$$

and Eq. (18.81) becomes

$$\frac{v_p^2}{v_0^2} = \left(2 - \frac{r_c^2}{r_p^2}\right)x^2 + 2\sqrt{1 - \frac{r_c^2}{r_p^2}}\,xz + 1 \tag{18.86}$$

Posthumus has examined these equations for the case of $\frac{r_c}{r_p} = 0$, for which the above equations simplify to

$$\frac{\dot{r}_p^2}{v_0^2} = x^2 + 2xz + 1 \tag{18.87}$$

and

$$\frac{v_p^2}{v_0^2} = 2x^2 + 2xz + 1 \tag{18.88}$$

The relations between these factors are shown in the curves of Fig. 18.21. These curves show the square of the radial velocity and the square of the total velocity as a function of the angular velocity xv_0 which is proportional to the square root of the plate potential and inversely proportional to the mode number n. Since the energy taken from

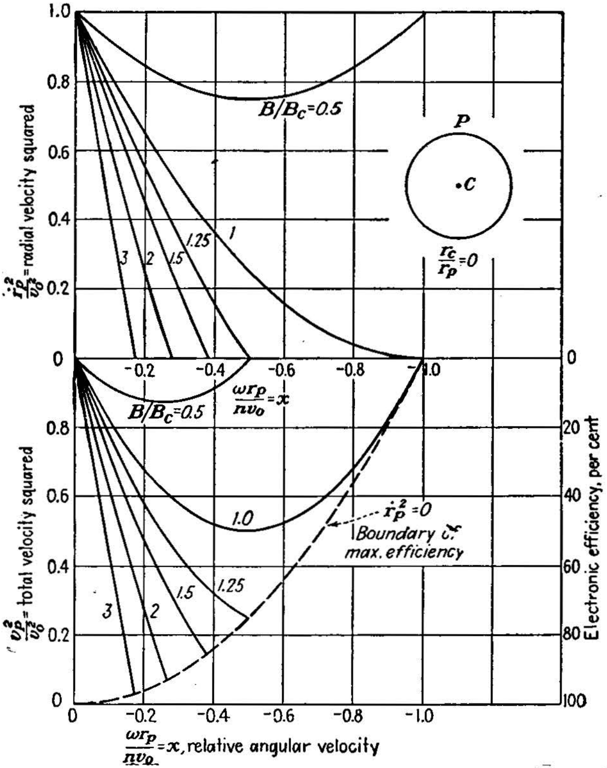


FIG. 18.21.—Total and radial electron velocity at the anode of a magnetron as a function of magnetic-flux density and angular velocity for a small ratio of cathode to plate radius. (*After Posthumus.*)

the potential source per electron is $\frac{1}{2}mv_0^2$ on the average, then the electron efficiency is

$$\text{Electron efficiency} = 1 - \frac{v_p^2}{v_0^2} \tag{18.89}$$

which means that an efficiency scale can be included on the curves of total velocity squared with a zero value of $\frac{v_p^2}{v_0^2}$ corresponding to 100 per cent efficiency. This efficiency does not, of course, include the effect of circuit losses. The curves indicate only the *maximum* efficiencies that can be obtained. Actual efficiencies will be less, since not all electrons are as favorably operated as those discussed in this analysis. The principal things to be learned from Fig. 18.21 are that higher efficiencies can be obtained at progressively higher mode numbers with higher d-c potentials and with magnetic-flux densities higher than the critical value. The dashed curve in Fig. 18.21 is for the case of $\dot{r}_p^2 = 0$. This corresponds to the case of electrons that have given up all their radial energy to the field and strike the plate at grazing incidence. This curve represents the highest efficiencies obtainable for any value of angular velocity and magnetic-flux density.

Equations (18.85) and (18.86) are a generalization of Eqs. (18.87) and (18.88) originally given by Posthumus and make possible an extension of this analysis to magnetrons with finite ratios of cathode to plate radius. Let a limiting small value of the ratio of cathode to plate radius be 0.707. Then Eqs. (18.85) and (18.86) become

$$\frac{\dot{r}_p^2}{v_0^2} = \frac{x^2}{2} + \sqrt{2}\,xz + 1 \tag{18.90}$$

$$\frac{v_p^2}{v_0^2} = 3\,\frac{x^2}{2} + \sqrt{2}\,xz + 1 \tag{18.91}$$

The corresponding curves for this high ratio of cathode to plate radius are shown in Fig. 18.22. These have the same general form as those of Fig. 18.21 except for some rather pronounced displacements. The limiting curve (shown dashed) for which the radial velocity of an electron at the plate is zero is the same for any ratio of cathode to plate radius and is simply

$$\frac{v_p^2}{v_0^2} = x^2 \tag{18.92}$$

However, for the larger ratio of cathode to plate radius of Fig. 18.22, oscillations can be had for a given mode and magnetic-flux density at a lower value of plate voltage but with a slightly lower efficiency.

In spite of some rather general assumptions made in this analysis, the results have considerable validity. Without question, the curves demonstrate correctly that it is possible to get higher efficiencies by going to higher mode numbers and magnetic-flux densities in excess of the critical value. Deductions as to the effect of the ratio of cathode and plate radius cannot be taken too seriously. The curves of Figs. 18.21 and 18.22 should be displaced upward by the amount of the integral of the radial force, which was neglected in Eq. (18.80). When this displacement upward is made, the minima of the efficiency curves play a more prominent role within the region of operation and specific deductions with regard to the effect of the dimensions definitely need to consider the effect of the radial electric forces.

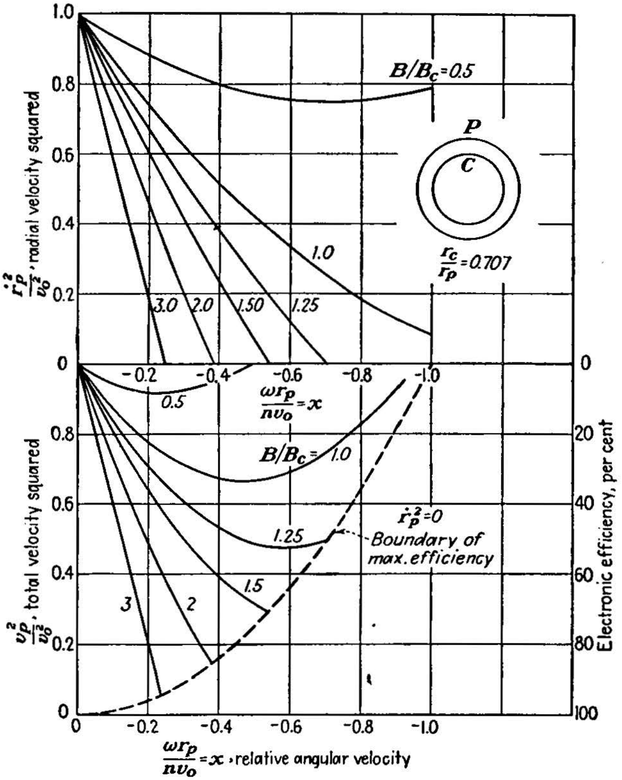


FIG. 18.22.—Total and radial electron velocity at the anode of a magnetron as a function of magnetic-flux density and angular velocity for a large ratio of cathode to plate radius.

18.9. Basic Relations for Multicavity Magnetrons. Of the various relations given for magnetrons thus far, the most important is the cutoff relation of Eq. (18.48). Further relations which have a bearing upon the a-c operation are semiempirical. Slater and his colleagues have shown by extensive calculations and tests that maximum bunching action and resultant efficiency occur in a tube when the area between cathode and plate and between two corresponding pole points, as shown by the shaded area in Fig. 18.1*c*, is approximately a curvilinear square. More specifically, the ratio of the radial to the average angular dimension of the four-sided figure shown should be $\frac{4}{\pi}$. In terms of the radii and the number of poles this gives

$$r_p - r_c = \frac{4}{\pi} 2\pi \left(\frac{r_p + r_c}{2N} \right) \tag{18.93a}$$

where N is the number of pole tips. This reduces to

$$\frac{r_c}{r_p} = \frac{1 - \frac{4}{N}}{1 + \frac{4}{N}} \tag{18.93b}$$

A curve of $\frac{r_c}{r_p}$ as a function of N is given in Fig. 18.23. For 4 plate segments or fewer, the ratio of plate to cathode radius should be zero. For a larger number of plate segments the optimum ratio increases but does so rather slowly and even at 12 segments is only at half its asymptotic value of unity. Under the condition of Eq. (18.93b) the dimensions of a multicavity magnetron will be such that the working electrons will traverse about two loops of a modified trochoidal path before being taken out of action at the plate with a small residual energy.

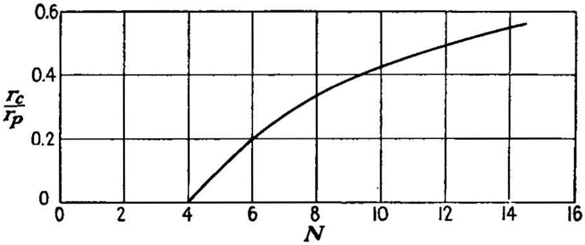


FIG. 18.23.—Optimum ratio of cathode to plate radius as a function of the number of plate segments. (*After Slater.*)

Another condition which ensures favorable action is that the electrons in their motion around the cathode move at the velocity of the traveling wave. The angular velocity of the electrons varies considerably from cathode to plate, and therefore let the angular velocity halfway between the cathode and plate be set equal to the velocity of the traveling wave. Referring to Eq. (18.40), setting the radial acceleration equal to zero, and letting the angular velocity $r\dot{\theta}$ be represented by v_θ,

$$\frac{dV}{dr} = v_\theta \left(B_z - \frac{mv_\theta}{er_0} \right) \tag{18.94}$$

where r_0 is to have the value corresponding to the halfway points between cathode and plate,

$$r_0 = \frac{r_c + r_p}{2} \tag{18.95}$$

If now, v_θ is to be equal to the wave velocity, then

$$v_\theta = r_0\dot{\theta} \tag{18.96a}$$

$$v_\theta = \frac{r_0\omega}{n} \tag{18.96b}$$

$$v_\theta = \frac{2\pi r_0 c}{n\lambda_0} \tag{18.96c}$$

$$v_\theta = \frac{\pi(r_c + r_p)c}{n\lambda_0} \tag{18.96d}$$

Assume now as a simplifying approximation that

$$\frac{dV}{dr} = \frac{V_p}{r_p - r_c} \tag{18.97}$$

for $r = r_0$. Then, with the substitutions of Eqs. (18.95), (18.96*d*), and (18.97), Eq. (18.94) becomes

$$V_p = \frac{\pi(r_p{}^2 - r_c{}^2)}{n\lambda_0} 3 \times 10^8 \left(B - \frac{0.010463}{n\lambda_0}\right) \tag{18.98a}$$

in mks units. For those who are more familiar with practical cgs units this will appear as the general relation

$$V_p = \frac{300\pi(r_p{}^2 - r_c{}^2)}{n\lambda_0} \left(B_z' - \frac{10{,}463}{n\lambda_0}\right) \tag{18.98b}$$

where V_p is in volts, B_z' is in gausses, and r and λ are in centimeters. If now the optimum ratio of cathode to plate radius of Eq. (18.93*b*) is introduced, the above equation becomes

$$V_p = \frac{300 r_p{}^2 \dfrac{16}{N}}{\left(1 + \dfrac{4}{N}\right)^2} \frac{\pi}{n\lambda_0} \left(B_z' - \frac{10{,}463}{n\lambda_0}\right) \tag{18.99}$$

for the optimum dimensions. This is the important relation that has been sought. It shows that there will be a linear relation between V_p and B_z for optimum operation on any one mode. A plot of Eq. (18.99) is usually referred to as the *mode line* or the Hartree line. The mode lines have slopes that vary inversely as the value of n, λ_0, and N. For any one tube there is a family of mode lines in the V-B plane that almost pass through the origin. Such a family of lines is shown relative to the cutoff parabola in Fig. 18.24.

The mode-line equation [Eq. (18.99)] may be solved for plate radius to give

$$r_p = \frac{\left(1 + \frac{4}{N}\right)\sqrt{nN}}{\sqrt{4{,}800\pi}} \sqrt{\frac{\lambda_0 V_p}{B_z' - \frac{10{,}463}{n\lambda_0}}} \tag{18.100}$$

where B_z' is in gausses. Tubes will ordinarily operate on the highest,

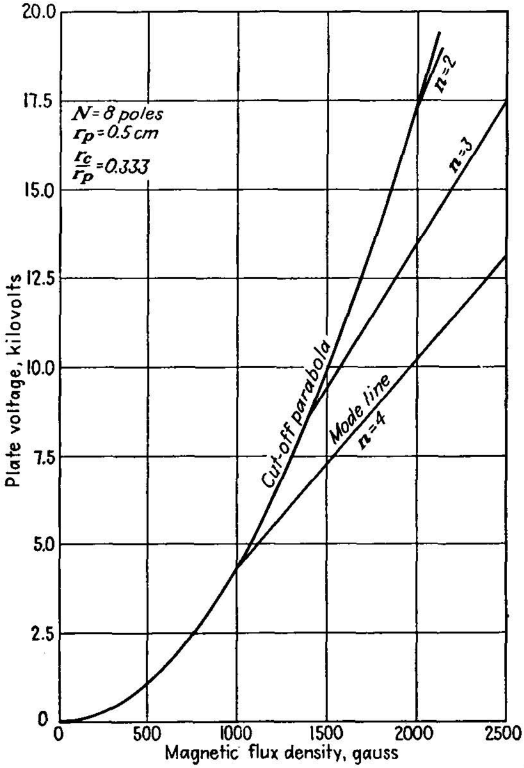


FIG. 18.24.—Mode lines in the voltage–magnetic-flux-density plane.

or π, mode for which $n = \frac{N}{2}$, in which case the above reduces to

$$r_p = \frac{N + 4}{\sqrt{9{,}600\pi}} \sqrt{\frac{\lambda_0 V_p}{B_z' - \frac{20{,}926}{N\lambda_0}}} \tag{18.101}$$

where B_z' is in gausses. The above may serve as an approximate design equation in determining plate radius of a magnetron. It assumes the optimum ratio of cathode to plate radius of Eq. (18.93*b*).

The cutoff relation itself may be rewritten to include the optimum electrode ratio of Eq. (18.93*b*). When restricted to the optimum ratio, the cutoff parabola becomes

$$V_p = \frac{B_z'^2 r_p^2}{45.48} \frac{\left(\frac{16}{N}\right)^2}{\left(1 + \frac{4}{N}\right)^4} \tag{18.102}$$

where B_z' is in gausses, and the relation is independent of the value of n since it is a static relation.

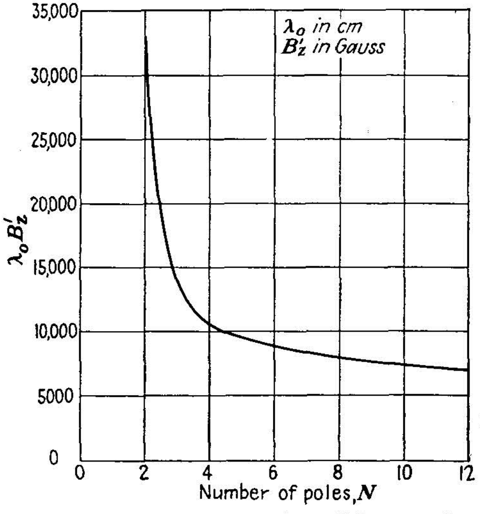


FIG. 18.25.—Values of $\lambda_0 B_z^1$ for different values of N.

Upon combining Eqs. (18.101) and (18.102), the quantities r_p and V_p may be eliminated and an expression obtained that gives $\lambda_0 B_z'$ in terms of the number of poles, N, for the case that $n = \frac{N}{2}$. This relation may be solved for the product $\lambda_0 B_z'$, which applies at cutoff for tubes with different numbers of poles N and operating in the π mode, $n = \frac{N}{2}$. The results are given by the curve of Fig. 18.25.[1] The advantage of large

---

[1] This and the other relations of this section follow the early work of J. C. Slater and colleagues. Details of the analysis, along with refinements on this elementary point of view, are given in the report of the wartime researches of the Radiation Laboratory, volume on magnetrons, McGraw-Hill, New York, 1948.

values of N in terms of low magnetic field is apparent. Actual operation will be best at flux densities considerably above cutoff.

Another condition for operation which may be specified is that the frequency of oscillation should be approximately equal to the cyclotron frequency. Examination of this condition shows that the optimum value of magnetic-flux density is approximately 33 per cent greater than the cutoff value for values of N greater than 4. Oscillations may occur almost anywhere in the V-B plane of Fig. 18.24, but greatest output will be obtained in the vicinity of the mode lines to the right of the cutoff parabola.

18.10. Dimensional Relations in Magnetrons. Many important deductions about the effect of the various parameters involved in magnetron operation can be made by examining the dimensionality of the basic differential equations involved, just as was done for the ultra-high-frequency triode.[1] The differential equations of motion of an electron under the influence of electric and magnetic fields in rectangular coordinates are

$$m\frac{dv_x}{dt} = ev_yB_z - eE_x \tag{18.103}$$

and

$$m\frac{dv_y}{dt} = -ev_xB_z - eE_y \tag{18.104}$$

for a magnetic field having only a z component and an electric field having no z component. Poisson's equation, which governs the space-charge relations, is

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = \frac{\rho}{\varepsilon_0} \tag{18.105}$$

The relations between current density, space-charge density, and velocity are

$$J_x = \rho v_x \tag{18.106}$$

$$J_y = \rho v_y \tag{18.107}$$

Let now the comparative operation of two tubes that are geometrically similar be considered. Let D be the dimension ratio and W be the wavelength ratio of the two tubes. Then, if an electron is moving between two corresponding points in the two tubes,

$$dt_2 = W\,dt_1 \tag{18.108}$$

and

$$dx_2 = D\,dx_1 \tag{18.109}$$

$$dy_2 = D\,dy_1 \tag{18.110}$$

---

[1] The analysis of this section follows the early work of A. M. Clogston, done at the Radiation Laboratory.

where

$$W = \frac{\lambda_2}{\lambda_1} \tag{18.111}$$

and

$$D = \frac{r_2}{r_1} \tag{18.112}$$

The equation of motion for tube 2 then becomes

$$m \frac{dv_{x2}}{dt_2} = ev_{y2}B_2 - eE_{x2} \tag{18.113}$$

or

$$m \frac{D}{W^2} \frac{d\acute{v}_{x1}}{dt_1} = e \frac{D}{W} v_{y1}B_2 - eE_{x2} \tag{18.114}$$

For these last two equations to be consistent it is necessary that

$$B_2 = \frac{1}{W} B_1 \tag{18.115}$$

and that

$$E_2 = \frac{D}{W^2} E_1 \tag{18.116}$$

Since potential is the product of gradient by distance and the distance ratio is D, then

$$V_2 = \frac{D^2}{W^2} V_1 \tag{18.117}$$

By an extension of this type of reasoning, ratios of all the critical quantities in the two tubes as a function of the factors D and W may be obtained. These are summarized in the table on page 667.

The quantities in Table XII enable the tube designer to tell how the various operating quantities in a tube that has been scaled from a given tube will compare with the corresponding quantities of the given tube. It further tells how the quantities in a single tube will change if the operating characteristics are changed. Thus, if a tube is enlarged by a factor D but is to work at the same wave length, then the factors in the Voltage scaling column apply. In this case the required magnetic-flux density is unchanged, the required voltage is increased by a factor of D^2, and so on. If a given tube is to be operated at a wave length greater by a factor of W than that for which the operating characteristics are known, then the required magnetic-flux density is $\frac{1}{W}$ times as great as before, the required voltage is $\frac{1}{W^2}$ times as great as before, and so on.

If with a known set of dimensions and operating characteristics the dimensions and wave length are changed in direct proportion, then the values in the Complete scaling column apply. The values in the General column take care of the general case.

TABLE XII
MAGNETRON SCALING FACTORS

Quantity	Ratio	General	Complete scaling $W = D$	Voltage scaling $W = 1$	Wave-length scaling $D = 1$
Magnetic-flux density	$\frac{B_2}{B_1}$	$\frac{1}{W}$	$\frac{1}{W}$	1	$\frac{1}{W}$
Voltage	$\frac{V_2}{V_1}$	$\frac{D^2}{W^2}$	1	D^2	$\frac{1}{W^2}$
Current density	$\frac{J_2}{J_1}$	$\frac{D}{W^3}$	$\frac{1}{W^2}$	D	$\frac{1}{W^2}$
Current	$\frac{I_2}{I_1}$	$\frac{D^2}{W^3}$	$\frac{1}{W}$	D^2	$\frac{1}{W^3}$
Power	$\frac{P_2}{P_1}$	$\frac{D^4}{W^5}$	$\frac{1}{W}$	D^4	$\frac{1}{W^5}$
Conductance	$\frac{G_2}{G_1}$	$\frac{1}{W}$	$\frac{1}{W}$	1	$\frac{1}{W}$
Gradient	$\frac{E_2}{E_1}$	$\frac{D}{W^2}$	$\frac{1}{W}$	D	$\frac{1}{W^2}$

18.11. Output Characteristics of Magnetrons. It is not possible to write simple formulas that describe the output characteristics of magnetrons as was possible for reflex-klystron oscillators. This is because no valid expressions for the equivalent electronic admittance of a magnetron have yet been proposed. From external measurements on magnetrons it has been established that the electronic conductance is negative for conditions of oscillation and decreases in magnitude as the r-f voltage increases, as was the case for reflex-klystron oscillators. However, the electronic admittance evidently depends upon the effective impedance presented by the cavity, whereas in the reflex klystron the beam admittance was independent of the cavity impedance. For this reason the only suitable way of representing magnetron characteristics is by means of a set of contours on some sort of load-impedance coordinates which show the way in which such quantities as efficiency, power output, and frequency depend upon the load impedance. In practice, magnetrons feed loads through transmission lines, and the effective impedance into which the magnetron works is determined by the standing-wave ratio and position of the minimum of voltage on the line. Accordingly, it is convenient to plot magnetron characteristics on special transmission-line-coordinate paper instead of on an impedance plane directly.

Although the exact nature of the electronic admittance of a magnetron (which corresponds to the beam admittance of a reflex-klystron oscillator) is not known, the nature of the conventional representations of magnetron characteristics may be understood from a brief analysis based upon the assumption that the electronic admittance of the magnetron is something like that of the reflex-klystron oscillator. With this assumption the equivalent circuit of a magnetron, resonant cavity, coupling loop, and line terminated in load, is that given in Fig. 18.26. The electronic admittance of the magnetron is represented by the admittance labeled Y_e. It will have a negative conductance component for a condition of oscillation. The electronic admittance is considered to be in shunt with the unloaded resonator, which is represented by a parallel combination of an inductance and capacity, and with the shunt resistance of the unloaded resonator. The resonator is assumed to be inductively coupled to a transmission line leading to a load. The impedance seen looking back into the coupling loop from the line is

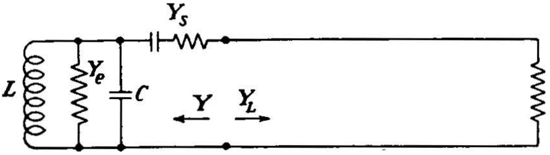


FIG. 18.26.—Simple equivalent circuit of magnetron oscillator, output coupling, line, and load.

$$Z = Z_l + \omega^2 M^2 (Y_r + Y_e) \tag{18.118}$$

where Z_l is the impedance of the coupling loop, ω is the operating angular frequency, M is the mutual impedance between the loop and the resonator, and Y_r is the unloaded admittance of the resonator at the operating frequency. The requirement for oscillation is that the impedance seen looking back into the coupling loop be the negative of the impedance seen looking into the line Z_L,

$$Z_L = -[Z_l + \omega^2 M^2 (Y_r + Y_e)] \tag{18.119}$$

While this equation is not capable of analytical solution, it is capable of graphical representation. This graphical representation will now be developed.

Assume that the electronic admittance is as shown in Fig. 18.27. The locus of the electronic admittance is given by the vector in the second quadrant. For zero r-f voltage the electronic admittance has the value given by the extremity of the vector. As the r-f voltage increases, the vector will be assumed to shrink but maintain its direction. This is not strictly true but will serve for a basis of discussion. It will further be assumed that the electronic admittance is not affected by the resonator admittance. Shown in the same figure is the unloaded resonator admittance. This has a locus that is approximately a straight line parallel to the susceptance axis, as was shown in the chapter on Velocity-modulated Tubes, or Klystrons. Different points along this locus correspond to different frequencies, frequency increasing upward. At unloaded cold resonance the resonator admittance is a pure conductance.

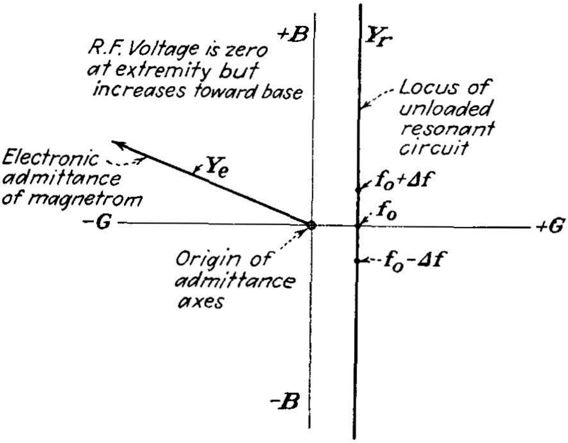


FIG. 18.27.—Loci of electronic admittance of a magnetron and the resonator admittance in an admittance plane.

Shown in Fig. 18.28 is the sum of the resonator and electronic admittance. The electronic-admittance vectors are shifted to the right by the resonator conductance and shifted up or down by departures in frequency from the cold unloaded resonant frequency of the resonator. The locus of $\omega^2M^2(Y_r + Y_e)$ will have the same form as that shown in Fig. 18.28 except that the scale will be changed and the locus will be plotted on an impedance plane with axes of resistance and reactance instead of conductance and susceptance. The locus of

$$Z_l + \omega^2M^2(Y_r + Y_e)$$

is shown in Fig. 18.29. The addition of the loop impedance merely shifts the previous representation strongly upward and a little to the right since the loop will ordinarily have a higher reactance than resistance.

The negative of the impedance seen looking into the loop is shown in Fig. 18.30 against coordinates of load impedance. In this representation it is possible to plot power contours. Ideally, the power will be constant along any vertical line since power output is given by $\frac{V^2}{R}$. Power is expected to be zero along the vertical line through the extremities of the transformed admittance vectors since here the r-f voltage is zero. It is

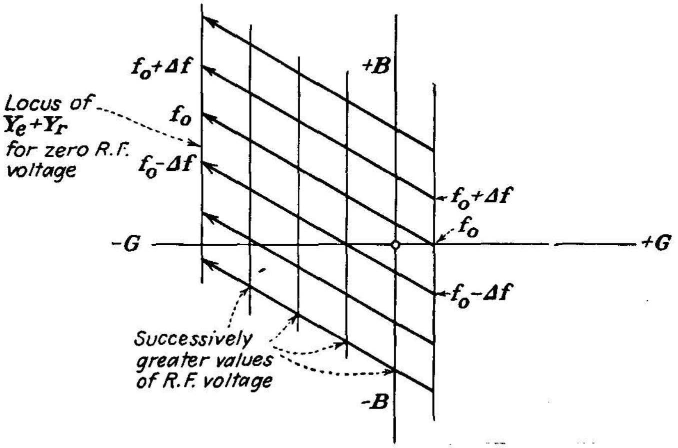


FIG. 18.28.—Sum of resonator and electronic admittance of a magnetron.

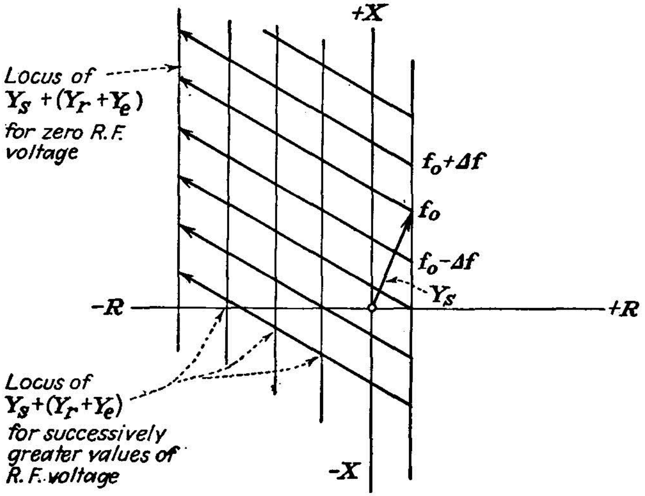


FIG. 18.29.—Locus of $Z_l + \omega^2 M^2(Y_r + Y_e)$ in an impedance plane.

also expected to be zero along the zero-resistance axis since the power that can be delivered to a zero resistance is zero. Between these limits the power will rise to a maximum. Actual contours of constant power are not straight vertical lines but elongated closed loops, for the electronic admittance of the tube apparently changes with the admittance into which the electrons work. Such elongated loops are shown dotted. They are closed about a point of maximum power output. Also shown in Fig. 18.30 are lines of constant frequency in the form of the transformed electronic-admittance vectors. For the reflex-klystron oscillator the slope of this line is related to the transit time in the repeller space. For the magnetron the slope of these lines is also probably related to the mean transit time, though the exact relation has not been definitely established. Shown in this same figure are some loci of constant standing-wave ratios on a transmission line that will produce the indicated load impedance. These loci are circles about the characteristic impedance of the line.[1]

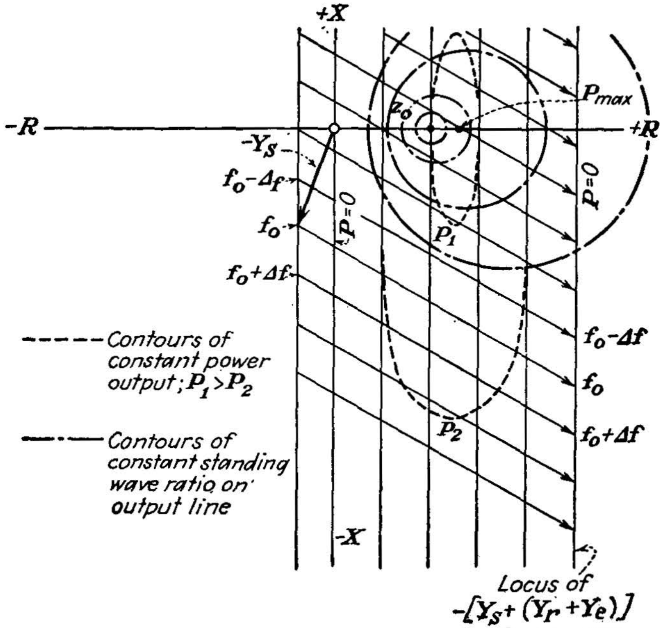


FIG. 18.30.—Locus of $-[Z_l + \omega^2 M^2 (Y_r + Y_e)]$ in an impedance plane.

---

[1] See KING, R. W. P., H. R. MIMNO, and A. H. WING, "Transmission Lines, Antennas and Wave Guides," McGraw-Hill, New York, 1945, for an introductory treatment of transmission-line impedance loci.

Along any circle about the characteristic impedance of the line the magnitude of the standing wave of voltage or current is constant, but the distance from the magnetron output to the minimum of voltage changes.

The contours of Fig. 18.30 are usually transformed to a representation on which the circles of constant standing-wave ratio are concentric about

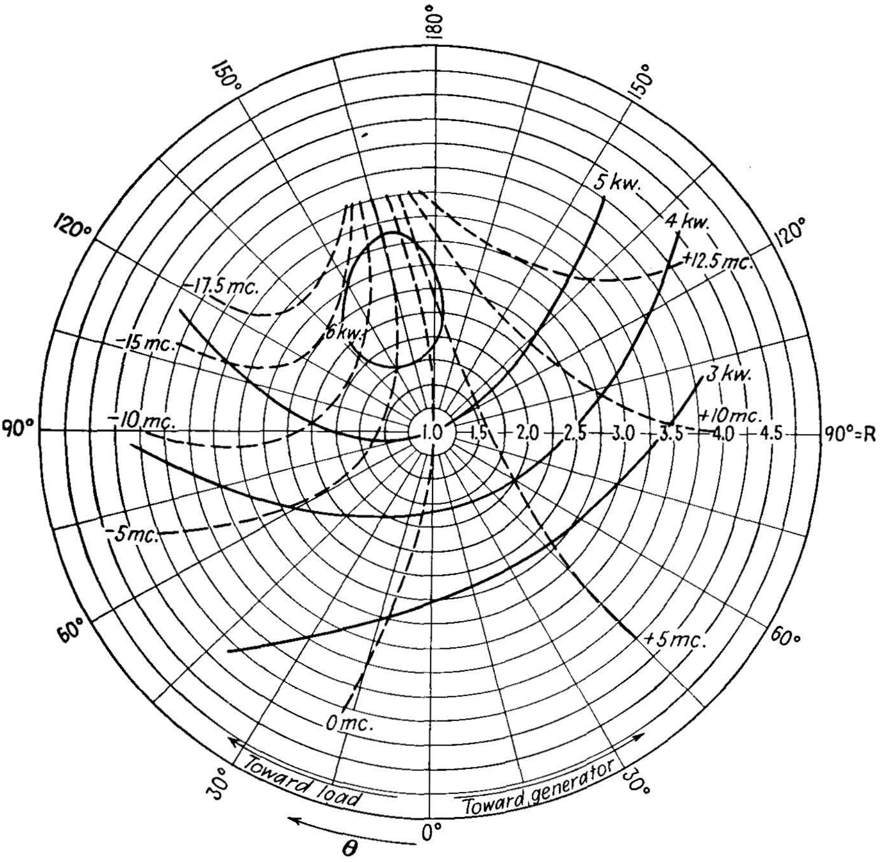


FIG. 18.31.—Rieke diagram of a Raytheon 2J38 magnetron.

the center of the plot and evenly spaced on a radial scale, the value of the standing-wave ratio at the center being unity. The contours of Fig. 18.30 are correspondingly deformed to give the representation of Fig. 18.31, which is known as a *Rieke diagram*. Positions of constant distance of a voltage minimum from the output loop in electrical degrees become radial straight lines in such a plot. The contours of constant power output are closed contours about a point of maximum output, though for low powers the contours are closed off the chart through regions of voltage standing-wave ratios greater than 5. Contours of constant

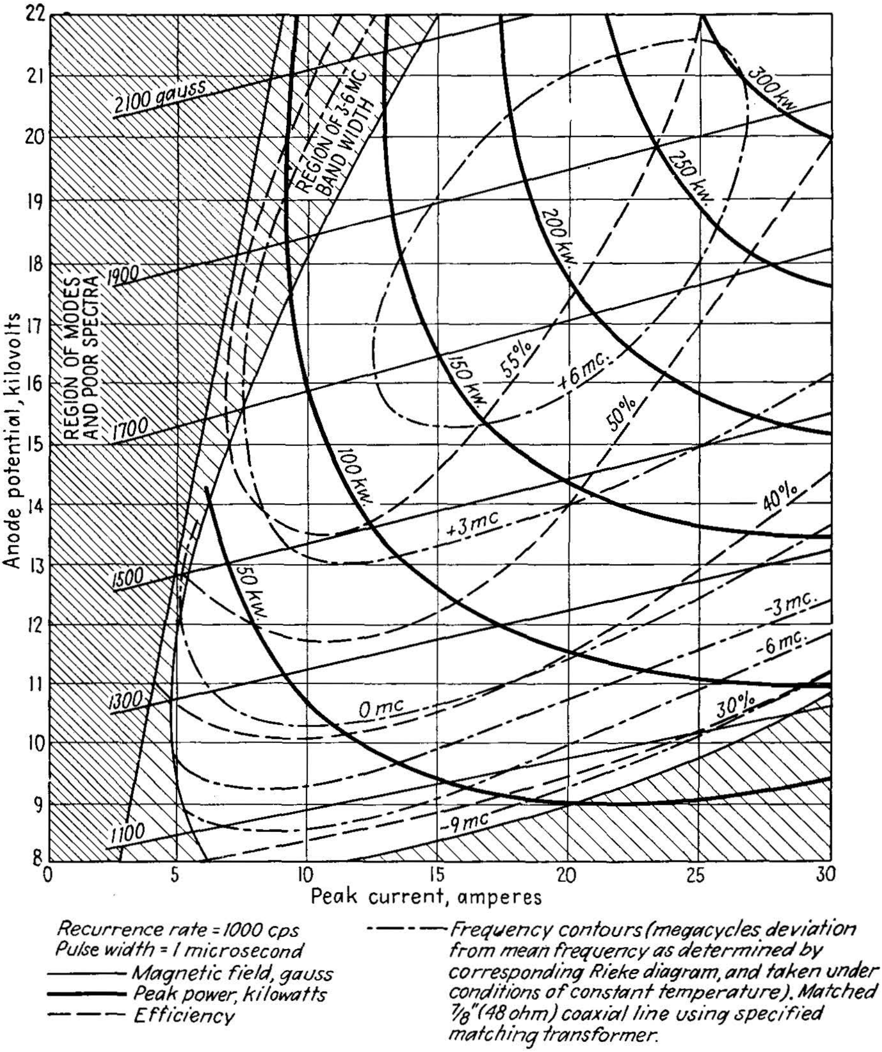


FIG. 18.32.—Voltage-current characteristics of a Raytheon 2J30 magnetron.

frequency are shown in this diagram, corresponding to those in Fig. 18.30. Ideally, these would intersect the constant-power contours at a constant angle. It is seen that for a load corresponding to a given standing-wave ratio of voltage and location of voltage minimum from the output loop the power output and frequency are specified. The position of the point of maximum power output relative to the center of the chart is determined primarily by the design of the output-coupling loop. It is not always desirable to operate the tube at maximum output, for here the frequency changes relatively rapidly with changes in load impedance, an effect known as *frequency pulling.* Accordingly, the output loop is usually designed so that the center of the chart falls at a point in the characteristic field which represents a suitable compromise between output and high-frequency stability. The amount by which the frequency changes for a given standing-wave ratio as the position of the minimum of the standing wave is changed is a figure of merit for the tube; the less the frequency change, the better the tube. In a good tube the amount of the frequency variation at a standing-wave ratio of 1.5 is less than $\frac{1}{10}$ of 1 per cent. Rieke diagrams are usually plotted for a condition of constant plate current and constant magnetic-flux density, the voltage being varied slightly to keep the current constant as the load is changed. The advantage of the Rieke diagram over other possible representations is that a change in the reference point from which the standing-wave maxima and minima are measured merely rotates the plot without changing its form.

Another representation of magnetron characteristics that is commonly given is a voltage-current plot as shown in Fig. 18.32. On this plot, known as the "performance characteristic," there are shown contours of constant magnetic field, output, efficiency, and frequency. The controlled variables are the magnetic field and the voltage, which determine the current and at which frequency, power output, and efficiency can be measured. Such plots are made for a constant load impedance, usually a flat line of the proper characteristic impedance.

CHAPTER 19

PHOTOELECTRIC TUBES

19.1. The General Form of Photoelectric Tubes. Photoelectric tubes, or, as they are now more frequently referred to, "phototubes," are at first glance very simple devices, though the preparation of the photosensitive surface involves some of the most delicate operations in modern electronic practice. The tube is generally housed in a small glass envelope and contains, in its simplest form, just two electrodes. The cathode, or photosensitive emissive surface, is usually in the form of a half cylinder. The anode, or electron collector, is usually in the form of straight wire on the axis of the cylindrical cathode. Great pains are taken to make the leakage resistance between the two electrodes as high as possible. In some tubes the leads to the two electrodes are brought out at different ends of the tube in order to achieve a high leakage resistance. The envelope of the tube is usually made of a special glass, which acts as a light filter to make the light absorption as low as possible in the desired light frequency band.

Applications of the phototube are too well known to require much discussion. Phototubes can be used to activate almost any kind of electrical or mechanical device through the medium of suitable amplifiers and relays. They can be used to cause a device to respond to almost any variation in light intensity. They can be made to respond to light of any color in the visible spectrum and to respond as well to radiation in the infrared and ultraviolet portions of the spectrum. Applications as door openers, counters, automatic light switches, and color sorters are well known.

19.2. Fundamental Photoelectric Relations. Phototube operation is based upon what are now the well-established properties of the photoelectric effect. These may be enumerated as follows:

1. Electrons are emitted from low-work-function surfaces when exposed to radiations in the visible or near-visible region of the spectrum.
2. The magnitude of the emitted photoelectric current is proportional to the intensity of the illumination.
3. Photoelectrons are emitted with finite velocities. The maximum

velocity of emission is independent of the intensity of the illumination of the emitting surface (time rate of flow of radiant energy).

4. Any photoemissive surface has a low-frequency limit of radiation beyond which no electrons are emitted regardless of the intensity of illumination.
5. The emission velocity of photoelectrons depends upon the work function of the emissive surface as well as upon the frequency of the illuminating radiation.

These various properties will be described in some detail in subsequent sections.

19.3. History of the Photoelectric Effect. The history of the discovery, theoretical development, and experimental verification of the photoelectric effect is so fascinating that it deserves at least a topical recapitulation. It is all the more remarkable in that the fundamental relations of the photoelectric effect were established before the existence of the electron was verified! Chronologically, the high spots in the history of the photoelectric effect are somewhat as follows:

1887 Hertz discovered the photoelectric effect in his experiments on electromagnetic waves. His experiments dealt with observations on the transmission of damped electromagnetic waves of a frequency of about 1,500 mc, generated with a spark coil and a suitable resonant circuit. Transmitted energy was picked up on a resonant circuit, and the intensity of the transmission was observed on a spark gap adjustable with a micrometer. Hertz found that his receiving circuit sparked more readily when the electrodes were illuminated by the spark from the transmitting gap. He further found that the effect was present only when the negative electrode (the gaps were polarized with a direct voltage) was illuminated. He verified that ultraviolet radiations were responsible for the effect and that the effect was independent of the source of the radiations.

1888 Hallwachs established that the effect consisted in the emission of negative particles of electricity.

1889 Elster and Geitel showed a relation between the contact potential of a surface and its long-wave limit of photoemission. They built the first photocells and made the first photometer.

1889 J. J. Thompson discovered the electron as a fundamental particle and constituent of matter. He established that the negative particles emitted from incandescent bodies were the same as the particles emitted photoelectrically. He deflected electrons electrically and magnetically and made the first determination of the ratio of the charge to the mass of the electron.

1889 Lenard showed that the magnitude of the photoelectric current was proportional to the intensity of the exciting illumination. He also discovered that the velocity of emission of photoelectrons was independent of the intensity of exciting illumination.

1905 Einstein applied the quantum theory enunciated by Planck in 1900 to the photoelectric effect. He predicted correctly the relation between the velocity of emission of photoelectrons, the work function of the emitting surface, and the frequency of the exciting radiation.

1912 Hughes verified the Einstein equation.

1916 Millikan checked the values of Planck's constant by photoelectric measurements.

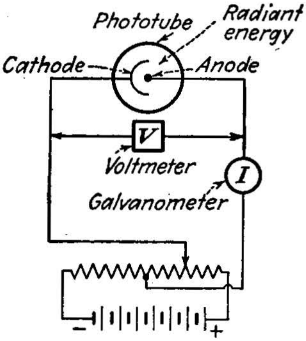


FIG. 19.1.—Circuit for observing the photoelectric effect.

19.4. Specific Photoemission Characteristics. The photoemissive properties of surfaces are usually investigated by means of the arrangement of Fig. 19.1. Here the phototube is shown by the circle containing a photosensitive cathode and an anode. The cathode is illuminated from an external source. The cathode and anode are connected to a source of direct potential in such a way that the anode can be made either positive or negative relative to the cathode. A sensitive current meter is connected in series with the tube and voltage source.

With the arrangement of Fig. 19.1 the current registered by the meter is a function of the intensity of the light and the electrode voltages, as shown in Fig. 19.2. From this figure it is observed that for any anode voltage positive relative to cathode voltage the photoelectric current is directly proportional to the intensity of the illumination. Let the difference between anode and cathode voltage be designated by V. (V includes the effect of contact potential.) Then for positive values of V the photoelectric current is constant for a fixed illumination. This means that the photoemission is constant and that the anode is collecting all the photoelectrons. When V is made negative at a fixed illumination, the current falls off, reaching zero at a value of V_0 that is independent of the intensity of the illumination. The explanation of these effects is apparently that photoelectrons are emitted with velocities ranging from zero to some maximum value. The number of electrons emitted is proportional to the rate of incidence of radiant energy, but the maximum velocity of the emitted electrons is independent of the intensity of illumination of a given spectral distribution.

The maximum velocity of emission does, however, depend upon the frequency of the light, as may be shown by illuminating the photoemissive surface with monochromatic light of a variable frequency but constant intensity. The results of such a test are shown in Fig. 19.3.

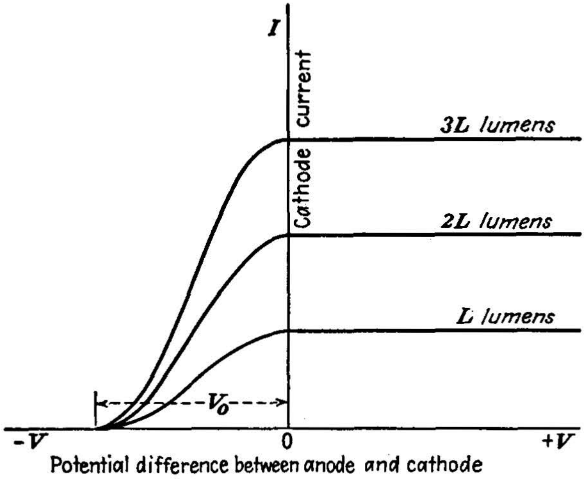


FIG. 19.2.—Photoemission current versus retarding voltage for various intensities of illumination.

The three curves shown give current against retarding voltage for equal intensities of illumination of three different frequencies of light such that $f_1 > f_2 > f_3$. The higher the frequency of the light, *i.e.*, the farther toward the short-wave-length (blue) end of the spectrum, the greater the maximum velocity of emission. Curves such as those of Fig. 19.3 are rather difficult to obtain, for it is necessary to measure radiant energy with a thermocouple or bolometer, correct the resultant curves for contact potential, stray light, and secondary emission, and be sure that the emissive surfaces are free from any contamination and totally outgassed.

The relation between the maximum velocity of emission and the frequency of the exciting radiation is given in Fig. 19.4. This shows that the maximum energy of emission of photoelectrons is linear with the frequency of the exciting radiation. There is a minimum frequency of light for any surface beyond which photoelectrons are simply not emitted. The curve of Fig. 19.4 is a good straight line, which Millikan has shown comes down to the axis with a definite angle and not asymptotically. The straight line of Fig. 19.4 may be represented by the equation

$$V_0e = \tfrac{1}{2}mv_m{}^2 = hf - w \tag{19.1}$$

where $-V_0$ is the intercept with the voltage axis of any curve in Fig. 19.3, v_m is the corresponding maximum velocity of emission, m is mass of the electron, $-e$ is charge of the electron, f is light frequency in cycles per second, h is the slope of the straight line of Fig. 19.4, and w is the frequency axis intercept of the straight line of Fig. 19.4. As given above, Eq. (19.1) is purely empirical. However, Eq. (19.1) is the equation predicted by Einstein on purely theoretical grounds, with h identified as Planck's constant and w identified as the work function of the photoemissive surface in electron volts. From Eq. (19.1), the minimum frequency of emission occurs when the velocity of emission is zero and is given by

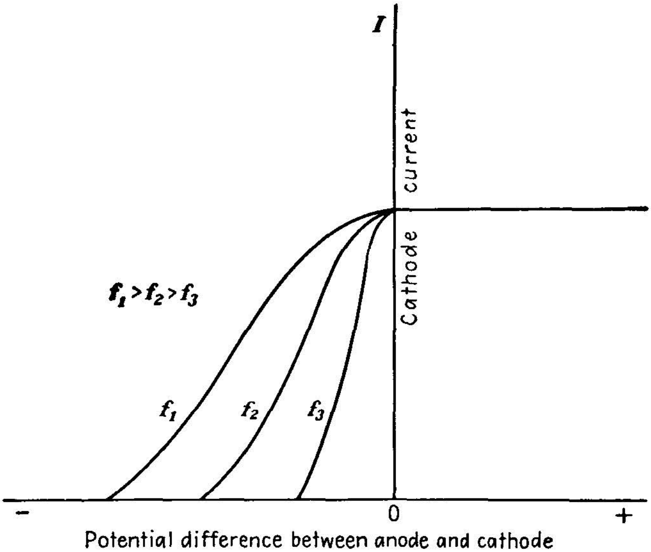


FIG. 19.3.—Photoemission current versus retarding voltage for various frequencies of illumination.

$$f_0 = \frac{w}{h} \tag{19.2}$$

The relation between the work function and the minimum frequency or maximum wave length of exciting radiation predicted by Einstein in the form of Eq. (19.1) has been verified experimentally. If we let

$$w = e\phi_p \tag{19.3}$$

where w is work function in electron volts, e is electron charge, and ϕ_p is the voltage equivalent of the work function as determined from photoelectric measurements, then the work function in volts should be inversely proportional to the maximum wave length in angstrom units. In Fig. 19.5 is given a plot on log-log paper of the relation between experimentally observed values of the thermionic work function and the maximum wave length of photoelectric emission for different materials. If the relation predicted by Einstein is correct, then the plot of the work function against the threshold wave length on log-log paper should be a straight line with a slope of −1. Reference to Fig. 19.5 shows that this relation is obeyed fairly well. Departures from the relation postulated are primarily due to the difficulty of getting an uncontaminated emitting surface. There are also some discrepancies due to a correction which must be made for the temperature of the emitting surface. The most extensive work in trying to correlate values of the work function as measured by thermionic and photoelectric methods has been done on platinum. It is the consensus of workers in this field that the photoelectric and thermionic work functions of platinum

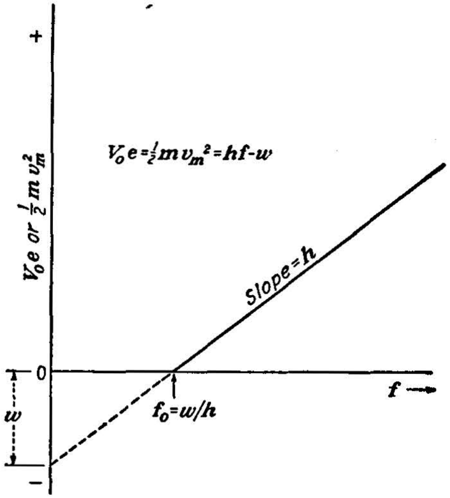


FIG. 19.4.—Maximum velocity of emission of photoelectrons as a function of frequency of exciting radiation.

are the same and that those of other metals would be revealed as the same if the measurements were sufficiently refined.[1,2]

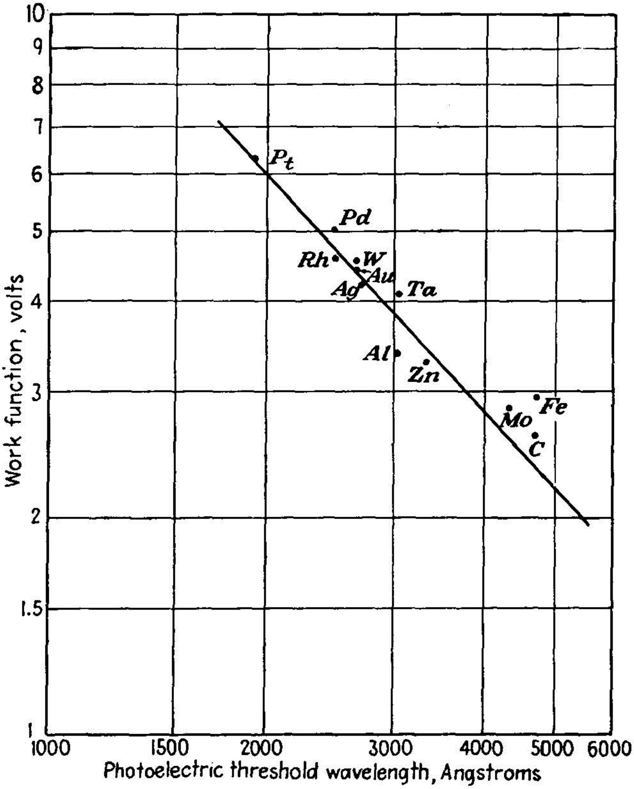


FIG. 19.5.—Relation between the thermionic work function of different metals and the threshold wave length of photoemission.

19.5. Fundamental Theory of Photoemission. The wave theory of light meets with considerable difficulty in explaining the various aspects of the photoelectric effect. The proportionality between the photoelectric current and the intensity of illumination is consistent with the wave theory, but the fact that the maximum velocity of emission is independent of the intensity of the illumination cannot be explained on the basis of the wave theory of light. When the independence of the velocity of emission

---

[1] The classical reference on all phases of photoelectricity is HUGHES, A. L., and L. A. DUBRIDGE, "Photoelectric Phenomena," McGraw-Hill, New York, 1932.

[2] An excellent elementary survey of the photoelectric effect is contained in RICHTMYER, F. K., and E. H. KENNARD, "Introduction to Modern Physics," 3d. ed., McGraw-Hill, New York, 1947.

and the intensity of illumination was discovered a furor was created among students of physics. The dilemma encountered in trying to explain the above-mentioned effect can be circumvented by postulating the dual nature of light; that is to say, light rays exhibit both a wave and a particle aspect. The wave nature of light cannot, however, be completely discarded on the assumption that light is corpuscular in nature, for some aspects of light behavior are very difficult to explain on this basis.

The corpuscular aspect of light rays has its basis in the quantum theory. The quantum theory had its origin in the study of heat-radiation phenomena. The quantum theory has proposed that energy flows, not continuously, but rather in small packages. The smallest unit of energy that can be involved in any transfer is called the "quantum." A quantum of energy has a size that is directly proportional to the corresponding frequency of radiation as given by

$$Q = hf \tag{19.4}$$

where Q is the quantum of energy, h is a universal constant having a value of 6.624×10^{-34} watt-second per cycle and known as "Planck's constant," and f is the frequency of the radiation in cycles per second. Thus if monochromatic orange light of wave length 6,000 angstrom units is involved (1 angstrom unit = 10^{-10} meter), the corresponding frequency of radiation is 5×10^{16} cycles per sec and the corresponding quantum of energy *for this frequency* is 33.12×10^{-18} watt-second. This means that light of this frequency delivers energy in units of 33.12×10^{-18} watt-second and cannot deliver any but an integral multiple of this amount of energy. Thus, just as the modern theory of matter postulates the indivisible particle, the electron, so, correspondingly, the quantum theory says that energy is finally delivered in minute but indivisible units of quanta.

A quantum of light is known as a *photon*. Light rays may be considered to be made up of photons, which have many of the characteristics of small particles in that each carries a discrete quantity of energy but which also have the characteristics of waves. When the quantum theory is applied to light rays, all the effects observed in connection with photoemission are readily explained.

If light rays consist of photons each of which carries a definite quantity of energy proportional to its frequency, then each photon on striking a surface may transfer to an electron in the surface *at most* a quantum of energy. This quantum of energy may give rise to emission of an electron, and the energy that the emitted electron will have will be *at most* the quantum of energy *minus* the work necessary to overcome the surface electrostatic forces. Hence the validity of Eq. (19.1), the Einstein photoelectric equation,

$$V_0e = \tfrac{1}{2}mv_m{}^2 = hf - w \tag{19.1}$$

The work necessary to overcome the surface electrostatic forces, w, is the work function of the metal in question.

By applying the quantum theory of light, all the photoelectric effects observed experimentally are completely explained. The threshold frequency of photoemission is that frequency at which the energy of the photon is converted into electron energy enabling the electron to just barely overcome the surface restraints and thus be emitted with zero velocity. The threshold frequency is accordingly proportional to the work function of the metal, as previously noted. The proportionality between photoelectric current and intensity of illumination follows from the fact that the *number* of photons is proportional to the intensity of the illumination for a given area.

19.6. Spectral Response Curves of Photoemissive Surfaces. The photoelectric emission of metal surfaces exhibits two important kinds of selectivity The first selectivity is a variation in emitted current with wave length of the exciting radiation. The second shows itself as a variation in emitted current with the polarization of the exciting radiation. The response to polarized light is much smaller when the electric vector of the exciting radiation is parallel to the surface than when the light is polarized at right angles to the surface. Of the two types of selectivity the first is by far the more important since ordinary photoemissive surfaces as used in commercial tubes are so rough that no differentiation with respect to polarization can be observed.

Every photoemissive surface exhibits peaks of sensitivity as the wave lengths of the exciting radiation are changed. Typical of the response characteristics of the pure metals are the curves for the alkali metals shown in Fig. 19.6.[1] Observation of these curves shows that, as the atomic number of the element increases, the maximum sensitivity decreases, the resonance peak becomes broader, and the wave length of the maximum sensitivity increases. No completely satisfactory quantitative explanation for the above relations seems to be available. The threshold wave length increases as the work function of the surface decreases in accordance with Einstein's photoelectric equation. Qualitatively it is expected that the wave length of maximum sensitivity would follow somewhat the same relation. An investigation from the point of view of quantum-mechanical considerations will no doubt some day give

---

[1] Seiler, E. F., Color-sensitiveness of Photo-electric Cells, *Astrophys. Jour.*, vol. 52, pp. 129–153, October, 1920.

the complete story. It is *possible* that the mechanism involved is similar to that which occurs for secondary emission, which yields a maximum emitted current for a given energy of excitation.

It is possible to make complex emitting surfaces that have lower work functions than the pure metals. The surface that gives maximum secondary emission also seems to give maximum photoelectric emission. Maximum emission is obtained with a surface of the type caesium on caesium oxide on silver.[1] Such a surface is prepared by oxidizing silver and then exposing it to caesium vapor. Photoemissive surfaces may also be prepared by sputtering metals, vaporizing metals, and electrolyzing metals through a glass envelope.

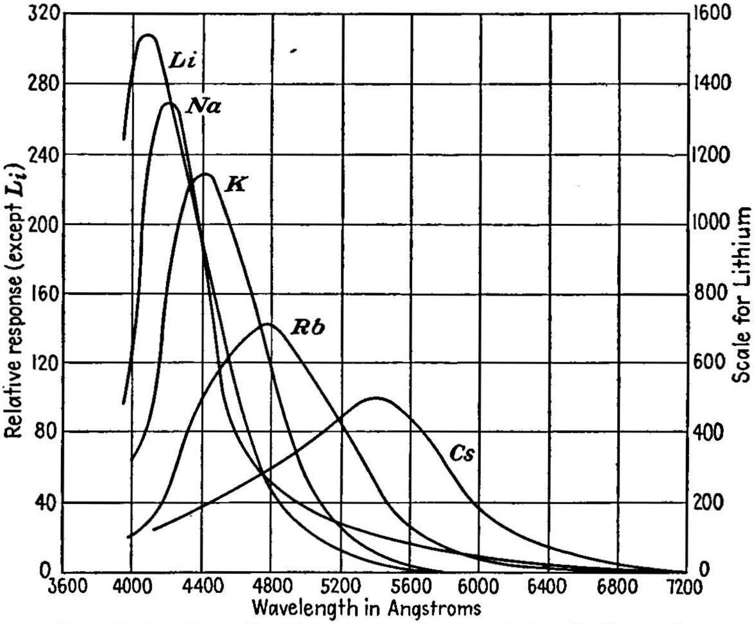


FIG. 19.6.—Photoelectric color sensitivity of the alkali metals.

19.7. Vacuum-phototube Characteristics. *Current-voltage Characteristics.* Vacuum phototubes exhibit characteristics that depend primarily upon the nature of the emissive surface and the transmission characteristics of the glass envelope. A typical set of vacuum-phototube characteristics is shown in Fig. 19.7. For a fixed amount of light flux from the exciting source the curves of current against voltage are similar to those of a diode. For very low voltages the current follows the three-

---

[1] ZWORYKIN, V. K., and G. A. MORTON, "Television," pp. 22–28, McGraw-Hill, New York, 1940.

halves-power law of variation with voltage. Because the emission current from a photosensitive surface is so small, this region is extremely small and most of the curve of current against voltage shows pronounced emission saturation. As a result, the emission current is almost constant over nearly the entire operating range. A load line may be constructed on the current-voltage characteristics of a phototube just as is done on a set of vacuum-tube characteristics. Several such lines are shown in Fig. 19.7. These lines have a slope that is the *negative reciprocal of the resistance* in series with the voltage supply and the phototube. Such load lines will always be straight lines regardless of the current-voltage characteristics of the device since they are simply a graphical representation of Ohm's law. The proportionality between current and light flux is almost exactly linear for any operating voltage, as shown in Fig. 19.8.

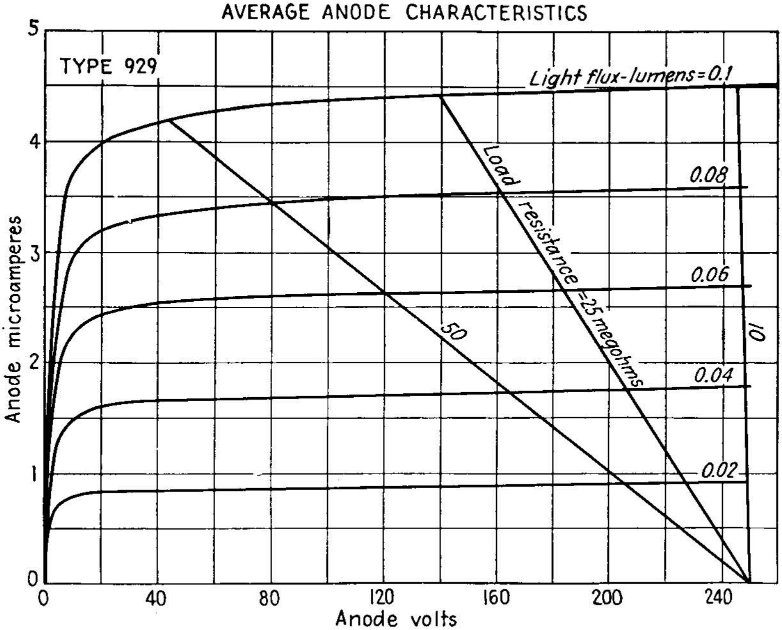


Fig. 19.7.—Current-voltage characteristics of the RCA 929 phototube (light from a tungsten filament at 2870°K).

The reaction of a photoemissive surface to illumination is almost instantaneous. Experiments show that less than 3×10^{-9} sec elapse from the time the photoemissive surface is illuminated until photoemission begins. The photoelectric current ceases in less than 10^{-8} sec after the illumination is cut off. Hence in a vacuum phototube the principal time factor involved is the transit time of an electron from cathode to anode. The transit time may be calculated from the curves of Fig. 8.14. This time will generally be very short.

Example: Determine the transit time of a photoelectron emitted from a semicylindrical cathode of radius 1 cm and collected at an anode of radius ½ mm. The ratio of cathode to anode radius is 20, and the distance between cathode and anode surfaces is 0.95 cm. From Fig. 8.14, the factor K is 1.344. Let the anode potential be 200 volts. Then the electron velocity at the anode is 243.5×10^6 cm per sec. The corresponding transit time from the formula $T = \frac{Kd}{v}$ is 0.00525 microsecond. This means that a vacuum phototube can handle any known type of light modulation.

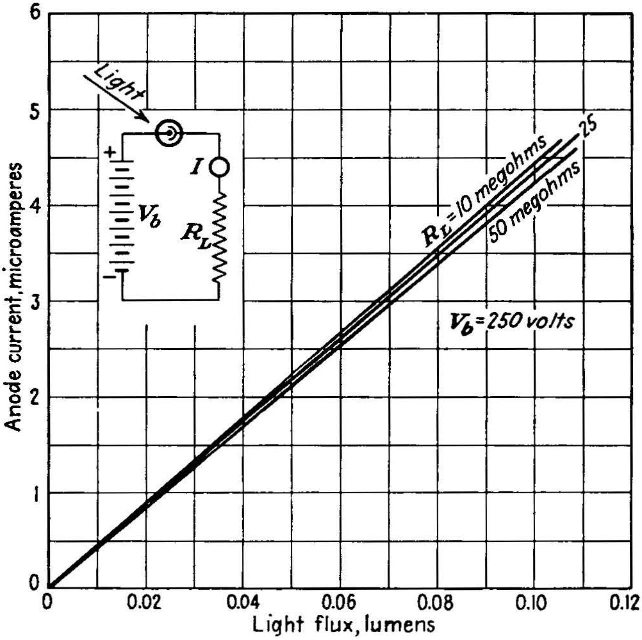
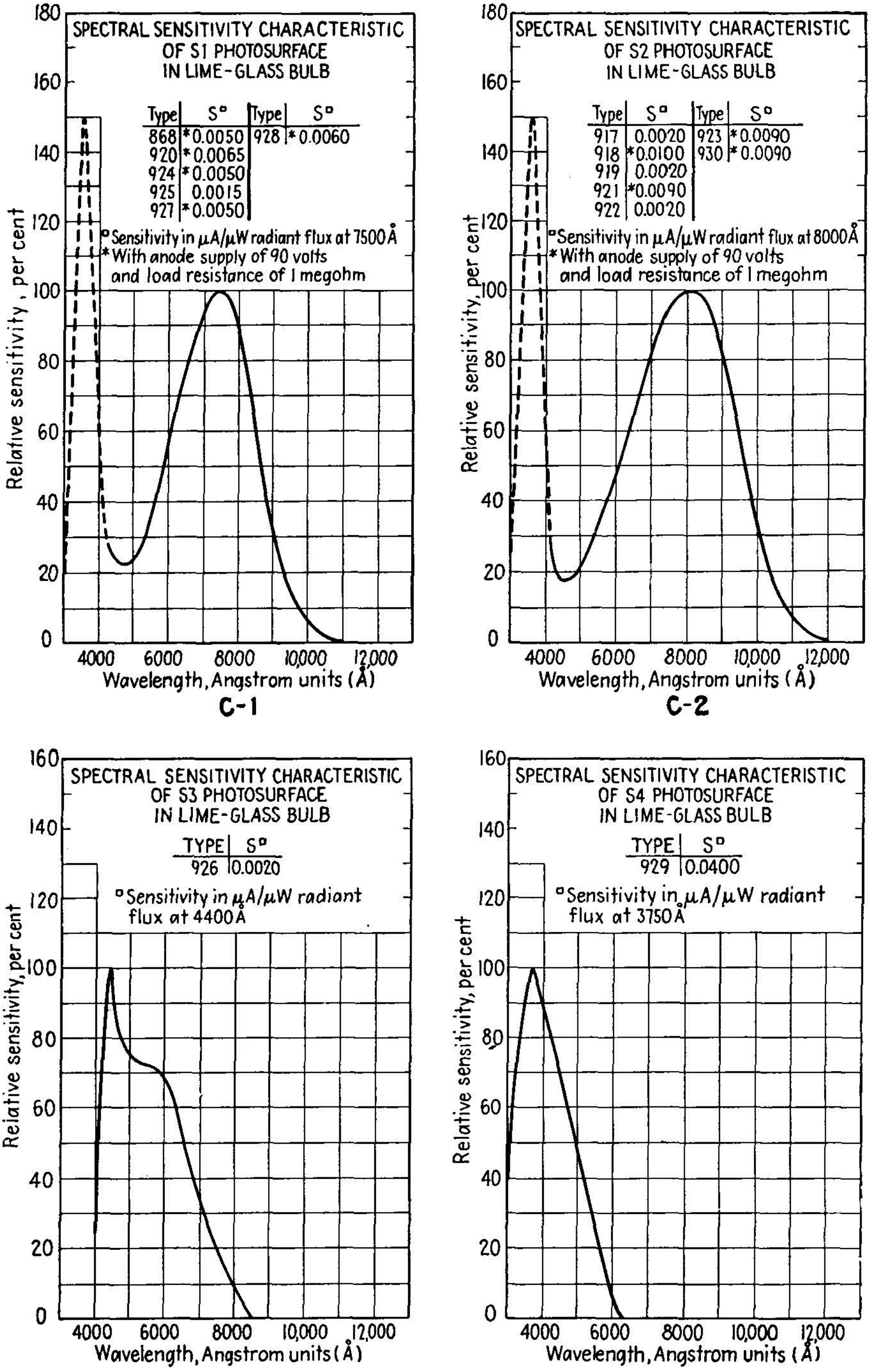


FIG. 19.8.—Current versus light flux of an RCA 929 phototube.

Spectral Characteristics. Phototubes are available with spectral sensitivities that cover the visible portion of the spectrum and carry well into the infrared and ultraviolet. In general, the response curves will be different from that of the eye, which is shaped something like a resonance curve, with a peak at 5,550 angstrom units (1 angstrom unit = 10^{-10} meter) and dropping to virtually zero at 4,000 and 7,000 angstrom units. Some typical spectral response curves of commercial phototubes are shown in Fig. 19.9. It is seen that there are tubes available which cover the visible spectrum, the short infrared rays, and the long ultra-



SPECTRAL SENSITIVITY CHARACTERISTIC OF S1 PHOTOSURFACE IN LIME-GLASS BULB

Type	S°	Type	S°
868	*0.0050	928	*0.0060
920	*0.0065		
924	*0.0050		
925	0.0015		
927	*0.0050		

°Sensitivity in μA/μW radiant flux at 7500 Å
*With anode supply of 90 volts and load resistance of 1 megohm

C-1

SPECTRAL SENSITIVITY CHARACTERISTIC OF S2 PHOTOSURFACE IN LIME-GLASS BULB

Type	S°	Type	S°
917	0.0020	923	*0.0090
918	*0.0100	930	*0.0090
919	0.0020		
921	*0.0090		
922	0.0020		

°Sensitivity in μA/μW radiant flux at 8000 Å
*With anode supply of 90 volts and load resistance of 1 megohm

C-2

SPECTRAL SENSITIVITY CHARACTERISTIC OF S3 PHOTOSURFACE IN LIME-GLASS BULB

TYPE	S°
926	0.0020

°Sensitivity in μA/μW radiant flux at 4400 Å

C-3

SPECTRAL SENSITIVITY CHARACTERISTIC OF S4 PHOTOSURFACE IN LIME-GLASS BULB

TYPE	S°
929	0.0400

°Sensitivity in μA/μW radiant flux at 3750 Å

C-4

FIG. 19.9.—Spectral response curves of typical RCA phototubes.

violet rays. The majority of phototube applications depend upon a tungsten filament as a source of illumination. The tungsten filament has its spectral characteristic centered in the infrared range, with appreciable radiation in the visible portion of the spectrum. Light filters may be used with phototubes where selective response with respect to color is desired. Where high sensitivity in the ultraviolet is desired, special envelopes must be used with the tube, for the ordinary glass does not transmit ultraviolet rays well. Such special envelopes usually take the form either of a glass envelope with an extremely thin window in front of the cathode or of a quartz envelope.

The spectral sensitivities of vacuum phototubes range from about 5 to 50 microamperes per lumen (1 lumen = 0.0016 watt for green light). The number of lumens, L, of light flux falling upon an area A of a surface a distance d from a point source of light of candle-power strength C is

$$L = \frac{CA}{d^2} \qquad \text{lumens} \tag{19.5}$$

where any units of length may be used provided only that they are the same for A and d^2.

19.8. Gas-phototube Characteristics. The sensitivity of a phototube can be increased by utilizing what is known as the *gas amplification* of the photoemission current. If a small amount of gas of the right kind and pressure is admitted into the phototube, then the photoelectrons in their travel from cathode to anode will strike some of the gas molecules, causing ionization. This ionization splits the gas molecule into a free electron and a negative ion. The free electron is now available to join the photoelectron in its travel toward the anode and may itself ionize other gas molecules, giving rise to more electrons, which can add to the effective current of the phototube. The positive gas ions formed will move toward the cathode and, in doing so, will constitute a current that is nearly equal to the electron current. In addition, the positive ions on impact with the cathode will create some secondary electrons, which will further increase the total current. As a result of the cumulative action of all the above effects, the net current to the anode of the phototube can be made as much as *ten times* the photoemission current.

The current-voltage characteristics of a typical gas phototube are shown in Fig. 19.10. For low anode voltages the characteristics are about the same as for the vacuum phototube, for at low voltages there is inappreciable ionization owing to the low energies of the photoelectrons. At higher anode voltages, ionization occurs, and the current increases rather rapidly with voltage. At sufficiently high voltages a glow discharge will be sustained between electrodes, as shown in Fig. 19.10, and the tube operation is impaired. Some appreciable departures from current linearity with light intensity are expected in the gas phototube and are indeed present, as shown in Fig. 19.11. The distortion resulting from this nonlinearity of the characteristics is, however, no greater than that encountered in ordinary vacuum tubes and does not prevent gas phototubes from being used to reproduce the sound recorded on film.

Factors in the Design of Gas Phototubes. There are a number of rather critical factors that must be properly adjusted in the gas photocell to obtain a good tube. These may be listed as follows:

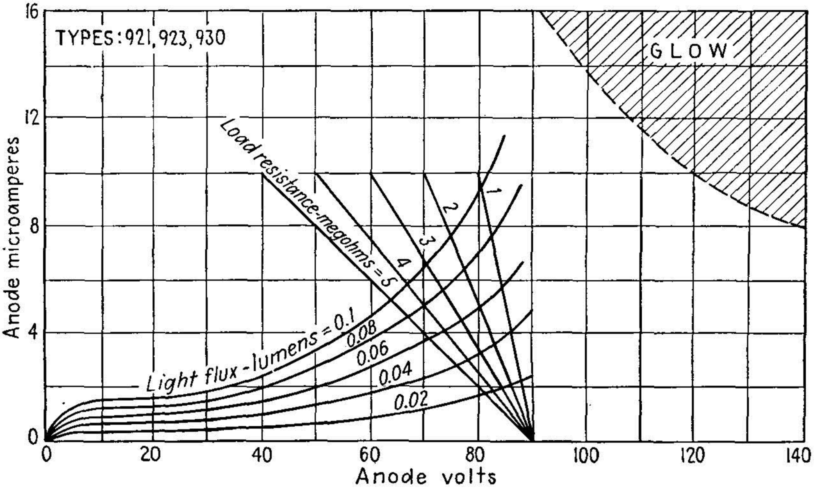


FIG. 19.10.—Current-voltage characteristics of a typical gas phototube.

1. Chemical properties of the gas.
2. Atomic weight of the gas.
3. Pressure.
4. Maximum allowable voltage.

The principal consideration involved in the choice of a gas is that it must not react with the photoemissive surface. The only gases that can be depended upon not to react with caesium surfaces are the inert gases helium, neon, argon, krypton, and xenon.

The atomic weight of the gas used is a factor, for if the gas is too heavy the transit time of the positive ions formed will be too great and the high-frequency response of the phototube will be poor. Correspondingly, the ionization potential, or potential of a striking electron that will free an electron from the gas molecule, must be low; otherwise, the potential across the tube will be so high that the cathode emission may be impaired by the bombardment of high-energy positive ions. The critical physical characteristics of the inert gases are listed below:

Gas	Atomic weight	Ionization potential, volts	Molecular diameter, cm
Helium	4.002	24.46	1.9×10^{-8}
Neon	20.183	21.47	2.35×10^{-8}
Argon	39.944	15.68	2.9×10^{-8}
Krypton	82.9	13.96	3.2×10^{-8}
Xenon	130.2	12.08	3.5×10^{-8}

From this tabulation it is seen that as the atomic weight decreases the ionization potential increases. A compromise must therefore be effected in realizing the requirements of low atomic weight and low ionization potential. The properties of argon represent a reasonable compromise, and this gas is the one most commonly used, though other gases may be and sometimes are used in special applications.

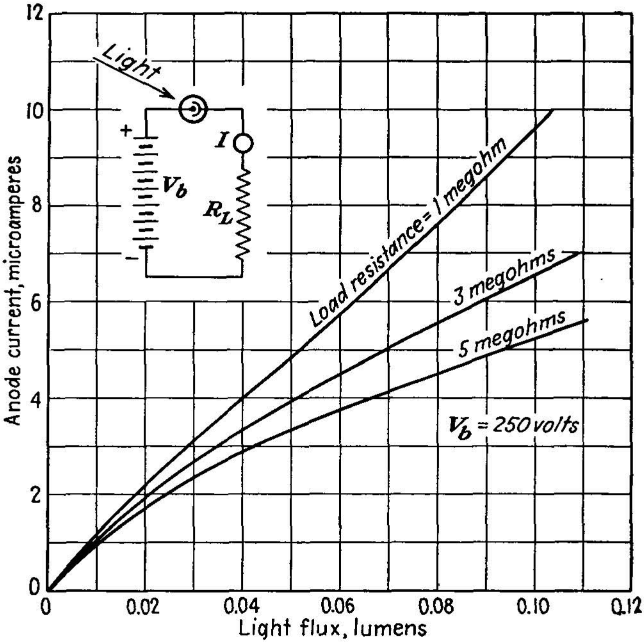


FIG. 19.11.—Current as a function of light flux in a typical gas phototube.

The gas amplification that can be realized in a gas phototube depends upon the gas pressure and the voltage involved. These factors determine the number of ionizing collisions of a photoelectron. The greater the pressure, the less the average distance between molecules but correspondingly the less energy the electron has at each collision. The average distance between collisions of molecules or electrons in a gas is known as the "mean free path." The mean free path of an electron moving among gas molecules is in turn related to the pressure, or number of molecules per cubic centimeter, and to the molecular diameter of the gas molecules by the relation

$$\text{Mean free path} = \frac{4}{\pi d_m{}^2 n} \quad \text{cm} \tag{19.6}$$

where d_m is the molecular diameter in centimeters and n is the number of molecules per cubic centimeter.[1] The number of molecules per cubic centimeter of a gas depends only upon the pressure and the temperature and is independent of the gas involved,

$$n = 7.244 \times 10^{15} \frac{P}{T} \tag{19.7}$$

where P is pressure in bars or dynes per square centimeter

(1 atmosphere = 10^6 bars = 760 mm of mercury)

and T is temperature in degrees Kelvin (273 + C°). Combining Eqs. (19.6) and (19.7) for argon and assuming room temperature to be 290°K,

$$\begin{array}{r}\text{Mean free path of electron} \\ \text{among argon molecules}\end{array} = \frac{60.7}{P} \quad \text{cm} \tag{19.8}$$

where P is in dynes per square centimeter or bars.

A pressure of 0.2 mm of mercury is commonly used in gas phototubes. This corresponds to a pressure of 263 bars and a mean free path of 0.23 cm. At every ionizing collision a new free electron is created that can itself produce more electrons by collision. Thus, if the original photoelectron in traveling from cathode to anode experiences n collisions each of which produces a single free electron, then 2^n free electrons reach the anode for each photoelectron emitted. The potential distribution must be such that each electron acquires enough energy to ionize another molecule in a distance equal to or slightly less than the mean free path. From the above figures it is seen that with a linear potential field it would be necessary to have a cathode-anode spacing of only about 0.8 cm and a total potential of only about 64 volts to ensure a gas amplification of at least fifteen times (since for every electron formed a positive ion is also formed that contributes to the current).

---

[1] Dow, W. G., "Fundamentals of Engineering Electronics," pp. 256–260, Wiley, New York, 1937.

Frequency Distortion in Gas Phototubes. Owing to the presence of the high-mass positive ions in the current flow of a gas phototube there is appreciable frequency distortion in such tubes. This arises from the time involved in the formation of the ions and in their large transit time. A typical response curve to a light ray that is sine-wave-modulated at a variable frequency is given in Fig. 19.12. Distortion is small enough so that it is tolerable in the a-f range. It may be equalized by using an amplifier with a characteristic that rises with frequency in such a way as to offset the distortion introduced by the gas tube. A little harmonic distortion is involved in the response of a gas phototube, too, but it is generally small enough so that it is not serious. If the light source is square-wave-modulated, the current output of the gas phototube will not be a perfect square wave but will have the form of the wave

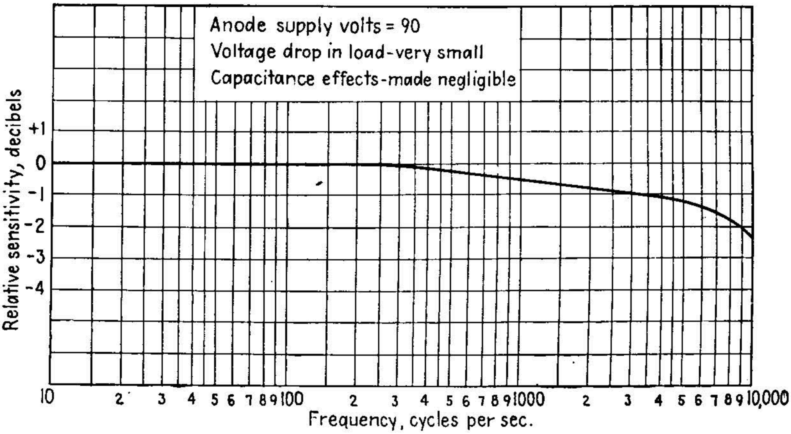


FIG. 19.12.—Response of a gas phototube to a constant illumination modulated at audio frequencies.

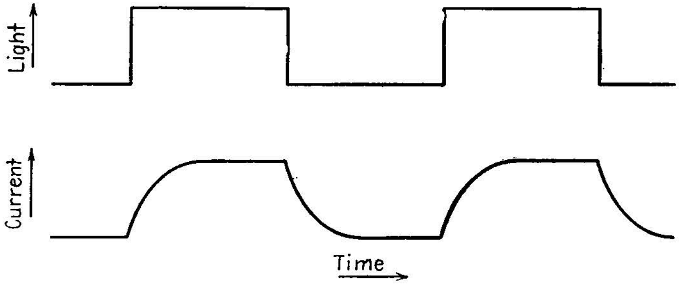


FIG. 19.13.—Response of a gas phototube to a light ray that is square-wave-modulated at a high audio frequency.

shown in Fig. 19.13. The current does not build up instantaneously to its maximum value. The principal cause of this time lag is the time required for the positive ions formed to reach the cathode. When the light source is cut off the current does not immediately drop to zero, for there are still positive ions floating about between electrodes. The time lag here is primarily due to the time required for the positive ions to diffuse to the electrodes or to disappear by combination with free electrons.

Summary of Gas-phototube Characteristics. As a result of introducing gas into a phototube, a gain in the luminous sensitivity by about a factor of 10 may be realized. A price must, however, be paid for this gain in sensitivity—the fact that the resultant tube characteristics are slightly nonlinear, introducing some harmonic distortion. Further, some frequency distortion is encountered, due to the time-lag effects in the tube.

The gas phototube must operate at much lower voltages than the vacuum phototube. This effects a considerable simplification of the power-supply circuit and is an advantage in many applications. However, there is a minimum resistance that can be used with the tube to avoid a glow discharge. The glow discharge is readily avoided by using a larger load resistance, but then the nonlinear distortion increases. Further, since the voltages at which the gas phototube are operated are of the order of one-half to one-fifth of the voltages used with the vacuum phototube and the load resistances are correspondingly lower, much of the gain in luminous sensitivity is lost. In general, gas phototubes are more suitable for low levels of illumination because of the greater luminous sensitivity, while vacuum phototubes are best suited for applications in which the amount of light or the size of the voltage supply is not a factor. Gas phototubes are further not as stable as vacuum phototubes and have a greater tendency to age rapidly and are more susceptible to injury from excessive light intensity or voltage.

19.9. Utilization of Phototube Characteristics. The output current of a phototube is so low that the phototube must always be used in conjunction with some other vacuum tube that can amplify the phototube current to a value large enough to operate a relay or other registering device. Generally this can be achieved with one stage of amplification of the voltage across the load resistor in series with the photocell and with the amplified voltage then applied to the grid of a small thyratron in whose plate circuit there is a relay. The amplifier and thyratron may be operated with either alternating voltage or direct voltage; in fact, the former arrangement has the advantage that the relay operation is generally better. If alternating voltages are used, then either a condenser must be put across the relay or a relay with a shaded pole must be used. Recently there have been developed some small screen-grid thyratrons, such as the RCA 2051, which have a sufficiently high control ratio and a low enough control-grid current so that they may be operated from either a vacuum or a gas phototube directly. In general, the electronic circuits associated with phototube control systems are quite simple and easy to build.[1–4]

19.10. Photomultiplier Tubes. Much attention has been devoted to the development of phototubes with a secondary-electron multiplier as part of the tube to increase the minute photoemission current to a larger value. [–9] Early attempts met with great difficulty in achieving stable secondary-emission surfaces that had low noise characteristics. Suitable secondary-emission surfaces were finally developed, and photomultiplier tubes are now available commercially.

The principle of the photomultiplier tube is illustrated by the partition type of tube shown in Fig. 19.14. This is a longitudinal section of a cylindrical structure, *i.e.*, the individual electrodes are noncircular cylinders generated by moving a line perpendicular to the paper. The tube contains a photocathode *PC*, from which electrons are drawn through a hole *H* in a mica shield to a first electrode 1, which has an electrostatic shield *S* attached. The photoelectrons striking the concave side of the first electrode, which is more positive than the photocathode by, say, 100 volts, give rise to secondary electrons, which are attracted to the second anode, 2, which is, say, 100 volts more positive than the

---

[1] Phototubes, *RCA Tech. Bull.* PT-20R1, pp. 4–41.

[2] HENNEY, KEITH, "Electron Tubes in Industry," 2d ed., McGraw-Hill, New York, 1937.

[3] SHEPHARD, F. H., JR., Application of Conventional Vacuum Tubes in Unconventional Circuits, *Proc. I.R.E.*, vol. 24, pp. 1573–1581, December, 1936.

[4] REICH, H. J., "Theory and Application of Electron Tubes," pp. 505–511, McGraw-Hill, New York, 1939.

[5] IAMS, H., and B. SALZBERG, The Secondary Emission Phototube, *Proc. I.R.E.*, vol. 23, pp. 55–64, January, 1935.

[6] RAJCHMAN, J. A., Le Courant résiduel dans les multiplicateurs d'électrons électrostatique, *Archives sci. phys. nat.*, [V] vol. 20, September–October and November–December, 1938. The same material is contained in Rajchman's doctor of science thesis from the Technical Institute of Zurich, 1938.

[7] ZWORYKIN, V. K., and J. A. RAJCHMAN, The Electrostatic Electron Multiplier, *Proc. I.R.E.*, vol. 27, pp. 558–566, September, 1939.

[8] RAJCHMAN, J. A., and R. L. SNYDER, An Electrically Focused Multiplier Phototube, *Electronics*, vol. 13, pp. 20–23, 58, 60, December, 1940.

[9] GLOVER, A. M., A Review of the Development of Sensitive Phototubes, *Proc. I.R.E.*, vol. 29, pp. 413–423, August, 1941.

first anode. The secondary electrons from the first anode are more numerous than the exciting photoelectrons. Likewise, the secondary electrons from the first anode on striking the second anode give rise to still more secondary electrons. Each successive anode is at a higher potential than its predecessor, and each electron striking one anode gives rise to several secondary electrons. If the secondary-emission ratio for any one electrode is r and the number of electrodes is n, then the output current is r^{n-1} times the photoelectron current. By this mechanism, current amplification of the order of 100,000 is possible. Voltages of the successive anodes are readily obtained from a voltage divider since the magnitude of the current is small.

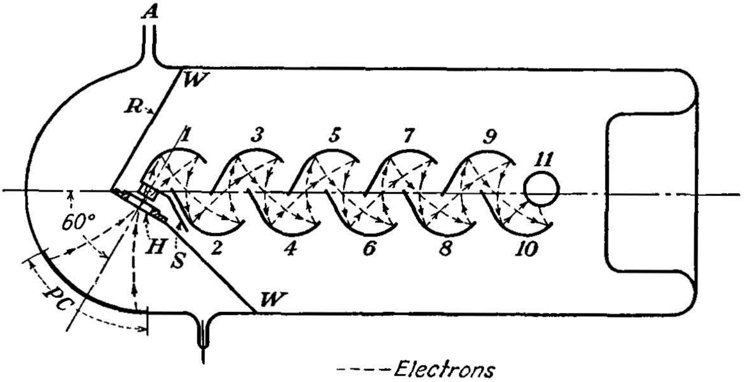


FIG. 19.14.—Structure of a partition-type photomultiplier tube.

The shape of suitable electrodes may be determined from membrane-model studies. Some typical electrode shapes and electron paths through the resultant field are shown in Fig. 19.15. For the case shown, the potential between successive electrodes is taken as 100 volts. The paths of the electrons are critical only to the extent that the action of successive electrodes produces a convergent focusing action which prevents the electrons from spilling over the edge of some later electrode. The focusing action of successive similar electrodes can be checked by plotting a curve of the striking position on an electrode as a function of the position of liberation on the previous electrode. A typical focusing curve is shown in Fig. 19.16. The liberation point is indicated by the parameter x in Fig. 19.15, while the corresponding arrival point is y (Figs. 19.15 and 19.16 are for similar but slightly different tubes). The crossover action evident in Fig. 19.15 gives rise to the peaked double-valued focusing curve of Fig. 19.16. The focusing action of the successive electrodes may be studied from the curve of Fig. 19.16 with the aid of a 45-deg construction line. An electron liberated from $x = 10$ on anode 1 will strike anode 2 at $y = 2.8$. Using the 45-deg construction line, an electron liberated from $x = 2.8$ on anode 2 will strike anode 3 at $y = 9.1$. An electron liberated from $x = 9.1$ on anode 3 will strike anode 4 at $y = 3.6$, and so on. The focusing action follows the rectangular spiral shown, with eventual convergence on point P. The electrode will have a convergent focusing action as long as the second derivative of y with respect to x of Fig. 19.16 is negative. The height of the focusing curve d is a figure of merit of the electrode shape because it determines the active portion of the multiplier electrodes. A large radius of curvature in the vicinity of the point P is desirable to prevent the electrons from bunching into the middle of an electrode too rapidly.

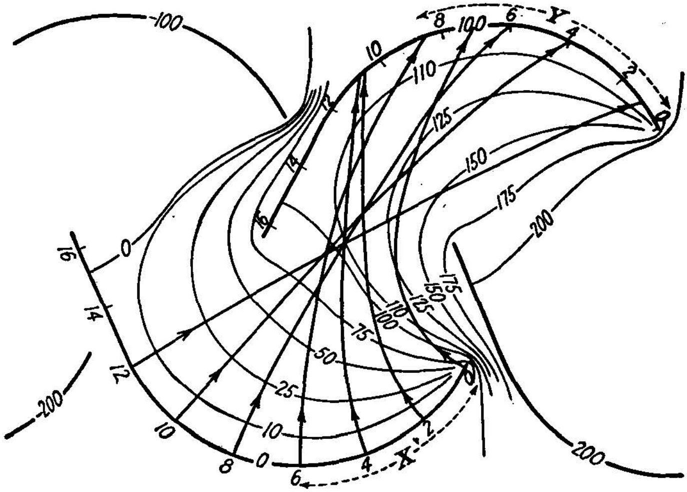


FIG. 19.15.—Electron paths in the partition-type photomultiplier tube.

In practical commercial tubes the circular structure of Fig. 19.17 is preferred because of its smaller space requirements. The action of this tube is the same in principle as that of the tube of Fig. 19.14. In the tube of Fig. 19.17 the same type of emissive surface is used for both the photocathode and the multiplier electrodes. It was the discovery of a surface with both good photoemission and good secondary-emission properties that made the commercial form of this type of tube practical. The photoemission sensitivity of this surface is about 15 microamperes per lumen. The secondary-emission multiplication ratio is about 3.5 at 100 volts per stage and about 4.0 at 125 volts per stage. For 10 multiplying anodes, or "dynodes" as they are sometimes called, this gives a total multiplication of 60,000 at 1,000 volts or 230,000 at 1,250 volts. Since a 25 per cent increase in voltage gives rise to a 200 per cent increase in current, the voltage must be regulated to 0.1 per cent if the output current is to be constant to 1.2 per cent. The luminous sensitivity of the 931 tube is 0.6 ampere per lumen for light from a tungsten filament at 2879°K and a 1,000-volt supply. The corresponding background, or "dark," current is 0.25 ampere. The dark current arises from (1) leakage resistance between electrodes, (2) secondary emission resulting from bombardment of the photocathode by positive gas ions, (3) field emission from all electrodes, and (4) thermal emission from all electrodes. Contributions to the dark current from all these sources can be reduced by careful design but can probably never be completely eliminated.

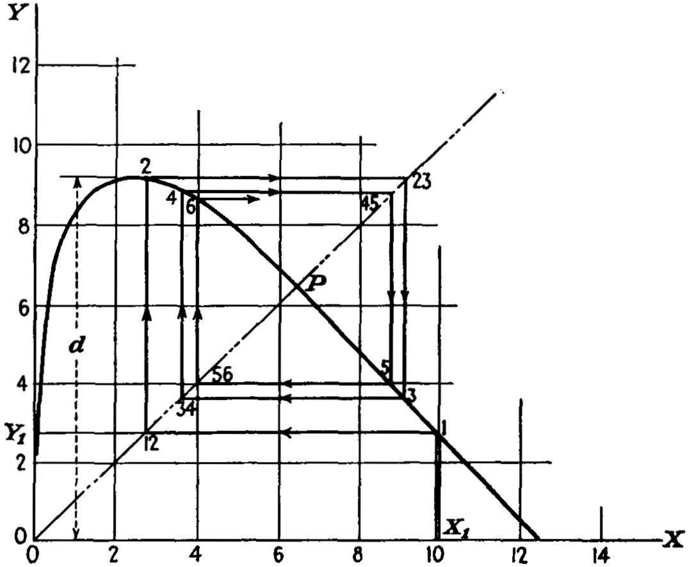


FIG. 19.16.—Focusing curve of photomultiplier electrodes.

The signal-to-noise ratio of the type 931 photomultiplier tube shown in Fig. 19.17 is superior to that of an ordinary phototube-resistor-amplifier combination. A comparison of the signal-to-noise characteristics of the 931 photomultiplier tube and a 929 vacuum phototube with amplifier is shown in Fig. 19.18. At threshold values of illumination the multiplier phototube is about 45 db superior to the vacuum-tube–resistor–amplifier combination. The signal-to-noise ratio of the photomultiplier tube increases 10 db for every factor of 10 in current, while the signal-to-noise ratio of the phototube-resistor-amplifier combination increases 20 db for every factor of 10 in current up to a point at which relatively large currents flow. At this point the signal-to-noise ratios of the two devices are about the same. The superiority of the multiplier phototube lies in the fact that noise is contributed only by shot effect, whereas in the phototube-resistor-amplifier combination there is a considerable contribution of noise from the rather large load resistor that must be used.

The nature of the signal-to-noise characteristics can be better understood from a study of the specific formulas involved. For a vacuum-phototube–resistor combination, the signal-to-noise ratio has the form

$$\frac{S_{\text{out}}}{N_{\text{out}}} = \frac{FM^2I_0^2R}{2eBI_0R + 4kTB} \tag{19.9}$$

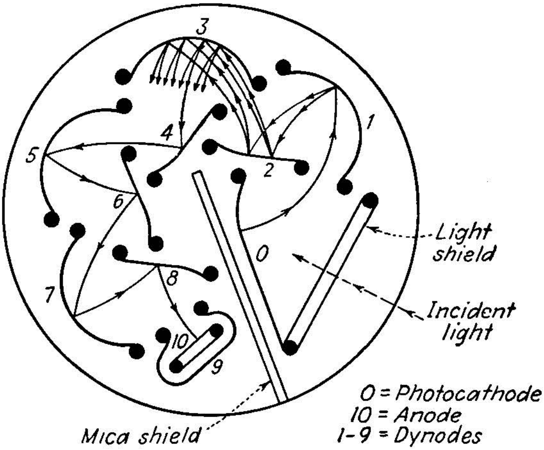


FIG. 19.17.—Structure of a circular photomultiplier tube.

where F is form factor of light modulation wave
M is percentage modulation of light wave
I_0 is direct photoelectric current, amperes
R is effective load resistance, ohms
e is charge of the electron, coulombs
B is band width, cycles
k is Boltzmann's constant, 1.372×10^{-23} watt-sec per °K
T is temperature, °K

The numerator of Eq. (19.9) represents the signal power. The denominator contains two terms, the first of which represents the shot-noise power originating in the phototube and the second of which represents the thermal-agitation noise arising from the load resistor. For low levels of illumination the first term in the denominator of Eq. (19.9) will be small, and the equation will reduce to

$$\frac{S_{\text{out}}}{N_{\text{out}}} = \frac{FM^2I_0^2R}{4kTB} \tag{19.10}$$

The signal-to-noise level for low levels of illumination is seen to vary with the square of the photoemission current (20 db for every factor of 10 in current). At high levels of illumination the second term of the denominator of Eq. (19.9) will be small compared with the first, and the equation will reduce to

$$\frac{S_{\text{out}}}{N_{\text{out}}} = \frac{FM^2I_0}{2eB} \tag{19.11}$$

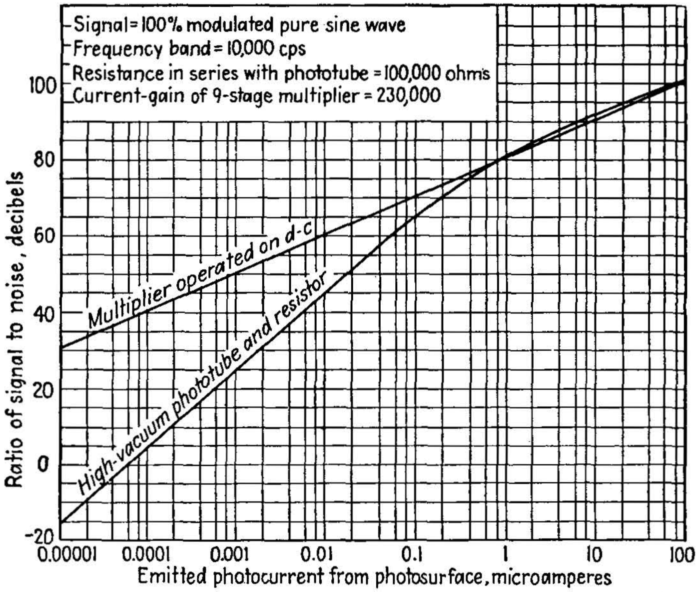


FIG. 19.18.—Comparative signal-to-noise ratios of a photomultiplier tube and a vacuum phototube with amplifier.

which is seen to be linear with photoemission current (10 db for every factor of 10 in current). The low and high level rates of variation of signal-to-noise level with photoemission current exhibited in Fig. 19.18 are thus explained.

It is of interest to examine the conditions under which either the shot noise or the thermal-agitation noise predominates in the phototube-resistor-amplifier combination. Noise contributions from the two sources will be equal when the two terms in the denominator of Eq. (19.9) are equal, *i.e.*, when

$$I_0 = \frac{2kT}{eR} \tag{19.12}$$

At room temperature, $T = 290°\text{K}$, this has the approximate value

$$I_0 = \frac{1}{20R} \tag{19.13}$$

where I_0 is in amperes and R is in ohms (or, more conveniently, I_0 may be taken in microamperes when R is in megohms). Thus the shot-noise contribution will equal the thermal-agitation-noise contribution from the resistor if the current is 1 microampere when the load resistance is 1/20 megohms. When the current is less than the value given by $\frac{1}{20R}$, the thermal-agitation noise from the resistor will be larger. When the current is greater than $\frac{1}{20R}$, the shot noise from the emitted electrons will predominate.

For the photomultiplier the signal-to-noise ratio is given by

$$\frac{S_{\text{out}}}{N_{\text{out}}} = \frac{FM^2S^{2n}I_0{}^2R}{2eB\,\dfrac{S^n(S^{n+1} - 1)}{S - 1}\,RI_0 + 4kTB} \tag{19.14}$$

where the symbols have the significance of Eq. (19.9) and S is the secondary-emission-current multiplication ratio per stage. The total power gain of the multiplier is thus S^{2n}. The numerator is evidently the same as those in the previous equations except for the factor of power gain. The first term of the denominator represents the photoemission shot noise multiplied by the noise amplification factor developed in Eq. (12.44), which includes the noise of subsequent secondary emission. The second term of the denominator of Eq. (19.14) represents the thermal-agitation noise in the frequency band B. Because of the rather considerable noise amplification, the first term in the denominator will generally be much larger than the second, and accordingly the equation reduces to

$$\frac{S_{\text{out}}}{N_{\text{out}}} = \frac{FM^2(S - 1)I_0}{2eBS} \tag{19.15}$$

This shows the signal-to-noise ratio to be linear with the photoemission current (10 db for every factor of 10 in current). The signal-to-noise ratio for this case is further seen to be the same as that of the vacuum-phototube–resistor combination as given in Eq. (19.11) except that it is smaller by the factor $\frac{S - 1}{S}$. Hence the observed behavior of Fig. 19.18, which shows the signal-to-noise ratio of the vacuum-phototube–resistor combination to be slightly less than that of the photomultiplier tube at high levels of illumination.

CHAPTER 20

SPECIAL TUBES

20.1. Introduction. It was inevitable that in the development of vacuum tubes there should arise the need for various special forms. Fortunately, tubes are now manufactured so easily that it is actually possible to get a tube tailor-made to suit almost any purpose. Attempts at standardization have held down the number of special tubes that would otherwise have come into existence. Also, the fact that tubes are quite versatile and can be used to give various operating characteristics by changing connections and the applied voltages has acted somewhat to restrict the number. During the Second World War the need for special tubes was so great that hundreds came into existence, but with time these will probably be reduced to a relatively small number.

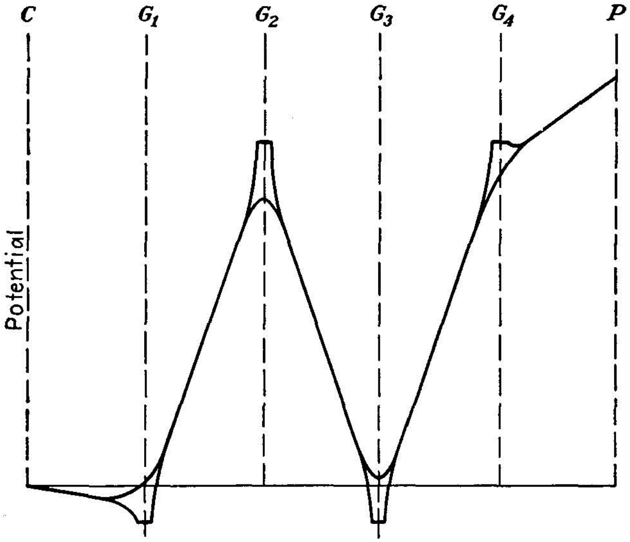


Fig. 20.1.—Potential profiles in a hexode.

In this chapter there will be discussed the principal special tubes not treated in the previous chapters. The characteristics of conventional tubes operated so as to produce special characteristics will also be discussed.

20.2. The Hexode. The hexode is a six-electrode tube. It consists of a cathode, four grids, and an anode. It is generally used as a mixer tube in superheterodyne receiver circuits. The nonlinear characteristics of the tube are used in such a way that when a r-f signal is applied to one grid and a signal from a local oscillator is applied to another grid the beat- or difference-frequency component appears in the plate current. Thus the mixer tube functions as a frequency converter.

For frequency-conversion purposes the hexode is invariably operated with the relative potentials shown in Fig. 20.1. Let the grids be numbered consecutively from the cathode to anode, and let the voltages applied to them be designated by the corresponding numerical subscripts in all the following discussion. The function and operating conditions of the various electrodes may be summarized by the following tabulation:

Electrode	Direct voltage	Function
Cathode	Zero	Source of electron current
Grid No. 1	Small negative	Injection of local oscillator voltage
Grid No. 2	Large positive	Screen grid to reduce electrostatic coupling between the signal and oscillator grids
Grid No. 3	Small negative	Injection of r-f signal voltage
Grid No. 4	Large positive	Screen grid to reduce electrostatic coupling between signal and output circuits
Plate	Larger positive	Collector of modulated electron current

The complete representation of hexode characteristics involves a large collection of characteristics, for many voltage combinations are possible. However, since the No. 2 and 4 grids and the plate are usually held at fixed potentials, the tube is well described by two characteristic curves. These are the I_p-V_1 and the I_p-V_3 characteristics. In addition, the I_p-V_p characteristics are of some interest.

The I_p-V_1 characteristics of a hexode are similar to the I_p-V_1 curves of a pentode for different values of suppressor-grid potential if the currents involved are small. The No. 1 grid under this condition will control the space current to the subsequent electrodes. This current will divide between these electrodes in a nearly constant fashion. The I_p-V_1 curves of an actual commercial hexode are shown in Fig. 20.2. These curves exhibit a maximum of plate current due to the formation of a virtual cathode in front of the No. 3 grid. When the virtual cathode forms, some of the space current will be reflected back to the No. 2 grid and the current transmitted to the plate will actually decrease as the space current increases. This action is very similar to that which occurred in the beam-power tube. As the No. 3 grid is made more negative, the virtual cathode will form at lower space current. The formation of the virtual cathode corresponds to the peak of plate current. In application, this tube must be operated to avoid the region of negative transconductance.

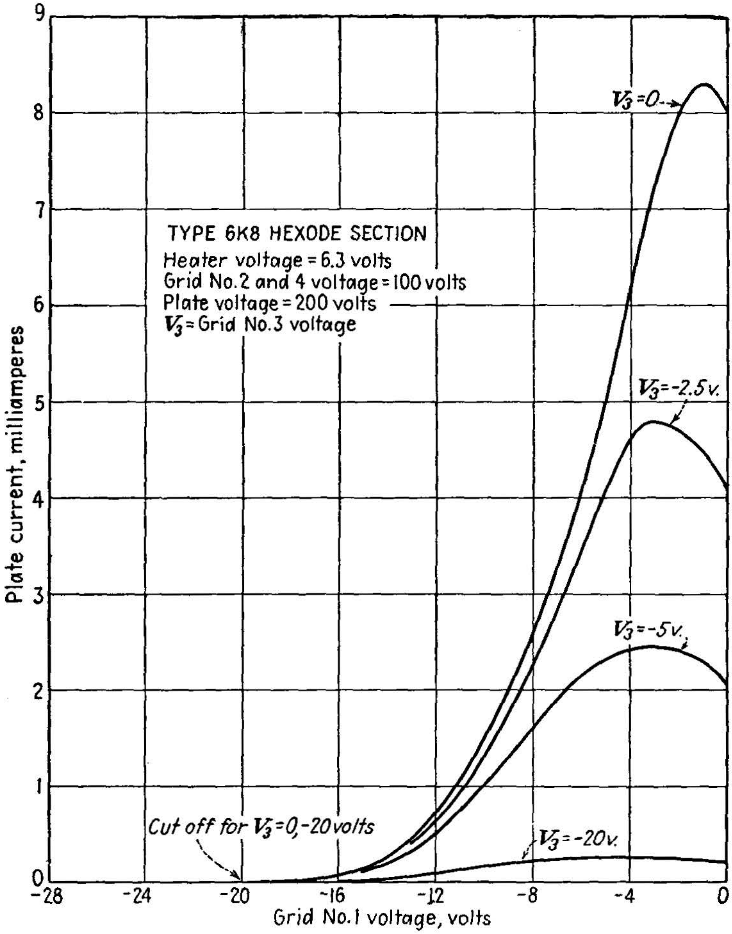


FIG. 20.2.—I_p-V_1 characteristics of a hexode.

The I_p-V_3 curves of a hexode will resemble those of a pentode if the current is small. Actual characteristics as shown in Fig. 20.3 may exhibit some crossovers due to the formation of a virtual cathode in front of the No. 3 grid.

The I_p-V_p characteristics of a hexode will resemble those of a tetrode if the current is small. There may be an interchange of secondary electrons between the No. 4 grid and plate since there is no shielding grid between them. Actual curves as shown in Fig. 20.4 will resemble beam-power-tube characteristics if the current is high enough, for then a potential minimum will form between the No. 3 grid and plate that will suppress the interchange of secondary electrons.

In addition to the static characteristics, several of the dynamic constants of the hexode are of interest. As with other multielectrode tubes, the amplification factor is of no particular significance. It is very high because of the shielding action of the screen grids. The plate resistance of the hexode is likewise of no great significance. It will tend to be high, of the order of the plate resistance in a tetrode but not as high as the plate resistance of a pentode. The control-grid transconductances of a hexode, however, are of considerable importance. The first of these transconductances is the first-grid–plate transconductance, which is defined by

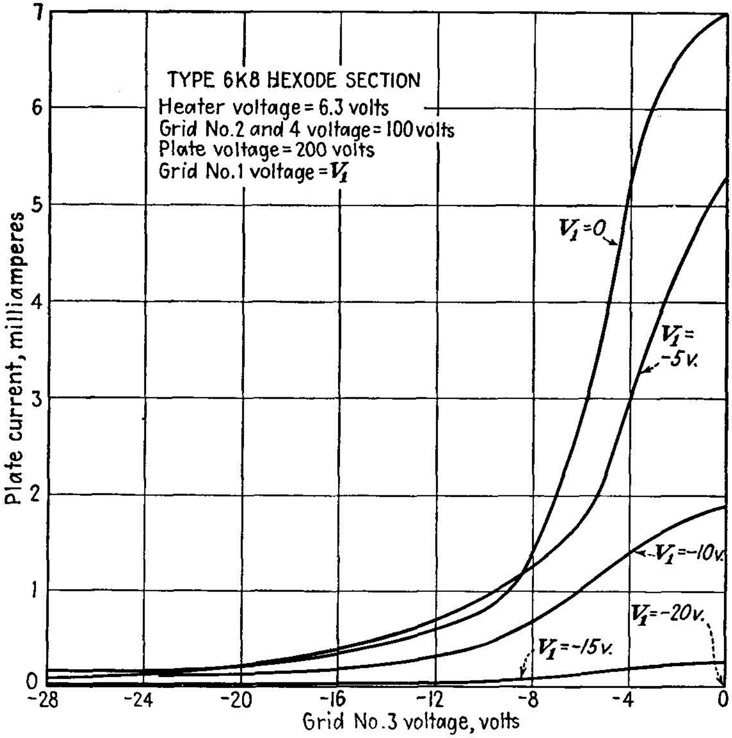


FIG. 20.3.—I_p-V_3 characteristics of a hexode.

$$g_{1p} = \frac{\partial I_p}{\partial V_1} = \left(\frac{dI_p}{dV_1}\right)_{V2, V3, V4, Vp\ =\text{const}} \tag{20.1}$$

This transconductance is equal to the slope of the characteristics shown in Fig. 20.2. The other transconductance of interest is the third-grid-plate transconductance. It is defined by

$$g_{3p} = \frac{\partial I_p}{\partial V_3} = \left(\frac{dI_p}{dV_3}\right)_{V1,V2,V4,Vp = \text{const}} \tag{20.2}$$

This transconductance is equal to the slope of the characteristics shown in Fig. 20.3. The transconductance g_{1p} will generally be greater than the transconductance g_{3p}.

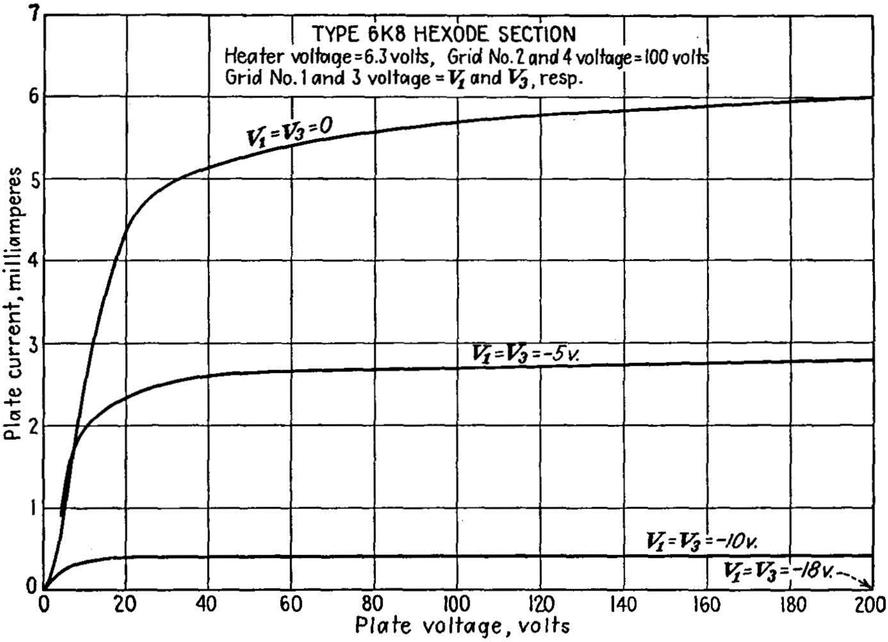


FIG. 20.4.—I_p-V_p characteristics of a hexode.

Another hexode constant that is of particular significance is the so-called "conversion transconductance." This is the ratio of the magnitude of the plate current of frequency $f_1 - f_2$ to the magnitude of voltage of frequency f_1 applied to one of the control grids (No. 1 or 3 in the case of the hexode) under the condition that a fixed voltage of frequency f_2 is applied to the other control grid and that all the direct electrode voltages are kept constant. Thus, in contrast with other tube conductances, the conversion transconductance is the ratio of an alternating component of plate current at the difference frequency $(f_1 - f_2)$ to the alternating component of signal voltage on one control electrode at a different frequency f_1 under the condition that a local oscillator voltage at a still different frequency f_2 be applied and maintained constant on another control electrode. The conversion transconductance is a function of the magnitude of the local oscillator voltage and passes through a maximum at a particular value of local oscillator voltage. In Fig. 20.5 are given some typical curves of conversion transconductance

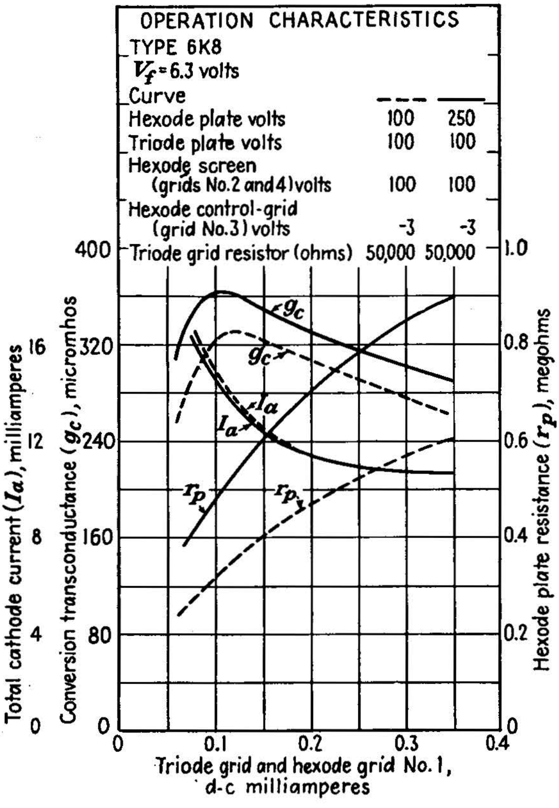


FIG. 20.5.—Conversion transconductance of a mixer tube.

of a hexode in terms of local oscillator grid current, which for a given bias resistor is proportional to local oscillator voltage.

Frequency conversion in any of the mixer type of tubes can be considered as a process of modulation of the oscillator frequency by the signal frequency, the intermediate frequency appearing as one of the sidebands. The modulation is accomplished through the medium of the electron stream in the tube. The electron stream will ordinarily experience a large amplitude variation at the oscillator frequency. The component of tube current at oscillator frequency is modulated in magnitude at the signal frequency. The degree of modulation is ordinarily so low that higher-order effects can be neglected and the basic relations studied by simple analysis.

As with an ordinary amplitude-modulated wave of current the amplitude of the side bands is smaller than the carrier by half the degree of modulation. Hence

$$I_{if} = \frac{m}{2} I_{osc} \tag{20.3}$$

where I_{if} is the component of plate current at the intermediate frequency, which is the difference between the signal and oscillator frequency; m is the degree of modulation, or the ratio of the change in the component of current at oscillator frequency to the magnitude of this component; and I_{osc} is the component of plate current at oscillator frequency. All values of current are peak rather than effective. The degree of modulation is given by

$$m = \frac{\Delta I_{osc}}{I_{osc}} = \frac{\partial I_{osc}}{\partial V_{sig}} V_{sig} \frac{1}{I_{osc}} \tag{20.4}$$

where V_{sig} is the peak value of the signal voltage. Hence, combining Eqs. (20.3) and (20.4),

$$I_{if} = \frac{1}{2} \frac{\partial I_{osc}}{\partial V_{sig}} V_{sig} \tag{20.5}$$

Since the conversion transconductance is defined as

$$g_c = \frac{I_{if}}{V_{sig}} \tag{20.6}$$

then, from Eq. (20.5),

$$g_c = \frac{1}{2} \frac{\partial I_{osc}}{\partial V_{sig}} \tag{20.7}$$

The component of plate current at oscillator frequency is given from the well-known Fourier integral

$$I_{osc} = \frac{1}{\pi} \int_{-\pi}^{\pi} I_p \cos \omega t \, d(\omega t) \tag{20.8}$$

where I_p is the instantaneous value of plate current as a function of time and ω is the oscillator angular frequency. Taking the derivative of this expression with respect to V_{sig} and substituting into Eq. (20.7),

$$g_c = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_{3p} \cos \omega t \, d(\omega t) \tag{20.9}$$

results, as may be seen by recalling the definition of g_{3p} given in Eq. (20.2). The above assumes that the oscillator voltage is applied to the

No. 1 grid of the hexode and that the signal voltage is applied to the No. 3 grid.

The third-grid–plate transconductance will vary over a large range of values as the oscillator voltage swings over its range of voltages applied to the No. 1 grid. The nature of the variation of the third-grid–plate transconductance with No. 1 grid voltage is shown in Fig. 20.6, curve *b*. The conversion transconductance can be evaluated graphically or numerically from this curve and Eq. (20.9). Also shown in Fig. 20.6 is the effect of the sinusoidal variation of oscillator voltage upon the plate current, which is proportional to the third-grid–plate transconductance for a fixed small signal voltage. The conversion transconductance is by the definition of Eq. (20.9) equal to half the fundamental component of the resultant curve of g_{3p} as a function of time shown in Fig. 20.6. From observation of Eq. (20.9) it is seen that the oscillator voltage should be adjusted so that the tube current is cut off in the interval that cos ωt is negative. Otherwise, there is a negative area under the curve of the integrand that reduces the conversion transconductance. The *maximum possible conversion transconductance* would be obtained if the curve of third-grid–plate transconductance rose sharply from zero to a maximum value as shown for curve *a* of Fig. 20.6. Such a transconductance would yield the square wave of plate current shown and a conversion transconductance of value

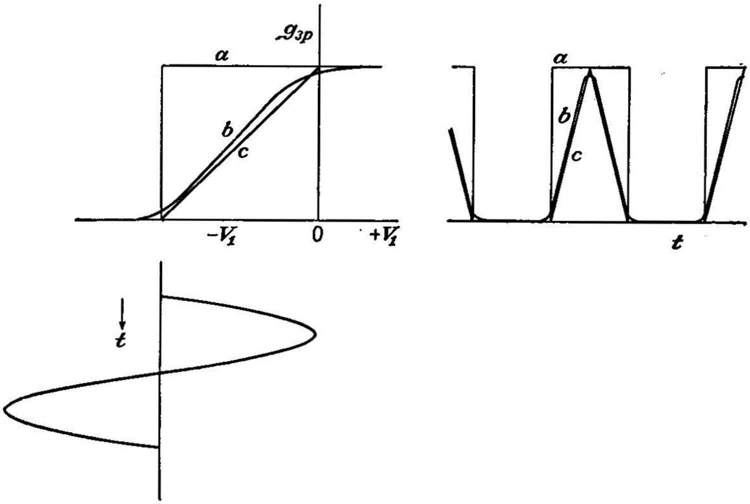


FIG. 20.6.—Variation of transconductance of a hexode with control-grid voltage.

$$g_c = \frac{g_{3p\,\max}}{\pi} \tag{20.10}$$

In other words, the maximum possible value of the conversion transconductance is about one-third the maximum value of the third-grid–plate transconductance. Actually, the curve of g_{3p} against V_1 has the form shown at *b*, which is an s-shaped curve that is almost straight. This s-shaped curve is closely approximated by the straight-line curve shown as *c*, which yields the triangular wave of plate current given. The Fourier integral of Eq. (20.9) for this case yields the value

$$g_c = \frac{g_{3p\,\max}}{4} \tag{20.11}$$

The actual value for the curve *b* will lie somewhere between the values given by Eqs. (20.10) and (20.11) but much closer to the latter. *Hence, in general it may be expected that the conversion transconductance will have a value approximately equal to one-fourth the maximum value of signal-grid–plate transconductance of any mixer-type tube.*

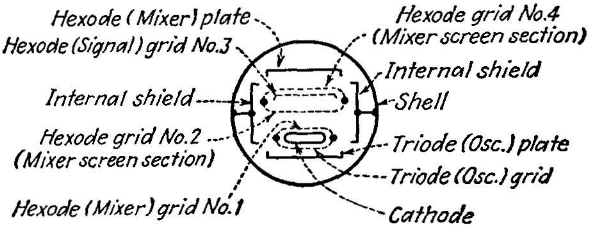


FIG. 20.7.—Electrode structure of the 6K8 triode-hexode.

The form of Eq. (20.9) indicates that there is generally an optimum value of oscillator voltage. If the oscillator voltage is too large, the plate current will not flow for a sufficiently large fraction of a cycle. If it is too small, the current will flow for more than a half cycle and the conversion transconductance will be reduced.

An example of a commercially available hexode is the 6K8, which contains a triode and a hexode in the same envelope. This tube is specifically designed to be operated as a mixer tube in a superheterodyne receiver. The tube is built so that the No. 1 grid is common to the hexode and the triode. The No. 2 and 4 grids are tied together internally. The triode is built on one side of a strip cathode, and the hexode is built on the other. A cross section of the tube electrode structure is shown in Fig. 20.7. As a result of this structure, the local oscillator voltage appears on the No. 1 grid of the hexode, and the r-f signal must be applied to the No. 3 grid.

The hexode was one of the first mixer-type tubes developed. More recently developed types exhibit better operating characteristics. Hexodes in general suffer from some interaction between the two input circuits, a relatively low conversion transconductance, and a relatively low plate resistance.

20.3. The Heptode. The heptode is a seven-electrode mixer tube with five grids. It has the construction shown in Fig. 20.8. The potential variation within the tube is shown in Fig. 20.9. The function and operating conditions of the various electrodes may be summarized in the following tabulation:

Electrode	Direct voltage	Function
Cathode	Zero	Source of electron current
Grid No. 1	Small negative	Injection of signal voltage, source of bias for automatic volume control
Grid No. 2	Large positive	Screen grid to reduce electrostatic coupling between signal and oscillator grids
Grid No. 3	Small negative	Injection of local oscillator voltage
Grid No. 4	Large positive	Screen grid to reduce electrostatic coupling between oscillator and output circuits
Grid No. 5	Zero	Suppressor grid to improve plate-current characteristics and further reduce electrostatic coupling between oscillator and output circuits
Plate	Large positive	Collector of modulated electron current

Since the No. 2, 4, and 5 grids of the heptode are generally held at fixed voltages, the static characteristics of the heptode may be represented by the I_p-V_1, I_p-V_3, and I_p-V_p characteristics.

The I_p-V_1 characteristics of a 6L7, which is a typical heptode, are shown in Fig. 20.10. These are similar to the I_p-V_1 curves of a variable-mu pentode for various suppressor-grid voltages. The No. 1 grid has the principal influence in determining the magnitude of the space current that is passed on to the subsequent electrodes.

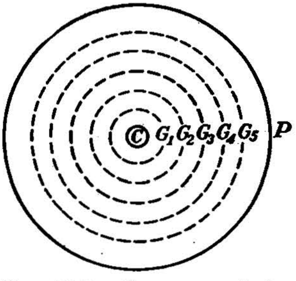


FIG. 20.8.—Structure of the heptode.

The I_p-V_3 characteristics of a 6L7 heptode are shown in Fig. 20.11. These curves are similar to the plate-current–suppressor-grid-voltage characteristics of an ordinary pentode. The potential of the No. 3 grid determines the fraction of the space current that is passed on to the plate.

The I_p-V_p characteristics of a 6L7 heptode are shown in Fig. 20.12. These curves are similar to the I_p-V_p characteristics of an ordinary pentode. They resemble pentode rather than screen-grid-tube characteristics because of the presence of a suppressor grid between the last screen grid and the plate.

Heptode characteristics are in general superior to hexode characteristics.[1] In the first place it is possible to use the No. 3 instead of the No. 1 grid for local-oscillator-voltage injection because of the extra shielding between the No. 3 grid and plate introduced by the presence of

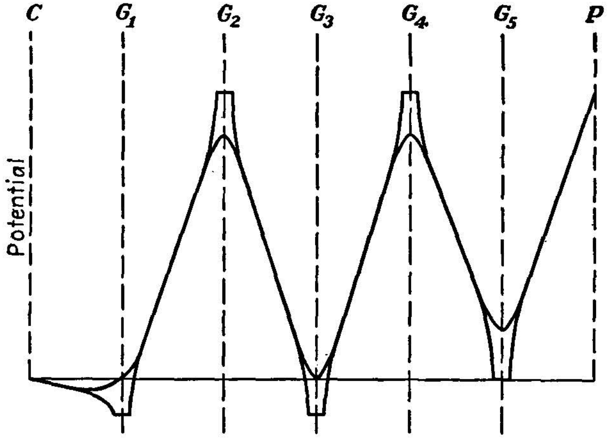


FIG. 20.9.—Potential profiles of the heptode.

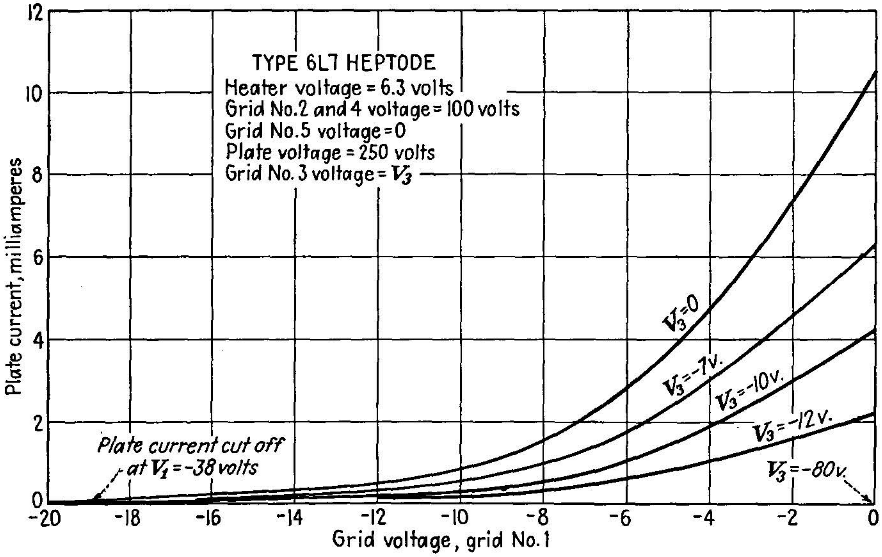


FIG. 20.10.—I_p-V_1 characteristics of the heptode.

---

[1] NESSLAGE, C. F., E. W. HEROLD, and W. A. HARRIS, A New Tube for Use in Superheterodyne Frequency Conversion Systems, *Proc. I.R.E.*, vol. 24, pp. 207–218, February, 1936.

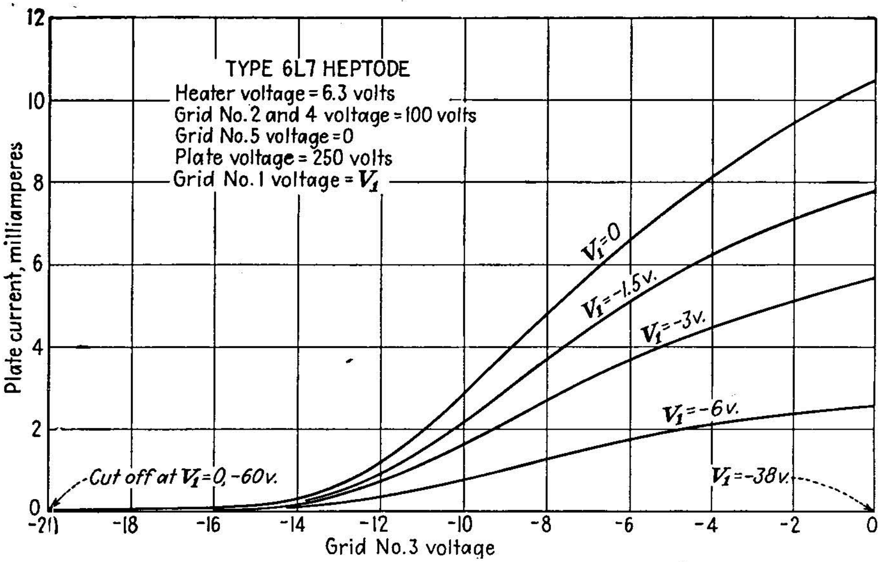


FIG. 20.11.—I_p-V_3 characteristics of the heptode.

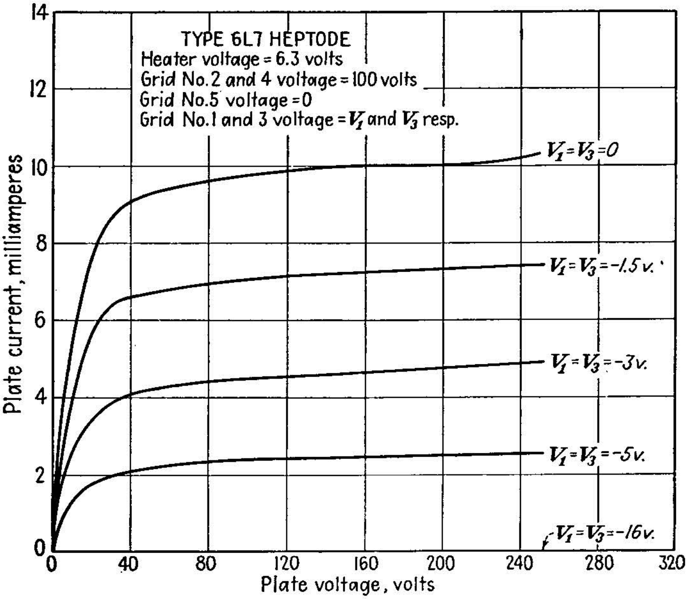


FIG. 20.12.—I_p-V_p characteristics of the heptode.

the No. 5 suppressor grid. This arrangement allows the No. 1 grid to be used for signal injection and makes it possible to obtain a variable-mu action from this grid, which in turn makes good automatic volume control possible. In general, it is very difficult to design a tube with a variable-mu characteristic on any but the first grid. This is because the subsequent control grids are necessarily coarse in order to pass a sufficient fraction of the space current, and being coarse do not allow a large enough range of amplification factor. The addition of the No. 5 suppressor grid also raises the plate resistance and thus allows higher screen-grid voltages, which in turn increases the No. 1 grid-plate transconductance and so raises the obtainable conversion transconductance. The increase in plate resistance improves the selectivity and gain of the tube.

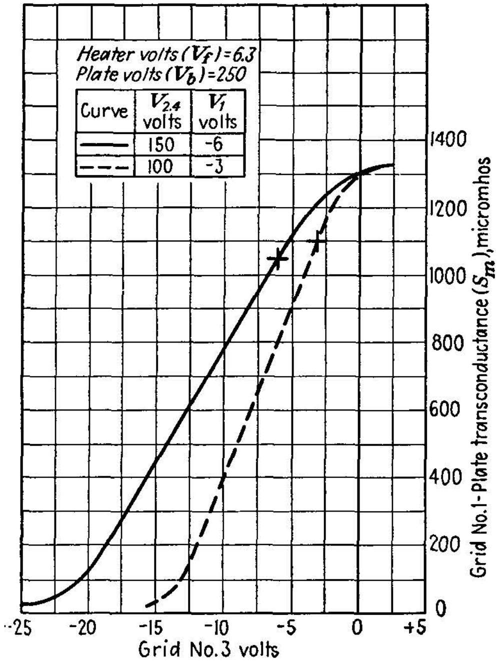


FIG. 20.13.—g_{1p}-V_3 characteristics of the heptode.

Aside from the above factors the mixer operation of the heptode is like that of the hexode. In Fig. 20.13 are shown the g_{1p}-V_3 characteristics of the 6L7 heptode. These transconductance curves are similar to the limiting curve c of Fig. 20.6. The resultant g_c-V_1 characteristics are shown in Fig. 20.14. Excellent automatic-volume-control characteristics are exhibited here. It should also be noted that the maximum conversion transconductance obtained is approximately one-fourth of the maximum No. 1 grid-plate transconductance.

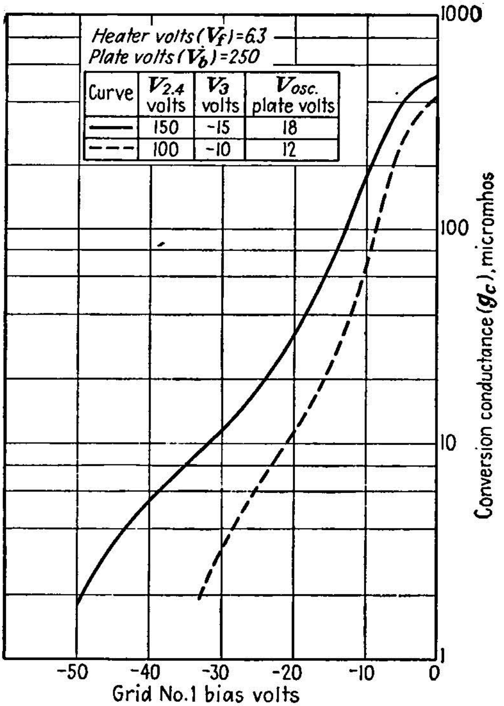


FIG. 20.14.—g_c-V_1 characteristics of the heptode.

20.4. The Pentagrid Converter. The pentagrid converter is a heptode as far as its static characteristics are concerned but is used as a mixer by connecting the cathode and first two grids as a triode oscillator. With this arrangement, the No. 2 grid acts as the triode plate, and the cathode cannot be operated at zero potential but must be allowed to have oscillator voltage on it. Furthermore, the local oscillator voltage is effectively introduced on the No. 1 grid, and the signal is introduced on the No. 3 grid. As with the heptode, the No. 4 grid is a screen grid, and the No. 5 grid functions as a suppressor. This arrangement has the advantage that it requires one less tube but has the disadvantage that bias for the automatic-volume-control action is more difficult to apply.

Typical potential profiles for a pentagrid-converter connection of a heptode are shown in Figs. 20.15 and 20.9. Two types of operation are possible. In Fig. 20.15 the No. 3 and 5 grids are operated as screen grids. This arrangement has the advantage that the reaction between the signal and oscillator circuits is reduced because the No. 2 and 3 grids exert a shielding action but has the disadvantage that the plate resistance is relatively low, with attendant loss of gain and selectivity. In the arrangement of Fig. 20.9 the No. 2 grid acts as triode plate and screen grid, the No. 3 grid is signal-injection grid, the No. 4 grid is a screen grid, and the No. 5 grid is a suppressor grid. This arrangement has better plate-resistance characteristics than the previous one but shows more interaction between the signal and oscillator circuits unless specially designed tubes are used.

In addition to the electrostatic coupling between the signal and oscillator circuits in mixer tubes there may be an electronic interaction. This occurs because with moderately large signal voltages the No. 3 signal grid may become negative enough on the negative half of the signal-voltage cycle to repel low-velocity electrons approaching it from the oscillator section of the tube. These electrons will be thrown back into the oscillator section and constitute an electronic loading that may change the local oscillator frequency appreciably.[1,2]

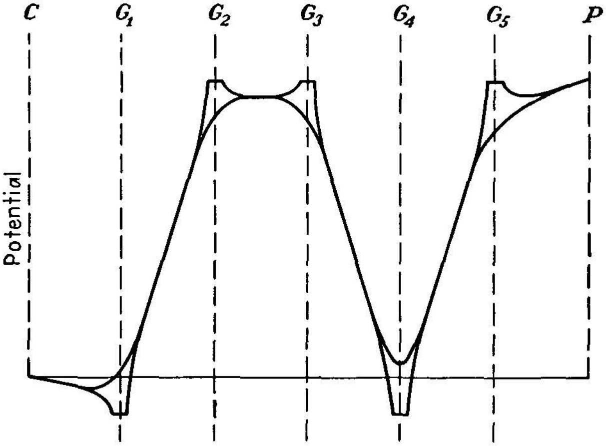


FIG. 20.15.—Potential profiles of a pentagrid converter.

The electronic interaction between the signal and oscillator circuits may be reduced by using a heptode with the special electrode structure shown in Fig. 20.16. This structure, typified in the 6SA7, differs from that shown in Fig. 20.8 by the addition of some curved collector plates which partly enclose the No. 2 grid and are connected to it. In addition, the No. 3 signal grid has a supporting wire opposite the opening in the

---

[1] Strutt, M. J. O., "Moderne Mehrgitter-Elektronenröhren," Vol. II, pp. 94–102, 112–114, Springer, Berlin, 1938.

[2] RCA Manufacturing Co., Operation of the 6SA7, *Application Note*, 100, 1938.

collector plates. This causes the potential opposite the collector-plate opening to be more negative than on either side of the supporting rod, and as a result electrons are deflected to one side. By virtue of this special electrode structure, electrons that are repelled from the No. 3 signal grid are deflected so that they are caught by the collector plates and prevented from returning into the active electron stream of the oscillator section. The collector plates further increase the electrostatic shielding between the signal and oscillator circuits, with attendant improvement of operation. The resulting operating characteristics are appreciably superior to those of the ordinary pentagrid tube.

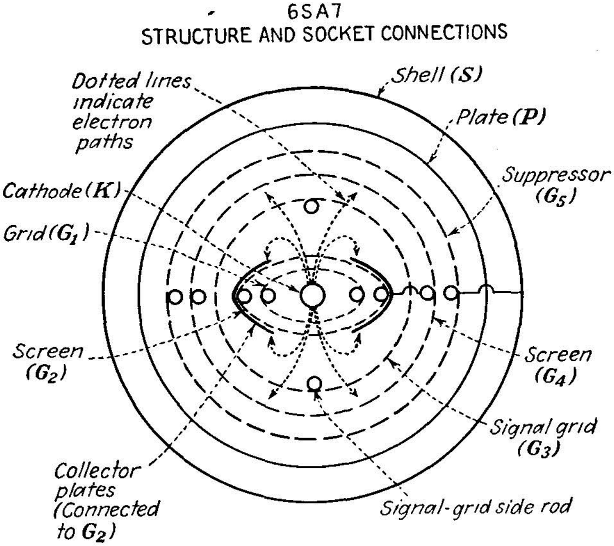


FIG. 20.16.—Structure of the 6SA7 special pentagrid converter.

20.5. The Octode. The octode is an eight-electrode mixer tube with six grids. As ordinarily used, the cathode and first two grids are connected as a triode oscillator. The No. 3 and 5 grids act as screen grids. Signal is injected into the No. 4 grid. The No. 6 grid is used as a suppressor grid. This arrangement achieves the desired effects of low electrostatic coupling between the signal and oscillator circuits while at the same time giving good plate-circuit characteristics. The 7A8 is an example of a typical octode. In addition, the electrodes may have the special structure described in the previous section, with the difference that the collector plates form the No. 3 electrode and are not connected to the No. 2 grid. The No. 3 grid is operated as a screen grid, and inasmuch as it has a separate connection the repelled electrons that are caught by it do not flow through the triode oscillator or No. 2 grid circuit. The Philips EK3 is an example of such a tube.[1]

20.6. Space-charge-grid Tubes. In all multielectrode tubes having signal grids operated at small negative potentials there is the possibility of the formation of a virtual cathode before the control grid if the space current is high enough. This virtual cathode acts like an ordinary cathode and has the advantage that it has a large area and so may give rise to a relatively high transconductance. An ordinary screen-grid tetrode can be operated as a space-charge-grid triode by connecting the No. 1 grid at a positive potential and using the No. 2 grid as a control grid. This gives rise to the potential distribution shown in Fig. 20.17. The No. 1 grid in this case is called the "space-charge grid" because it draws a high enough current from the cathode to form the virtual cathode in front of the No. 2 grid. Typical current-voltage characteristics are shown in Fig. 20.18. The I_p-V_2 characteristics are similar to those of a triode except that they exhibit greater curvature and hence more distortion in amplifier applications.

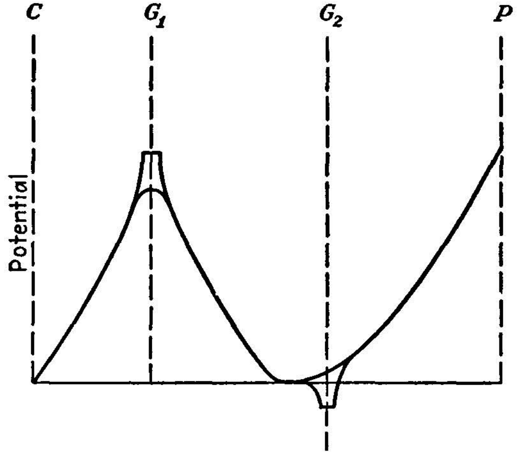


FIG. 20.17.—Potential profiles in the space-charge-grid triode.

The space-charge-grid principle may be applied to pentodes and other multielectrode tubes as well as to tetrodes. The space-charge principle finds its chief application where tube operation is restricted to low voltages, as with certain types of battery-operated circuits. When the voltages available are of the order of 50 volts or less, appreciably higher effective transconductances can be obtained than can be had with

---

[1] See STRUTT, *op. cit.*

conventional tube connections. For voltages above 50 volts there is no gain, and the space-charge-grid principle finds little application.

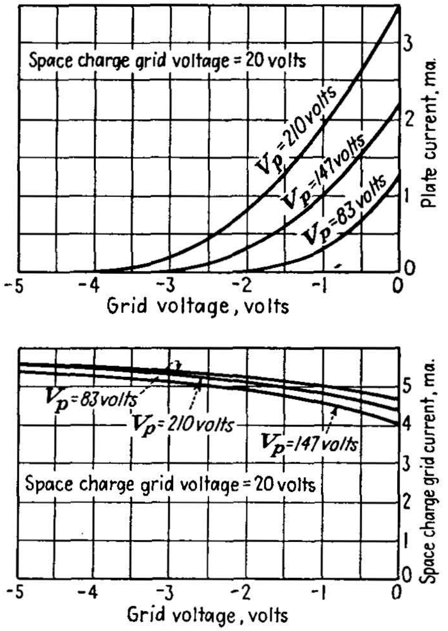


FIG. 20.18.—Current-voltage characteristics of the space-charge-grid triode.

20.7. Negative-resistance Tubes. Many applications in electron-tube circuits require the existence of a negative resistance. By a negative resistance is meant a circuit which is such that an increase in current through it produces a decrease rather than an increase in voltage across it. Any device having a current-voltage characteristic that exhibits a negative slope has a negative resistance in the region of the negative slope. In general, circuits will have a negative resistance over only a part of their operating characteristic. Just as a positive resistance consumes power, so, correspondingly, a negative-resistance element delivers power. This means that negative-resistance devices must always have a source of power associated with them.

Means of obtaining negative-resistance characteristics from tubes are almost too numerous to mention.[1] A few of the devices are fundamental and important enough to deserve mention.

Glow-discharge Tubes. In an arc or glow-discharge tube an increase in current produces an increase in ionization so that a smaller voltage is required to maintain the discharge. Almost every two-element glow tube exhibits a negative resistance somewhere in its operating characteristic. The usefulness of this type of negative resistance is limited by the fact that the magnitude of the resistance changes with temperature and life of the tube. Also, the time lag associated with the positive ions present limits the frequencies at which the negative-resistance characteristic is available to low audio values..

The Dynatron. The negative resistance that is available over part of the plate-current–plate-voltage characteristic of an ordinary screen-

---

[1] HEROLD, E. W., Negative Resistance and Devices for Obtaining It, *Proc. I.R.E.*, vol. 23, pp. 1201–1223, October, 1935. Contains extensive bibliography.

grid tube as shown in Fig. 10.2 is known as a "dynatron characteristic." The negative-resistance characteristic results from the transfer of secondary electrons from plate to screen grid. When the screen grid is more positive than the plate, an increase in plate voltage will attract more primary electrons to the plate but relatively more secondary electrons are lost to the screen grid so that the net plate current decreases rather than increases. If a parallel resonant circuit is placed in the plate circuit of a screen-grid tube and the tube operated at voltages that will give a negative-resistance characteristic, oscillations will occur in the plate circuit provided that the magnitude of the resistance of the parallel resonant circuit is greater than the magnitude of the negative resistance of the tube plate circuit. Oscillations will in general build up to the point where the magnitude of the negative resistance as averaged over the cycle of oscillation equals the positive resistance of the parallel resonant circuit.

The negative-resistance characteristic obtainable from a screen-grid tube is subject to change as the tube ages and as the secondary-emission characteristics of the plate change from any of a number of causes. For this reason this type of negative resistance is not extensively used.

Direct-coupled Negative-resistance Devices. The most interesting and stable types of negative resistances are those which are obtained from judicious interconnections of standard vacuum tubes. Such devices are dependent not upon gas or secondary-emission characteristics but rather upon current division between electrodes, space-charge effects, or a feedback action, all of which are quite stable and are capable of giving lower magnitudes of negative resistance and a wider range of operating voltages than are available by other methods.

Negative Screen-grid Resistance of a Pentode. The screen grid of an ordinary pentode exhibits a negative-resistance characteristic if it is connected to the suppressor grid in such a way that an increase in screen-grid voltage is accompanied by an equal increase in suppressor-grid voltage. This is evident from the curves of Fig. 20.19. This family of curves shows the I_2-V_2 characteristics of a pentode for various values of V_3, where the numerical subscripts refer to the grid number in order from the cathode to plate. The solid curves show the I_2-V_2 characteristics. As the No. 3 (suppressor) grid is made more negative, the No. 2 (screen) grid current decreases. If the No. 2 and 3 grids are connected so that there is a constant difference of potential between them, the dotted curves shown in Fig. 20.19 result. The screen current decreases as the suppressor grid is made more positive; for the latter then transmits a greater fraction of the space current that approaches it, and as a result less current is returned to the screen grid. This decrease in reflected current more than offsets the increase in directly intercepted space current that is taken on by the screen grid as a result of its more positive potential. The two dotted curves of Fig. 20.19 are for differences of No. 2 and No. 3 potential of 54 and 90 volts, respectively. The magnitude of the negative resistance made available by this means is of the order of 3,500 ohms, which is considerably less than that obtainable from a dynatron, which is usually of the order of 10,000 ohms. The region of negative resistance is limited at low voltages by the condition that the suppressor grid is returning all the electrons which approach it and beyond this condition the suppressor grid has virtually no influence. Correspondingly, the region of negative resistance is limited at high voltages by the condition that the suppressor grid is passing all the current which approaches it and so again loses control.

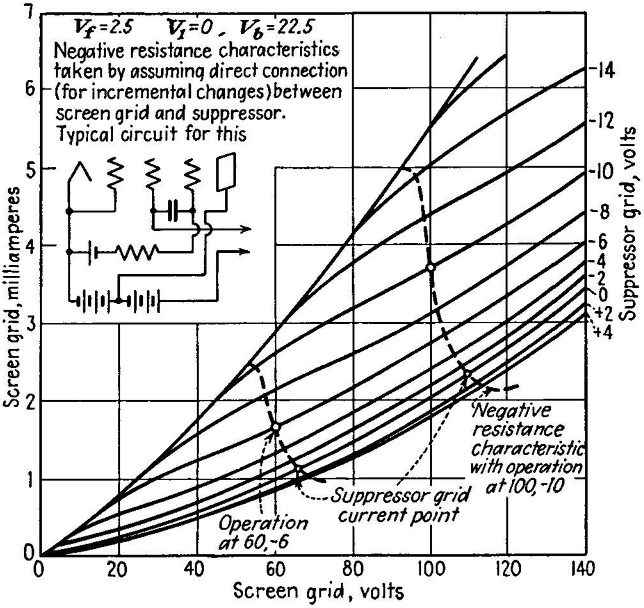


FIG. 20.19.—I_2-V_2 characteristics of a pentode.

In actual applications the screen and suppressor grids are separately biased and fed through separate resistors but are coupled by a large capacity connected directly across the tube leads. This means that the No. 2 and 3 grids are connected together as far as voltage variations are concerned over a large band of frequencies. The negative-resistance characteristic is available from low audio frequencies, dependent upon the size of the coupling condenser compared with the size of the resistors in series with the electrodes, to frequencies of the order of 60 mc, at which transit-time effects disturb the relations.

With proper connections the negative screen resistance of a pentode can be made to furnish either sinusoidal or square waves.[1,2] Likewise, trigger and flip-flop characteristics can be made available.

Push-pull Negative-resistance Circuit. It is possible to connect two triodes or two pentodes in a push-pull arrangement by which a very good negative-resistance characteristic is made available. The circuit and resultant characteristics are shown in Fig. 20.20. The current which flows through the input terminals shown consists of two components,

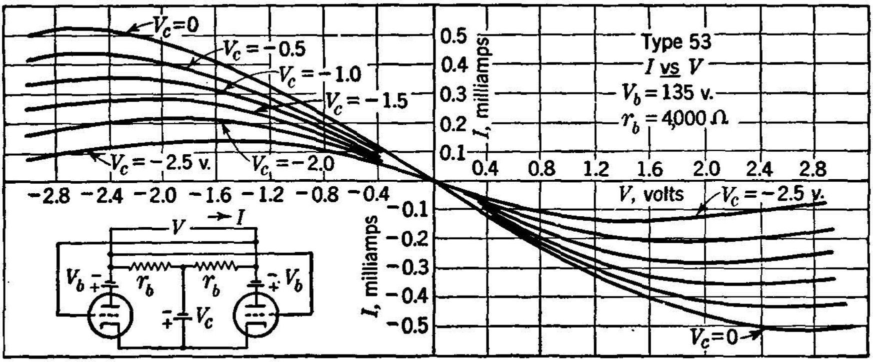


FIG. 20.20.—Current-voltage characteristics of the push-pull circuit.

that produced by the applied voltage, which is in one direction, and that produced by the tubes, which will be in the opposite direction because of the cross connection of the grids. The latter component of current can be made much larger than the former by tapping sufficiently high on the plate resistor of the other tube, usually across the entire resistor. The negative-resistance characteristic shown results. The resistance available has the approximate value of

$$R = \frac{2r_p}{\dfrac{r_b + r_p}{r_b} - \mu k} \tag{20.12}$$

for small values of impressed voltage, where R is the effective value of the resistance at the input terminals, r_p is the dynamic plate resistance of the tubes, r_b is the value of the plate resistor, μ is the amplification

---

[1] REICH, H. J., "Theory and Application of Electron Tubes," 2d ed., Chap. X, McGraw-Hill, New York, 1944.

[2] BRUNETTI, C., The Transitron Oscillator, *Proc. I.R.E.*, vol. 27, pp. 88–94, February. 1939.

factor of the tubes, and k is the fraction of the voltage developed across the plate resistors that is applied to the other tube. If k is sufficiently large, the effective resistance will usually be negative.[1]

In application, the cross connection between grids is made through a large capacity. By proper connection either sinusoidal or square waves may be obtained. The negative-resistance characteristic is available from low audio frequencies to frequencies of the order of megacycles. This is the basic circuit of the Eccles-Jordan trigger circuit and multivibrator.

Feedback Circuits. The push-pull tube connection just described may be thought of as a feedback circuit composed of a two-stage direct-coupled amplifier that has its output connected to its input. More complicated arrangements can be used as well to give negative reactances as well as resistances.[2] In principle, these methods use a feedback connection so that an increase in voltage between two terminals causes a current to flow in the opposite direction and thus gives rise to a negative-impedance characteristic.

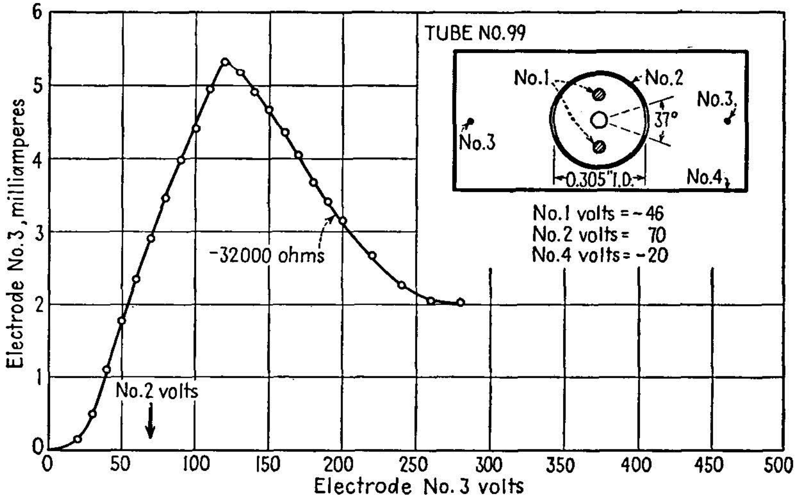


FIG. 20.21.—Structure and characteristics of a special negative-resistance tube.

Special Negative-resistance Tubes. It is possible to design special tubes so that they will have particularly good negative-resistance characteristics. Such tubes will in general make use of rather well-known electronic action. Already mentioned has been a negative resistance that

---

[1] See REICH, *op. cit.*

[2] GINZTON, E. L., Stabilized Negative Impedances, *Electronics*, vol. 18, pp. 140–150, 138–148, 140–144, July, August, September, 1945.

depends upon reflection of electrons from a grid. Such action can be enhanced by making the electrons approach the grid at a small angle. In addition, it is possible to get a negative resistance by deflection of an electron beam. Shown in Fig. 20.21 is a special electron tube that obtains its negative-resistance characteristic from the change of focal length of an electron beam.[1] Maximum current will be intercepted by the No. 4 wire electrode when the voltage on it is just right to focus on it the electron beam produced by the other electrodes. For higher or lower voltages the current will drop off, thus giving rise to a negative-resistance characteristic for voltages higher than that required for a focus on the wire.

20.8. Negative-transconductance Tubes. Two instances in which negative transconductances appear in tubes have already been mentioned. One case is that of the beam-power tube in which as shown in Fig. 10.13 the curve of plate current versus current injected into the screen-grid–plate region exhibits a negative slope over a portion of its characteristics. Since the space current in a beam tube increases as control-grid voltage increases, it will be true that over an appreciable portion of the available characteristics the plate current can be made to decrease as the control-grid voltage increases provided only that the space current is high enough so that a virtual cathode forms between the screen grid and plate. This gives rise to a negative control-grid–plate transconductance but with an ordinary 6L6 requires that the tube be operated at its maximum dissipation.

A negative suppressor–screen-grid transconductance also exists in the ordinary pentode, as is evident from the characteristics of Fig. 20.19. As suppressor-grid voltage is raised from a negative value, the screen-grid current decreases; for a greater fraction of the space current is passed on to the plate, and hence a smaller current is reflected back to the screen grid. The negative transconductance here is available only in the range of suppressor-grid voltages between which the suppressor grid reflects all current and passes all current.

20.9. Electron-ray Indicator Tubes. The electron-ray tube is an indicator tube designed originally as a tuning indicator for radio receivers but capable of a wide range of applications. It is something of a cross between a triode and a cathode-ray tube. Basically, in its commonest form, it is a triode with one grid wire. The cathode is cylindrical, the grid is of the form of a narrow metal strip, and the plate is of the form of a wide-angle cone, which is coated with a fluorescent material and faces the end of the glass envelope of the tube so that the observer looks

---

[1] THOMPSON, H. C., Electron Beams and Their Application in Low Voltage Devices, *Proc. I.R.E.*, vol. 24, pp. 1276–1297, October, 1936.

in toward the apex of the cone. The arrangement of the electrodes is shown in Fig. 20.22.

When the grid, or control electrode, of the electron-ray tube described is sufficiently positive relative to cathode potential, electrons will move in all directions, giving a complete circle of illumination from the fluorescent material on the plate. As the control electrode is made negative, it will repel electrons from its immediate vicinity and cause a sectorlike shadow on the plate. As the control-electrode potential is made more negative, the shadow angle increases. It is possible to make a tube in which the shadow angle can be varied from 0 to 100 deg as the control-electrode potential is varied from a suitable positive value to zero.

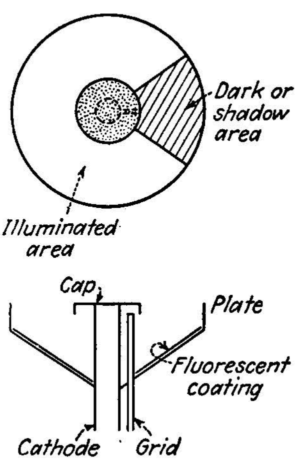


FIG. 20.22.—Construction of the electron-ray tube.

Characteristics of the 6AF6-G, which is an electron-ray tube with two identical and separate control electrodes, are shown in Fig. 20.23. Because of the relatively large variation in control-electrode potential required to give an appreciable change in shadow angle, the commonest types of electron-ray tubes contain a directly connected triode in the same envelope as the indicator electrodes. Examples of such tubes are the 6E5 and the 6G5. The characteristics of the 6E5 are shown in Fig. 20.24. These characteristics represent the combination effect of the triode amplifier and indicating device. The characteristics of the 6G5 are similar except that the sensitivity is about half as great. Many other electrode structures exhibit the property of having an electron ray whose position or width varies with the electrode potentials.[1]

20.10. Directed-ray Electron Tubes. In this class fall all the tubes in which the direction as well as the magnitude of the electron current is important. Several tubes in this category have already been mentioned. An example is the beam-power tube, in which the electron current is formed into parallel sheets by means of aligned control and screen grids. Likewise, the orbital beam tube, the 6SA7 special mixer heptode, and the electron-ray indicator tubes make use of electron rays directed in a specific manner to obtain the desired characteristics.

The directed-ray tubes fall into several classes according to the function that the directed ray is intended to perform. The possible functions include obtaining high current densities, a selective action between

---

[1] *Ibid.*

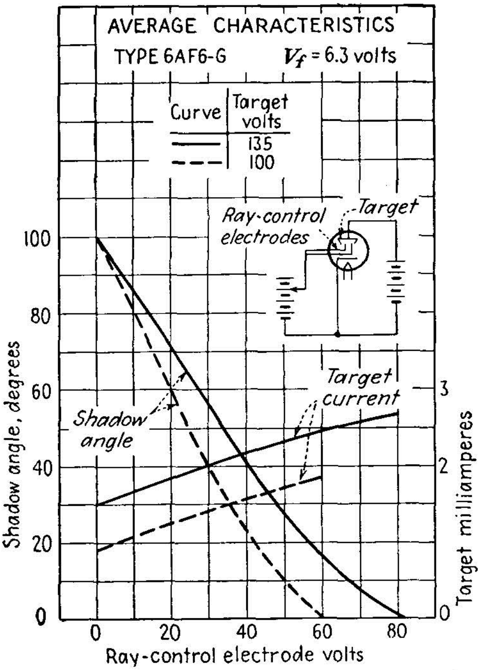


FIG. 20.23.—Electron-ray-tube characteristics.

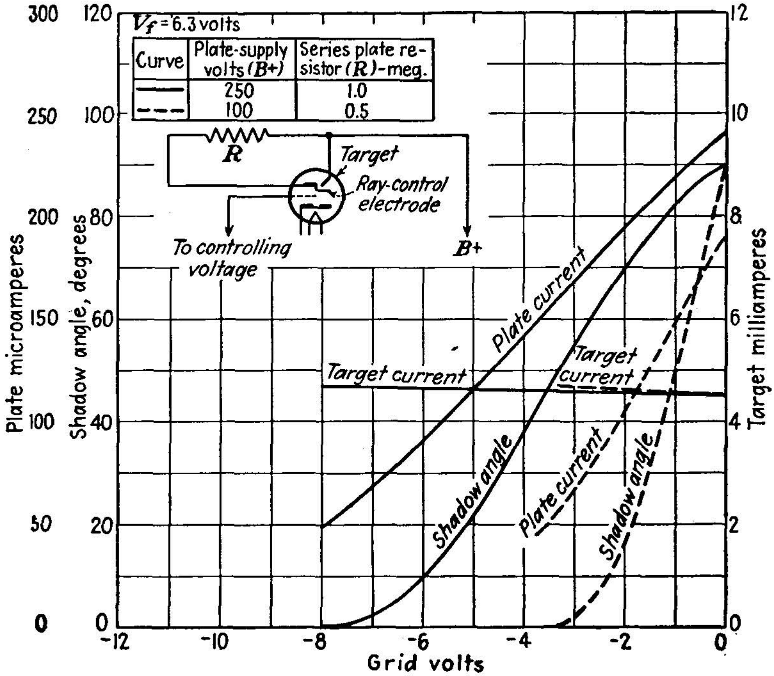


FIG. 20.24.—Triode-ray-tube characteristics.

electrodes by a focusing action, a variable electrode current by changing beam width, a discrimination between electrons of different velocities, a high or low fractional electron-current discrimination with respect to any particular electrode, and negative-resistance and negative-transconductance characteristics. This list is by no means complete.

It is a relatively simple matter to form electrons into definite beams of either sheet or circular form. In the ordinary triode the action of the control grid is such that the electrons leaving it tend to form into sheets. By proper use of grid wires and specially shaped electrodes a rather wide variety of beam patterns can be had. In Fig. 20.25 are shown a number of structures that can be used to produce specific electron-ray patterns. With all such devices, the angle of the electron beam will be a function of the electrode potentials, and the rate of change of angle or electrode current with any electrode voltage will be relatively slow because of the low amplification factors associated with a low number of grid wires.[1]

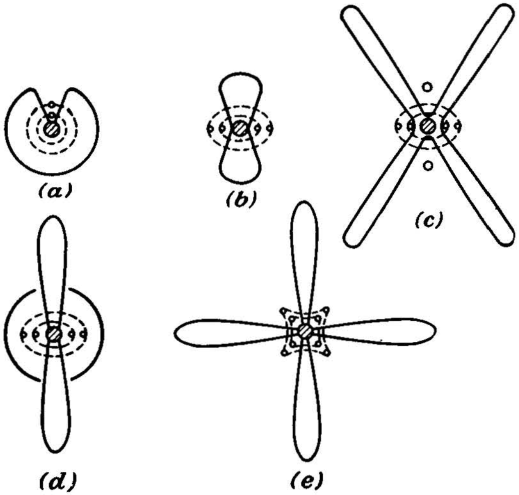


FIG. 20.25.—Directed-ray electron-tube structures.

Of particular interest are arrangements by which current to a set of grid wires may be kept low. In Fig. 20.26 are shown two such arrangements. The arrangement of Fig. 20.26*a* makes use of a cathode surface the cross section of which is scalloped in shape and against the points of which the grid wires are aligned. The curved equipotentials associated

---

[1] KNOLL, M., and J. SCHLOEMILCH, Elektronoptische Stromverteilung in gittergesteureen Elektronenröhren, *Arch. Elektrotech.* vol. 28, pp. 507–516, August, 1934.

with the scallops of the cathode surface focus the electrons so that they pass mostly between the grid wires even when the control grid is positive.[1] In the arrangement of Fig. 20.26*b* the emitting material is painted on a cylinder in a helical trace. The grid is likewise a helix of the same pitch as the emitting helix and positioned so that the wire lies opposite a non-emitting portion of the cathode cylinder. By these arrangements the control-grid current can be kept to less than 1 per cent of the space current when the control grid is as positive as the plate for a triode structure! In spite of this remarkable operating characteristic the difficulties of constructing such cathodes are great enough so that they have not found commercial application. In Fig. 20.27 is shown a low-screen-current tetrode. The low screen current is obtained by using squirrel-cage control and screen grids of the same number of wires and simply aligning the grids. With this arrangement the screen-grid current can be kept to a value as low as 0.2 per cent of the plate current.[2]

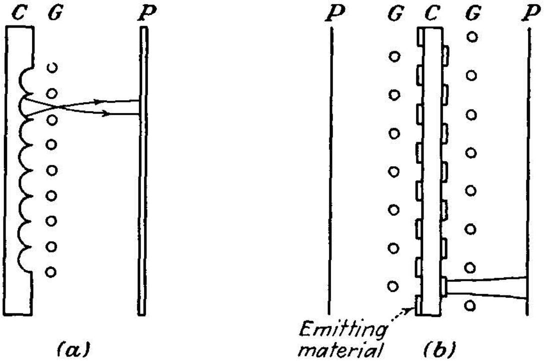


FIG. 20.26.—Electrode arrangements for reducing control-grid current.

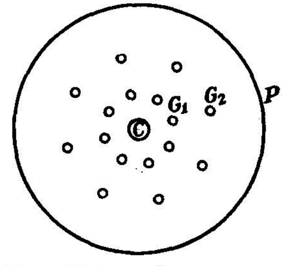


FIG. 20.27.—Low-screen-current tetrode.

20.11. Deflection Tubes. Most electron tubes make use of the variation of magnitude of electron current with electrode potentials to obtain their control characteristics. It would, however, be entirely feasible to obtain control characteristics from deflection of an electron beam rather than from

---

[1] KNOLL, M., Verstärker und Senderöhren als elektronenoptisches Problem, *Zeit. für Tech. Phys.*, vol. 15, pp. 584–591, December, 1934.

[2] THOMPSON, *op. cit.*

change in magnitude of an electron current.[1,2] In Fig. 20.28 are shown two forms of deflection tubes. In the tube of Fig. 20.28*a* a cylindrical cathode and two four-wire grids are used to form a four-lobed electron-current pattern The resultant lobes are then deflected by the trough-shaped external-corner electrodes so that the electron-beam lobes are effectively switched between the sections of the plate. In the arrangement of Fig. 20.28*b* a plane cathode and a parallel-wire control grid are used to form sheet electron beams, which are then deflected by an interleaved double screen grid so that the electron sheets are switched between the sections of an interleaved plate in the form of a double-strip grid. Alternate sections of the screen grid and plate are connected to opposite ends of the driving and output circuits, respectively.

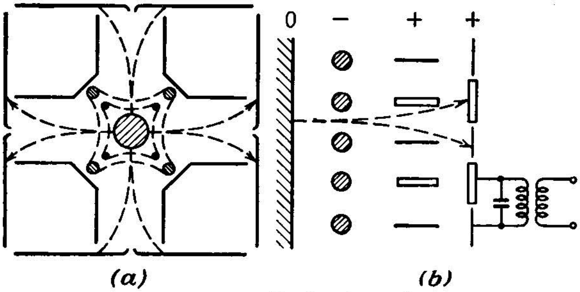


FIG. 20.28.—Deflection tubes.

It might be thought that it would be possible to make a deflection tube which would exhibit a nearly infinite transconductance by creating a beam with a sharp edge and then deflecting this past the edge of a collector electrode. This property does not seem realizable in practice, for two reasons. (1) Thermal velocities place a limit upon the maximum current density that can be achieved in a beam and upon the sharpness of the edge, as discussed in Sec. 15.5. (2) To realize a high effective mutual conductance it is necessary to place a large resistance in series with the collector electrode, and current flow through this resistance develops a voltage change that is in the direction to repel the electron beam being directed toward the electrode.

20.12. Television Camera Tubes. In a class by themselves are the television camera tubes. These tubes are means of electronically scanning a visual picture. They tax the tube designer's and the tube maker's art to the utmost, for they represent the most complicated

---

[1] HAZELTINE, A., Deflection Control Tubes, *Electronics*, vol. 9, pp. 14–16, March, 1936.

[2] ROTHE, H., and W. KLEEN, Die Bedeutung der Elektronenoptik in der Technik der Verstarkerrohren, *Zeit. für Tech. Phys.*, vol. 17 (No. 12), pp. 635–642, 1936.

assemblage of purely electronic components in existence. They involve the preparation of sensitive photoelectric and secondary-emissive surfaces. They involve beam formation and deflection, invoking all the tricks of combination electrostatic and magnetic manipulation. They are probably the most difficult of any tubes to make, and their successful development is a triumph of the application of fundamental electronic principles.

The Image-dissector Tube. One of the earliest purely electronic television viewing tubes developed was the so-called *image-dissector tube,*[1] a diagram of which is shown in Fig. 20.29. This tube contains a large photoelectric cathode upon which the picture to be imaged is focused. The photoelectrons liberated from the photocathode are attracted toward an anode in the form of a nickel wall coating designed so that the electrons from the photocathode are brought to a focus in a plane at the other end of the tube. In this way the electrons reproduce the current-density pattern corresponding to the original picture. In early tubes of this type the focusing was achieved by means of an axial magnetic field, but in later tubes by purely electrostatic means. In the plane of the focus of the photoelectrons there is located a pickup electrode shielded by an aperture so that this electrode picks up only the current corresponding to a small element of cathode area. The pickup electrode is followed by a secondary-electron multiplier to increase the sensitivity. Picture

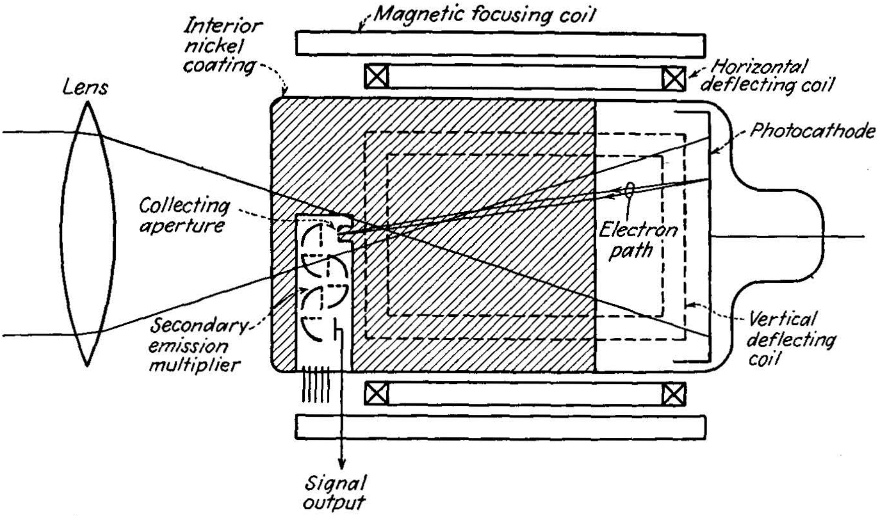


FIG. 20.29.—The Farnsworth image-dissector tube.

---

[1] FARNSWORTH, P. T., Television by Electron Image Scanning, *Jour. Franklin Inst.*, vol. 218, pp. 411–444, October, 1934.

scanning is achieved by deflecting the entire field stream of electrons from the photocathode so that the aperture in front of the collector electrode successively collects electrons from every portion of the picture-activated photocathode. Deflection is satisfactorily achieved by two pairs of magnetic coils external to the tube and producing fields at right angles to each other. The light image is projected upon the photocathode through the collector end of the tube, the obstruction caused by the collector structure and associated electron multiplier being negligible.

While the linearity of response of the image-dissector tube is virtually perfect, its sensitivity is very low, of the order of 50 microvolts per millilumen per cm^2 of cathode area. As a result, the tube is suitable only for outdoor televising and reproduction of motion pictures where the brightness of the objects to be viewed is quite high.

The Iconoscope. The iconoscope was the first of a series of television camera tubes developed to make use of a charge-storage principle.[1–4] The tube derives its name from the Greek derivatives "icon," meaning image, and "scope," signifying to view. The primary element of the tube is a mosaic of photoactive silver particles. These are deposited on a mica sheet and insulated from each other and from the sheet but are capacitively coupled to a metal backing on the other side of the mica sheet. The picture to be viewed is focused on this mosaic. The mosaic is scanned by an electron beam injected from an electron gun mounted at an angle with the mosaic. There is also a collector electrode in the same envelope. The video signal is obtained between the electron-gun anode and the conducting sheet backing the mosaic. The structure of the iconoscope and the associated electrical connections are shown in Fig. 20.30.

As the light image falls upon the mosaic screen, the various elements of the screen will emit photoelectrons in proportion to the intensity of the light falling upon them. The mosaic elements will thus become positively charged as they lose photoelectrons to the collector electrode. The mosaic elements act like a number of individual photoelectric cells all connected by capacity to the common signal plate, which is the conductive backing to the mica support. The elements of the mosaic are

---

[1] Zworykin, V. K., The Iconoscope, *Proc. I.R.E.*, vol. 22, pp. 16–32, January, 1934.

[2] Zworykin, V. K., Television, *Jour. Franklin Inst.*, vol. 217, pp. 1–37, January, 1934.

[3] Zworykin, V. K., Iconoscopes and Kinescopes in Television, *RCA Rev.*, vol. 1, pp. 60–84, July, 1936.

[4] Zworykin, V. K., G. A. Morton, and L. E. Flory, Theory and Performance of the Iconoscope, *Proc. I.R.E.*, vol. 25, pp. 1071–1092, August, 1937.

successively struck by the scanning beam. This process restores to them the charge that they have lost by photoemission and releases a corresponding charge to the signal backing plate to the mosaic. There thus flows through the signal-plate lead a current that is proportional to the light intensity of the areas scanned by the electron beam. A considerable gain in sensitivity is achieved by this arrangement by virtue of the fact that each of the picture elements is storing up charge continuously and so ideally the signal current is amplified by the number of picture elements over that obtained from such a tube as the image dissector. Actually, the process is only about 5 instead of 100 per cent efficient because of various detrimental effects to be described, but even at that the sensitivity of the tube is about 200 times as great as that of the image-dissector tube.

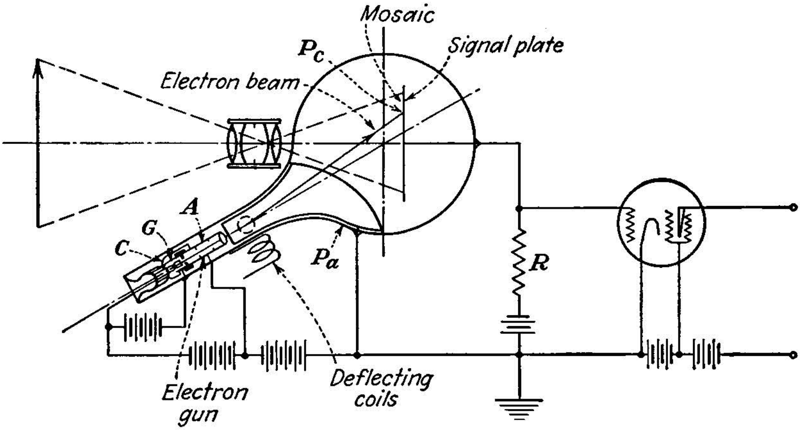


FIG. 20.30.—Iconoscope structure and circuit; C, cathode; G, control electrode; A, accelerating electrode; P_a, collector; P_c, photocathodes; R, load resistor.

The equivalent circuit of a mosaic picture element made up of many globules is given in Fig. 20.31. In effect, the mosaic globules in any picture element are equivalent to a photoelectric cell that is capacitively connected to an output resistor. As light falls upon the cell, electrons are passed slowly by the cell but gradually build up an appreciable charge and corresponding voltage upon the coupling condenser. The action of the beam is that of a separate circuit which dis-

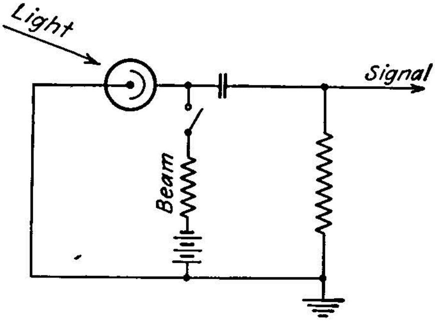


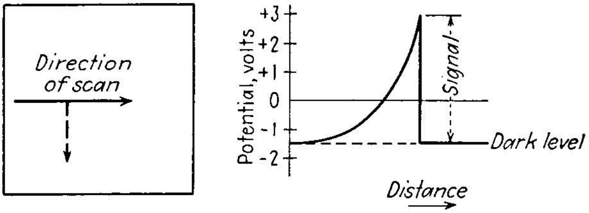
FIG. 20.31.—Equivalent circuit of a mosaic picture element.

charges the condenser periodically and thus releases a peaked current pulse through the output resistor.

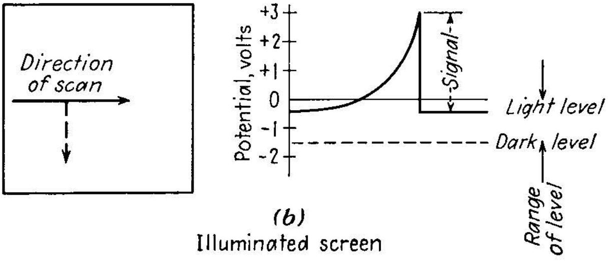
Several effects enter into the action of the iconoscope that prevent it from being perfect in its operation. For one thing, the photoelectric emission from the mosaic elements is space-charge limited. Also, the charge built up by the photoelectric emission is partially neutralized by a rain of secondary electrons all over the mosaic, originating from the impact of the beam primary electrons. Further, there is some loss of charge by surface leakage. All these effects combine to make the efficiency about 5 to 10 per cent of the theoretical maximum. The average brightness of field builds up the mosaic potential to a level such that the change in potential which the beam can effect is not the maximum value. As a result, the sensitivity of the tube is nonlinear and is only about one-fourth as much for high levels of illumination as for threshold values.

The potential behavior of the mosaic elements described earlier applies only if the secondary emission from the globules is negligible, which requires that the scanning electrons have only a few volts of energy. Ordinarily, the scanning electrons will have enough energy to produce five to seven secondary electrons for every primary electron, and the action of the mosaic screen will be quite different from that previously described. Since there are more electrons leaving than arriving on any mosaic element, when the scanning-beam potential is appreciable, the potential of the picture element being scanned will become positive instead of negative. Further, under the conditions of appreciable voltage of the scanning beam, the rain of secondary electrons falling upon elements of the mosaic not being scanned will exceed the number of photoelectrons being emitted, and the unscanned portions of the picture will assume a negative rather than a positive potential. The iconoscope still works under these conditions because the level of the negative potential assumed by the unscanned portions of the mosaic depends upon the picture illumination being relatively more positive (though still negative) in the illuminated areas than in the unilluminated areas.

The action described above is shown in Fig. 20.32 Figure 20.32*a* shows the potential response of an unilluminated portion of mosaic to the scanning beam. The rain of secondary electrons depresses the potential of such portions of the mosaic to about 1.5 volts negative relative to the collector. At this point a stable potential condition exists, for any further depression of potential would cause the mosaic to repel the secondary electrons. Thus in front of the scanning beam a potential of -1.5 volts exists. The mosaic elements under the beam become charged to about 3 volts positive, relative to the collector, by the emission of secondary electrons. This is also a stable potential, for if the element became more positive than this there would be a negative gradient of potential at the mosaic surface, which would prevent secondary-electron emission. The line behind the scanning beam rapidly becomes negative in potential again because of the acquisition of secondary electrons. Figure 20.32*b* shows the response of an illuminated section of mosaic to the scanning beam. Because of the emission of photoelectrons the potential of an unscanned portion of illuminated mosaic is less negative than for no illumination. The potential will, however, not become positive however much the mosaic is illuminated,



(a)
Unilluminated screen



(b)
Illuminated screen

FIG. 20.32.—Potential reaction of mosaic to scanning.

for then a negative gradient of potential would exist, which would prevent photoemission. The portions of the illuminated mosaic under the scanning beam rise to the same maximum, about 3 volts positive, as was the case for the unilluminated portions. Hence the signal is derived from the difference between the change in potential that exists between unscanned portions of illuminated and unilluminated portions of mosaic. For the figures quoted, it is seen that the maximum change in potential which can be achieved by illumination is 1.5 volts out of 4.5 volts, with the added condition that the change in potential level is nonlinear with illumination. From this factor alone, there is a loss of 67 per cent in efficiency. For ordinary levels of illumination restricted to the region of linear response the loss is about 80 per cent. Leakage and other factors bring the efficiency down to 5 to 10 per cent.

The structure of the mosaic is shown in Fig. 20.33. It consists of silver globules on a mica sheet of about 0.001 to 0.003 in. in thickness. The globules are formed by sifting a finely ground silver oxide powder onto the mica sheet and then reducing the silver by heating. The silver particles form into little globules ranging in size from 0.0005 cm in diameter down to particles microscopic in size. There are thus hundreds of mosaic particles scanned at any one instant by a beam of, say, 0.02 cm diameter, and as far as the beam action is concerned the mosaic may be considered as continuous. The conducting signal plate on the back side of the mica is formed by vaporizing or sputtering silver on that side of the mica. The globules on the front side of the mica are activated by much the same process as is used in making photoelectric surfaces. The silver globules are first oxidized and then exposed to caesium vapor to give a photoemissive surface. The signal plate is made just thick enough to become reasonably conducting and is usually backed by a heavy sheet of mica to prevent buckling or warping of the mosaic. The capacity of the mosaic to the signal plate ranges from 50 to 300 micromicrofarads per cm^2, and a value of 100 micromicrofarads per cm^2 is generally assumed in calculations. The photoelectric sensitivity of the emissive surfaces is of the order of 7 to 10 microamperes per lumen. The activated surface of the globules exhibits secondary emission as well as photoemission, the ratio of secondary to primary electrons ranging from five to seven for the usual conditions of operation. If the globules are activated with care, the insulation between them is very good, though some loss of charge is experienced from the fact that the resistance is not infinite.

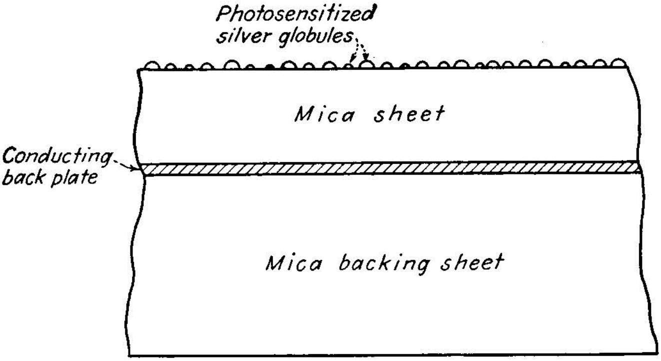


FIG. 20.33.—Structure of the iconoscope mosaic.

The Image Iconoscope. Various arrangements have been tested in the attempt to improve the efficiency of the iconoscope. One of these arrangements is found in the image iconoscope.[1,2] The structure of the image iconoscope is indicated in Fig. 20.34. This tube makes use of a mosaic screen as in the iconoscope but charges it by secondary-electron action. The mosaic is excited not with a light field but with a field of electrons of which the intensity pattern corresponds to that of the picture to be viewed. This field is generated by a transparent photocathode that is excited by the picture. This tube uses an electron lens, which is rather difficult to design, to focus the output of the photocathode upon the mosaic.[3] The charge on the mosaic is derived from secondary-electron emission, which exceeds the primary-electron current and also exceeds the photoelectric current in the ordinary iconoscope. Such tubes are capable of giving sensitivities as high as 5 millivolts per millilumen per cm^2. These tubes have not found great use because of the difficulty in constructing an electron lens that will reproduce the light image upon the mosaic without distortion.

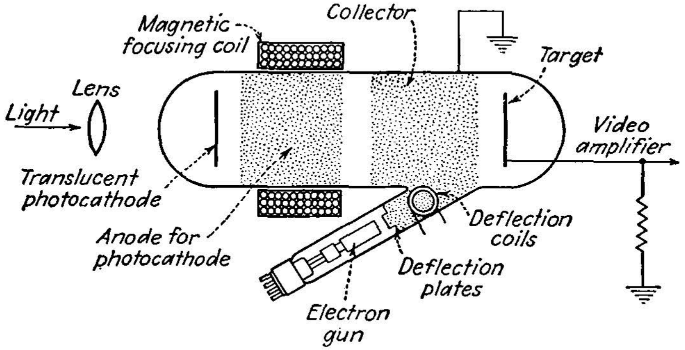


FIG. 20.34.—Structure of the image iconoscope.

The Orthicon. Another arrangement by which the low efficiency of the iconoscope is increased is the orthoiconoscope, usually abbreviated to orthicon.[4] The name is derived from the Greek prefix "ortho," meaning straight, which has reference to the linearity of characteristic. The orthicon overcomes some of the limitations of the iconoscope by

---

[1] *Ibid.*

[2] IAMS, H., and A. ROSE, Television Pickup Tubes with Cathode-ray Beam Scanning, *Proc. I.R.E.*, vol. 25, pp. 1048–1070, August, 1937.

[3] MORTON, G. A., and E. G. RAMBERG, Electron Optics of the Image Tube, *Physics*, vol. 7, pp. 451–459, December, 1936.

[4] ROSE, A., and H. A. IAMS, The Orthicon, *RCA Rev.*, vol. 4, pp. 186–199, October, 1939.

scanning the mosaic with electrons of such low velocity that secondary electrons are not created and hence do not neutralize the mosaic charge. A diagram of the orthicon is given in Fig. 20.35. The picture to be viewed is focused upon the mosaic, where it causes a variation in charge, as in the iconoscope. The scanning beam is generated by a flying spot of light upon a photocathode, which releases low-velocity electrons. Focusing of low-velocity electrons is difficult but is achieved in the orthicon by making use of the fact that low-velocity electrons will move in a tightly wrapped spiral around a strong magnetic-flux line. The scanning electrons generated by the light spot on the photocathode are guided to the mosaic by a strong curved magnetic field. When the scanning electrons approach a brightly illuminated spot on the mosaic, they are confronted with a positive potential, which draws them in and neutralizes the positive charge. When they approach a dark spot on the mosaic, they are repelled and return to the photocathode. Sensitivities of the order of 2 millivolts per millilumen per cm^2 have been attained with the orthicon. The conversion of the photoelectric scanning current into signal is believed to be nearly 100 per cent.

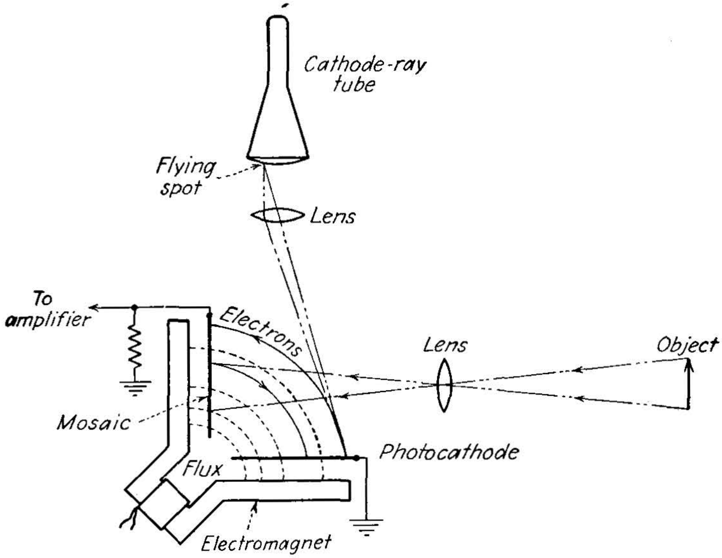


FIG. 20.35.—Structure of the orthicon.

The Image Orthicon. Still greater sensitivity can be obtained with a tube known as the "image orthicon," a diagram of which is given in Fig. 20.36. This tube combines features of the image iconoscope and orthicon. The light image is focused upon a transparent photocathode. The emitted current from the photocathode carries current-density variations, corresponding to the light-intensity variations in the picture to be viewed. The current field from the photocathode is focused upon a two-sided target by means of a suitable electron lens. The target surface is charged up according to the density of the exciting current by secondary-electron emission. The target is scanned with a low-velocity beam, which is deflected by magnetic means. The velocity of the scanning electrons is low enough so that no secondary electrons are created in the scanning process. The relative potentials of the mosaic and beam electrons are adjusted so that the scanning electrons will be attracted to neutralize the charge of the brightly illuminated areas but will be reflected from the dark areas. The signal is derived from the reflected electrons whose number will be an inverse function of the original picture illumination. In addition, the reflected electrons are multiplied by an electron multiplier, which helps increase the sensitivity and reduce the noise figure of the device. This tube is so sensitive that it can be used to view scenes illuminated with as little brightness as 0.01 candle per ft^2. The characteristics of the photocathode also make it possible to observe objects from their infrared radiation alone.[1,2]

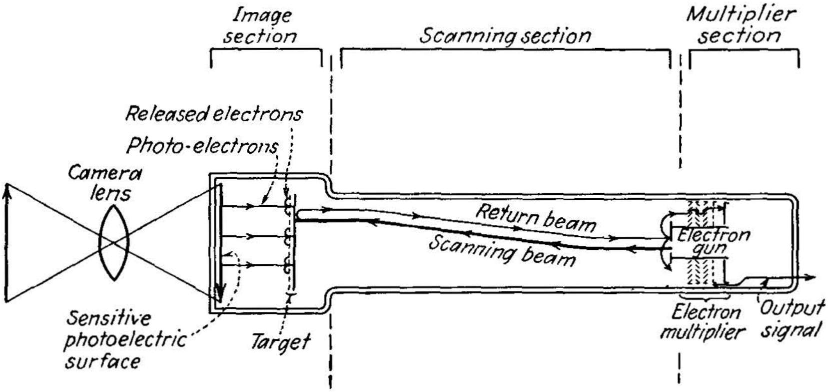


FIG. 20.36.—Structure of the image orthicon.

The Monoscope. The monoscope is not a camera tube but simply a standard picture-signal-generating tube. It is similar to the iconoscope in construction except that instead of the mosaic it has a fixed pattern printed on the oxide coating of an aluminum sheet. The secondary emission of the unprinted portions of the oxide coating is fairly high, while that of the printed portions is low. The printed pattern is scanned with an electron beam and gives rise to a variable secondary-electron

---

[1] Sensitive Television Camera Tube, *Electronics*, vol. 18, p. 330, December, 1945.

[2] ROSE, A., P. K. WEIMER, and H. B. LAW, The Image Orthicon, *Proc. I.R.E.*, vol. 34, pp. 424–432, July, 1946.

current to the collector electrode which corresponds to the printed signal. The signal from the collector electrode has the same polarity as that from the backing plate in an iconoscope. A negative picture signal may be obtained from the backing plate of the monoscope. The tube is used exclusively as a picture-signal generator for circuit-testing purposes.

20.13. The Electron Microscope. The electron microscope is not really a vacuum tube in the sense that it is available in sealed-off form that can be plugged into a circuit, but it is of sufficient importance to deserve a brief comment. The electron microscope extends the electron-optical analogy to the logical limit by actually using electrons to obtain an expanded image of an object. Use is made of the fact that electrons exhibit a wave as well as a particle behavior. Since electrons have appreciable path lengths only in a vacuum, it is necessary that the specimen be placed in a vacuum and the specimen is viewed by shooting electrons through it![1]

Structure of the Electron Microscope. In its usual form the electron microscope consists of a source of electrons, a condensing lens for the electron beam, and a specimen holder followed by two magnifying lenses. The lenses may be either electrostatic or magnetic. The electron source is usually a tungsten filament shielded by a cathode electrode so that emission occurs from only a small area of the filament. The electrons are accelerated by a unipotential gun structure constructed so that the angle of the electron ray is small. Since the magnetic type of electron microscope is thus far that most extensively used, the remainder of the remarks of this section will apply to it. A separate section will be devoted to the electrostatic type of microscope.

A magnetic condensing lens of the type shown in Fig. 14.8*e* is used to focus the electron beam upon the specimen. Just beyond the specimen to be viewed is placed an objective lens, also magnetic and of the same form as the condensing lens. Another magnetic lens, called the "projecting lens," is used to focus the image formed by the objective lens upon either a fluorescent screen or a photographic plate, both in vacuum. An intermediate image can be obtained before the projecting lens. The general structure involved is shown in Fig. 20.37.

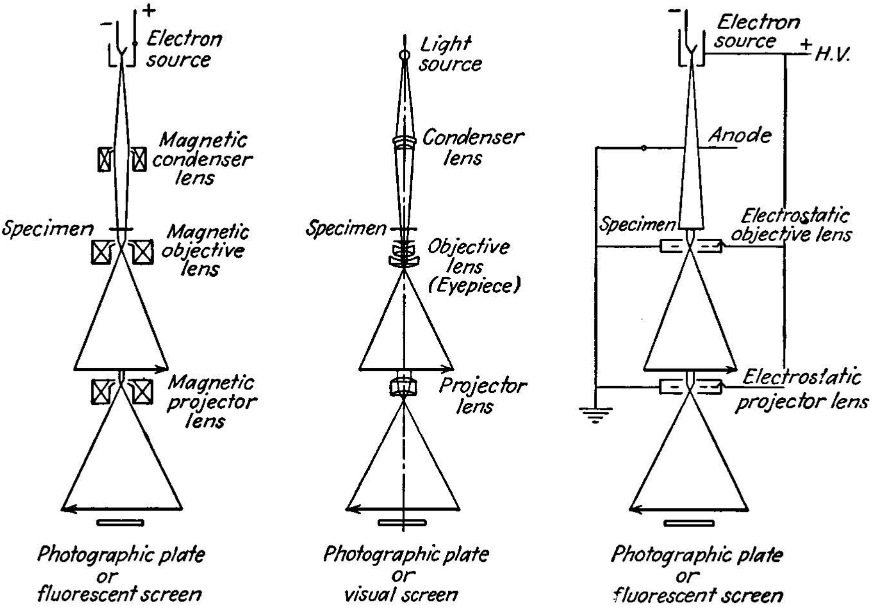
The specimen is applied to a thin film of collodion, which is supported

---

[1] The literature on electron microscopes runs into hundreds of articles and a score of books. The reader is referred to bibliographies on the subject by C. Marton and S. Sass, published in the *Journal of Applied Physics*, which list articles and books since the development of the electron microscope and are periodically extended.

See MARTON, CLAIRE, and SAMUEL SASS, A Bibliography of Electron Microscopy, *Jour. Appl. Phys.*, I, vol. 14, pp. 522–531, October, 1943; II, vol. 15, pp. 575–579, August, 1944; III, vol. 16, pp. 373–378, July, 1945.

upon a fine mesh screen, of 100 to 400 wires per inch. The collodion film is prepared by placing a drop of the material in liquid form on the surface of a dish of water, then raising the supporting screen through the film that forms from below so that a single layer of the film becomes attached to the screen, and then cutting away the excess collodion. The specimen to be observed is then deposited upon the collodion film. The specimen is admitted into the microscope through a rather intricate arrangement in the form of a vacuum lock with attendant manipulating levers. The microscope is left on a vacuum pump at all times, and when the specimen is removed a door is first opened into a small vacuum chamber and the specimen put into this chamber. The door between the chamber and the main body of the microscope is then closed, and another door opening to the outside is opened and the specimen removed. By this arrangement the entire microscope does not have to be evacuated every time a specimen is admitted or removed; rather, the vacuum pumps need remove only the small volume of the air admitted from the intermediate chamber. Photographic plates are admitted and removed by the same general scheme.



MAGNETIC ELECTRON MICROSCOPE

OPTICAL MICROSCOPE

ELECTROSTATIC ELECTRON MICROSCOPE

FIG. 20.37.—Structure of the electron microscope.

Equivalent Wave Length of Electrons. One of the revelations of modern physics is that there is a dual aspect of matter and energy. This applies to light rays, which may exhibit the properties of either waves or particles. Likewise, particles in motion exhibit the properties of light rays. It is the wave aspect of the electron that is utilized in the electron microscope. One of the teachings of quantum mechanics is that a particle in motion exhibits an equivalent wave length, known as the "De Broglie wave length," given by

$$\lambda = \frac{h}{mv} \tag{20.13}$$

where λ is the equivalent wave length in meters, h is Planck's constant whose value is 6.624×10^{-34} watt-sec², m is the mass of the particle in kg, and v is the velocity in meters per sec. For an electron at low velocities, the mass is constant, and the velocity is proportional to the square root of the potential through which it has been accelerated. Invoking the physical constants and Eq. (6.7*a*), the equivalent wave length of an electron that has been accelerated through a potential V is

$$\lambda = \frac{12.26}{\sqrt{V}} \qquad \text{angstrom units} \tag{20.14}$$

where V is in volts and 1 angstrom unit equals 10^{-10} meters. This expression is accurate within 1 per cent for voltages up to 20,000 volts. For higher voltages the relativity change of mass and the departure of velocity from the dependence upon the square root of potential must be considered. From Eq. (6.40*a*), the mass of an electron is seen to increase linearly with the potential through which it has been accelerated. Figure (6.3) and Eq. (6.39) give the dependence of electron velocity upon potential. Upon combining these last two relations with Eq. (20.13) there results the general equation for equivalent wave length of an electron that has been accelerated through V volts,

$$\lambda = \frac{12.26}{\sqrt{V}\,\sqrt{1 + 0.9788 \times 10^{-6}V}} \qquad \text{angstrom units} \tag{20.15}$$

This expression reduces to that of Eq. (20.14) for low voltages. A curve of equivalent wave length as a function of voltage is given in Fig. 20.38.

Theoretical Resolving Power of the Electron Microscope. The maximum useful magnification that can be obtained from a microscope is limited by the so-called "resolving power" and the lens defects. The resolving power of a microscope is measured by the least distance between two points that can just be distinguished and is determined by diffraction laws, which in turn depend upon the wave length of the light used. The least resolved distance of a microscope is usually given in terms of the Abbe formula, which has the form

$$d = \frac{0.5\lambda}{n \sin \alpha} \tag{20.16}$$

where d is the minimum distance separating two points which can be resolved, λ is the wave length of the light used, n is the index of refraction of the medium in which the object is situated, and α is the maximum angle which a ray leaving the central point of the object and entering the objective lens makes with the optical axis of the system. The quantity $n \sin \alpha$ is known as the "numerical aperture" of the lens. From the Abbe formula it is seen that the least resolved distance d is decreased as the wave length is decreased. This is supported experimentally by the fact that the highest resolution with optical microscopes is obtained with ultraviolet radiations. With ultraviolet wave lengths of the order of 2,500 angstrom units, refractive media with indices of refraction of a maximum value of about 1.6, and a maximum value of sin α, it is seen that the best that can be hoped for in the way of optical resolution is of the order of 1,000 angstrom units.

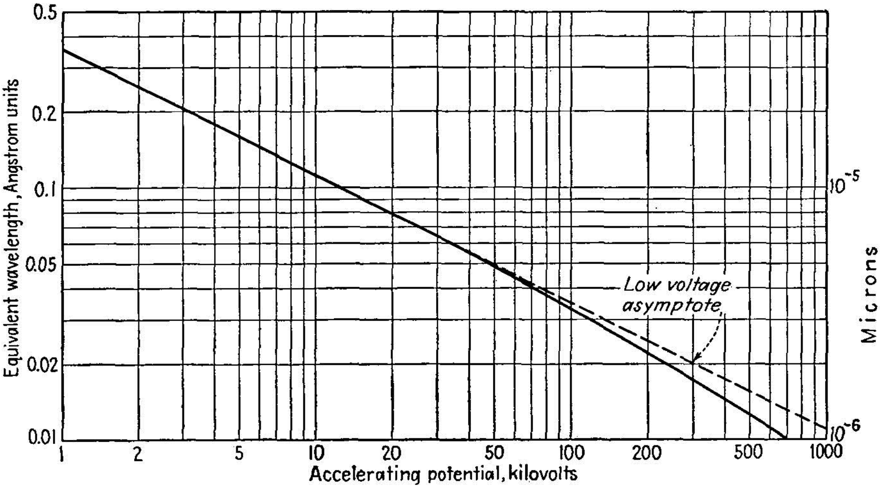


FIG. 20.38.—Equivalent wave length of an electron as a function of voltage.

It may be noted from Fig. 20.38 that the equivalent wave length of even low-voltage electrons is much less than the wave length of the shortest usable near-visible radiations. Hence the electron microscope has a tremendous opportunity of increasing resolution and magnification.

This it does, though as yet the maximum resolution is limited by the small apertures involved. The maximum useful magnification obtainable with an optical microscope is about 3,000 diameters. Electron microscopes have given useful magnifications as high as 100,000 diameters.

That the resolving power of an electron microscope is given by the Abbe formula may be demonstrated from a simple consideration of the Heisenberg principle of uncertainty. The principle of uncertainty states that the product of the error in determination of position and the error of determination of velocity of a particle is a constant. In other words, the more accurately the position is known, the less accurately the velocity can be known, and vice versa. Specifically, the principle of uncertainty states

$$\Delta x \, \Delta p = h \tag{20.17}$$

where Δx is the uncertainty of position, Δp is uncertainty of momentum mv, and h is Planck's constant. Upon applying this to a particle that is struck by an electron moving at an angle α with the axis of an electron microscope it is expected that the product of the indeterminancy of x of the particle by the indeterminancy of the x component of momentum will equal Planck's constant. If it is assumed that the point of contact of the electron with the particle is such that the tangent through this point is parallel to the optical axis, then the x component of momentum imparted to the particle is $2mv \sin \alpha$, where m is the electron mass and v is the electron velocity, and hence

$$\Delta x \, \Delta p_x = \Delta x 2mv \sin \alpha = h \tag{20.18}$$

Substituting the value of the De Broglie wave length from Eq. (20.13),

$$\Delta x = \frac{\lambda}{2 \sin \alpha} \tag{20.19}$$

which is the same as the Abbe formula of Eq. (20.16).[1] As an example, suppose that the numerical aperture of a lens is 0.0025 and that a 50,000-volt electron beam is involved. Then the De Broglie wave length, from Fig. 20.38, is about 0.05 angstrom unit and the corresponding theoretical least resolved distance is 10 angstrom units. The least resolved distance would be of the order of atomic dimensions if it were not for the lens defects. Actual least resolved distances of electron microscopes are of the order of 20 angstrom units.

Operating Principle of the Electron Microscope. The electron microscope produces an image because certain of the electrons in the beam on passing through the specimen have been scattered rather than refracted or absorbed. The scattering mechanism is entirely a dynamical one

---

[1] This derivation is attributed by Marton to Henriot.

resulting from the reaction of the electron charge with the electrostatic fields of the atoms and molecules that the electron approaches. Some of the scattered electrons are intercepted by the apertures. Others undergo single or plural scattering and remain in the field. Variations in density of the resultant picture occur because of the scattering process, which shifts some of the electrons from the portions of the picture corresponding to dense parts of the specimen. Thus the electron-microscope picture resembles an X-ray picture more than an optical picture. Some photographs of soap-curd fibers obtained with the electron microscope are shown in Fig. 20.39.

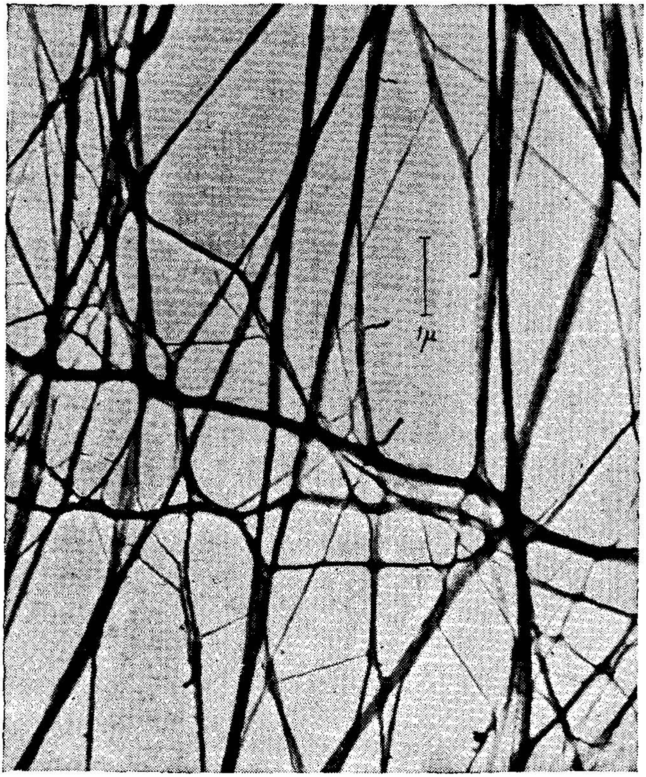


FIG. 20.39*a*.—Soap-curd fibers. Picture taken with the electron microscope. Magnification is about 10,000 times: $1\mu = 1$ micron $= 10^{-6}$ meters. (*Courtesy of J. W. McBain.*)

Limits of Resolving Power. A number of factors conspire to make the resolving power of the electron microscope less than the theoretical value. In the first place all the voltages and currents of the microscope are subject to some variation, which introduces a fuzziness in the pictures.

This can, however, be virtually eliminated by stabilizing the voltages and currents to a sufficient degree. The degree of stability required for a least resolved distance of 10 angstrom units is rather large, being of the following order:

Voltage stability	1 part in 7,000
Current stability	
Condenser lens	1 part in 1,000
Objective lens	1 part in 14,000
Projection lens	1 part in 1,500

FIG. 20.39*b*.—Soap-curd fibers. Picture taken with the electron microscope. Magnification is about 100,000 times. (*Courtesy of J. W. McBain.*)

Furthermore, any stray magnetic fields must be reduced by shielding so that components normal to the axis are weaker than 5×10^{-6} gauss.

In addition to the above there are all the various lens aberrations to be coped with.[1] Most of these are extraaxial so that they can be reduced

---

[1] An excellent discussion of the limitations of the resolving power of electron microscopes is given by MARTON, L., and R. G. E. HUTTER, The Transmission Type of Electron Microscope and Its Optics, *Proc. I.R.E.*, vol. 32, pp. 3–12, January, 1944.

by using a small beam angle, with attendant loss of resolving power. Whereas beam angles in optical microscopes may be quite large, values of α for electron microscopes are of the order of 10^{-2} to 10^{-5} radian. Some lens errors are axial and cannot be eliminated. These are spherical and chromatic aberration and also the diffraction defects. Scattering also imposes some irremediable limitations.

Chromatic aberration is proportional to the beam angle and thus may be reduced by keeping the beam angle low. It is principally due to changes in velocity incurred when the electrons pass through the specimen. It may be reduced by using thin specimens and supporting films and materials of low atomic number.

Spherical aberration increases as the cube of the beam angle and with the beam voltage because the minimum focal lengths of the magnetic lenses increase with beam voltage, owing to saturation effects.

Diffraction errors are proportional to the equivalent wave length and inversely proportional to the beam angle. Since the diffraction error decreases with beam angle, while the spherical aberration increases with beam angle, there is an optimum resolution which occurs at the angle at which the two errors are approximately equal.

Because the De Broglie wave length decreases with increasing electron energy, it might be thought that increasing improvement in resolution could be achieved by simply going to higher voltages. This is not realized in practice, largely because the magnetic lenses lose strength through saturation, as a result of which spherical aberration and diffraction actually increase. A reduction of chromatic aberration and an increase in penetration power are realized; but, in spite of this, pictures obtained with beam voltages greater than 100 kv are not noticeably superior to those taken with beam voltages between 50 and 100 kv except for the greater penetrating power evidenced.[1]

Electrostatic Electron Microscopes. The structure of an electrostatic electron microscope is also shown in Fig. 20.37.[2] A shielded tungsten filament and a unipotential cathode gun are used to produce and accelerate the electrons. Three lenses are used, and these have the same function as the corresponding lenses in the magnetic type. The electrostatic lenses are of the *Einzel* lens type and are dimensioned so that the center electrode of each lens is operated at cathode potential. This makes the lens action independent of the voltage used, for the focal length depends only on the shape of the field. As a result, the lens voltage supply

---

[1] ZWORYKIN, V. K., J. HILLIER, and A. W. VANCE, Preliminary Report on the Development of a 300 Kilovolt Magnetic Electron Microscope, *Jour. Appl. Phys.*, vol. 12, pp. 738–742, October, 1941.

[2] BACHMAN, C. H., and SIMON RAMO, Electrostatic Electron Microscopy, *Jour. Appl. Phys.*, vol. 14, pp. 8–18, 69–72, 155–160, January, February, April, 1943.

does not have to be as carefully stabilized as is the case with the magnetic type of electron microscope. Focusing is achieved in the General Electric model by moving the specimen physically without changing the lens characteristics. All the lens defects encountered in the magnetic type of microscope appear in the electrostatic type and are distinctly greater in magnitude. As a result, the least resolved distance as yet obtained with an electrostatic microscope is about 80 angstrom units. Hence the electrostatic types developed thus far are inferior in magnification to magnetic types by about a factor of 4. This limitation is offset by an appreciable reduction in cost and size.

CHAPTER 21

HIGH-VACUUM PRACTICE

21.1. Introduction. The construction of vacuum tubes requires a high degree of skill and a great knowledge of the techniques associated with obtaining and maintaining a high vacuum. It may be said without exaggeration that the problem of producing a good vacuum tube depends about 90 per cent upon the knowledge of high-vacuum techniques. Without a knowledge of these techniques a knowledge of the theory of vacuum-tube design is useless. Much has been written on the subject of high-vacuum techniques, but probably as much knowledge exists that has never been recorded. This brief chapter cannot do more than collect the most important relations and facts concerning high-vacuum practice.[1]

In answer to the question as to what is meant by high vacuum it is first necessary to define the units in which vacuums are measured. There are a number of systems of units that are commonly used to represent the degree of vacuum. Vacuums may be described in terms of a fraction of atmospheric pressure, 760 mm of mercury column. They may be described in terms of the absolute gas pressure in units of bars, 1 bar being nearly equal to 1 dyne per cm^2 and 1,000,000 bars being nearly equal to atmospheric pressure, actually 750 mm of mercury column. Vacuums may also be measured in terms of the height to which the gas whose pressure is being measured will raise a column of mercury. This height is 760 mm at atmospheric pressure and will be proportionately less for gases whose pressure is less than atmospheric. This method of representing pressure of vacuum has a definite physical significance in that it is possible to devise an apparatus which will give direct measurements in terms of a mercury column for heights of the mercury column as low as 10^{-4} mm. Sometimes the height of the mercury column is expressed in

---

[1] The most useful books devoted entirely to high-vacuum practice are

DUSHMAN, S., "The Production and Measurement of High Vacuum," *General Electric Review*, Schenectady, New York, 1922.

DUNOYER, L., "Vacuum Practice," Van Nostrand, New York, 1926.

ESPE, W., and M. KNOLL, "Werkstoffkunde der Hochvakuumtechnik," Springer, Berlin, 1936.

YARDWOOD, J., "High Vacuum Technique," Chapman & Hall, Ltd., London, 1943.

units of microns, 1 micron being $\frac{1}{1,000,000}$ meter. This has the virtue that the numbers are a little easier to write for low vacuums. In Fig. 21.1 is given a chart comparing the different scales for measuring vacuums.

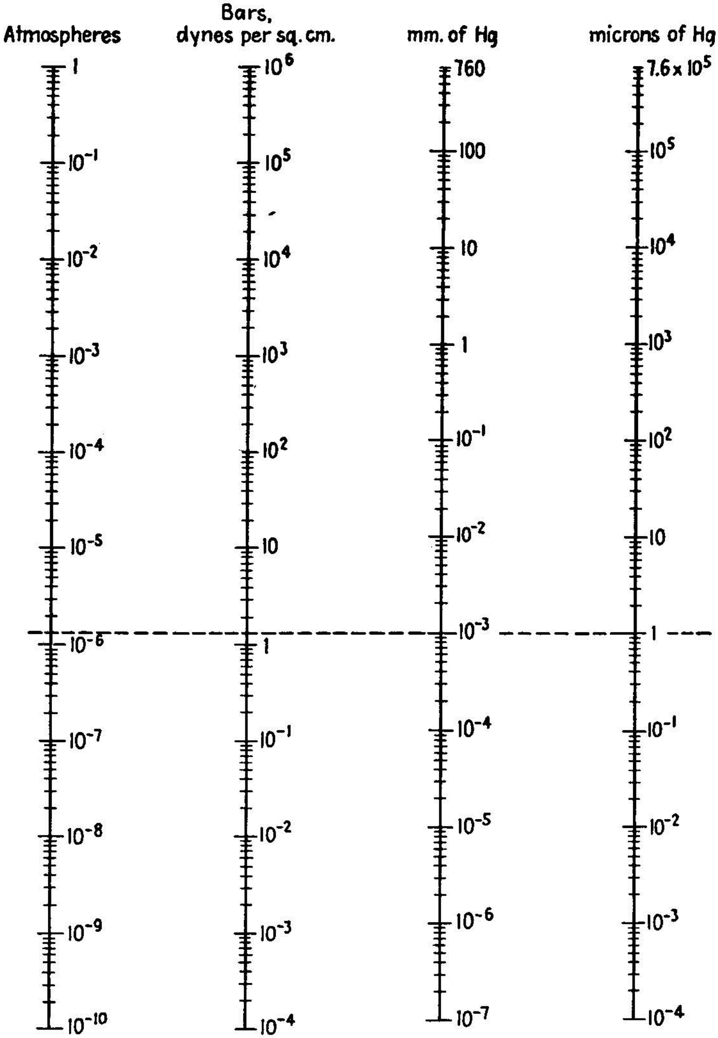


FIG. 21.1.—Comparison of pressure scales.

The values along any horizontal line correspond to the same gas pressure. Thus 1 micron equals 10^{-3} mm of mercury and corresponds to 1.333 bars or 1.318×10^{-6} atmosphere.

Pressures of 10^{-4} mm of mercury or less are referred to as "high vacuums." A pressure of 10^{-7} mm of mercury is referred to as a "hard vacuum." One of 10^{-3} mm of mercury or less is referred to as a "soft" or "low vacuum." An idea of the scale of vacuums encountered in electronic devices is given by the following tabulation:

Pressure, Mm of Mercury	Characteristic or Device
760	Atmospheric pressure
100	Gas-filled Lamps
10	
1	Spark streamers in electrical discharge
10^{-1}	Glow discharge, neon lamps
10^{-2}	Lower limit of glow discharge
10^{-3}	Glass fluoresces under electron bombardment
10^{-4}	Bad receiving tube
10^{-5}	Old receiving tube, operating
10^{-6}	New receiving tube, operating Old transmitting tube, operating
10^{-7}	New transmitting tube, operating
10^{-8}	New tube, cold

21.2. Fundamental Gas Laws. Since all vacuum processes are merely operations upon gases at pressures less than atmospheric, it is important to review the laws governing the behavior of gases. There will also be included in this section comments upon the behavior of molecules in a gas and of electrons in a gas.

Boyle's Law. Boyle's law states that the volume which a given mass of gas at a fixed temperature occupies varies inversely as the pressure to which the gas is subjected. Mathematically this is stated by

$$P_1V_1 = P_2V_2 = \text{const} \tag{21.1}$$

Charles's, or Gay-Lussac's Law. Charles's, or Gay-Lussac's, law states that the volume which any mass of gas occupies at a given pressure varies directly with the temperature. Mathematically this is stated

$$\frac{V_1}{T_1} = \frac{V_2}{T_2} = \text{const} \tag{21.2}$$

Avogadro's Law. Avogadro's law states that the number of molecules in equal volumes of gases at the same temperature and pressure are equal. More specifically, the number of molecules in a *mole* or in a mass in grams of substance numerically equal to its molecular weight is always the same regardless of the kind of gas. The number of molecules in a mole is known as *Avogadro's number* and is equal to 6.023×10^{23}.

General Gas-expansion Law. Boyle's, Charles's, and Avogadro's laws can be combined into a single law,

$$PV = RmT \tag{21.3}$$

where P is pressure in bars or dynes per square centimeter, V is volume in cubic centimeters, R is a universal gas constant having a value of 8.314×10^7 ergs per deg per mole, m is the mass of the gas in moles, *i.e.*, the mass in grams divided by the molecular weight, and T is the temperature on the absolute scale, which is 273° plus the number of degrees centigrade. From this equation and Avogadro's number it is readily calculated that the number of molecules in 1 cm³ at a pressure of 10^6 bars (750 mm of mercury, nearly atmospheric) and 0°C is 2.654×10^{19}. Even at the very low pressure of 10^{-7} mm of mercury, about 10^{-10} atmosphere, the number of molecules per cubic centimeter of a gas is about 3,000,000,000. In general, the number of molecules per cubic centimeter of a gas is given by

$$n = 7.244 \times 10^{15} \frac{P}{T} \quad \text{molecules per cm}^3 \tag{21.4}$$

where P is pressure in bars and T is absolute temperature.

Distribution of Velocities in a Gas. The heat energy that a body of gas contains exists in the kinetic energy of motion of the gas molecules. As the temperature is increased, the heat energy increases and the velocity of the molecules increases. The molecules will have velocities in all directions and with all magnitudes, but most of them will have velocities grouped around a most probable velocity. Maxwell has shown from application of the theory of probability that the distribution of velocities of molecules in a gas is given by

$$y = \frac{4}{\sqrt{\pi}} x^2 \epsilon^{-x^2} \tag{21.5}$$

where x is the ratio of velocity to the most probable velocity and y is the corresponding probability that a molecule will have a velocity x. A plot of Eq. (21.5) is given in Fig. 21.2. The area under the curve between any two values of x, say x_1 and x_2, divided by the total area under the curve gives the fraction of the total number of molecules that have velocities in the interval between x_1 and x_2. The coefficient of Eq. (21.5) is chosen so that the total area under the distribution curve is unity. This causes the maximum ordinate to be other than unity but makes the estimate of the fraction of the total number of molecules having velocities in any given velocity interval very simple. Thus the area under the curve between values of velocity x_1 and x_2 gives the fraction directly and may be estimated by simply counting squares, *each square representing 1 per cent of the total number* for the scale divisions given. Thus only about 5 per cent of the total number of molecules have velocities greater than twice the most probable velocity.

The distribution curve of Fig. 21.2 holds for all temperatures, the only difference being that the most probable velocity increases with the square root of the absolute temperature.

$$v_p = 12{,}900\sqrt{\frac{T}{M}} \qquad \text{cm per sec} \tag{21.6}$$

where T is the absolute temperature and M is the molecular weight of the gas. Also of interest is the average velocity, which is 1.124 times the most probable velocity,

$$v_a = 1.124v_p \tag{21.7}$$

where v_a is the average velocity and v_p is the most probable velocity.

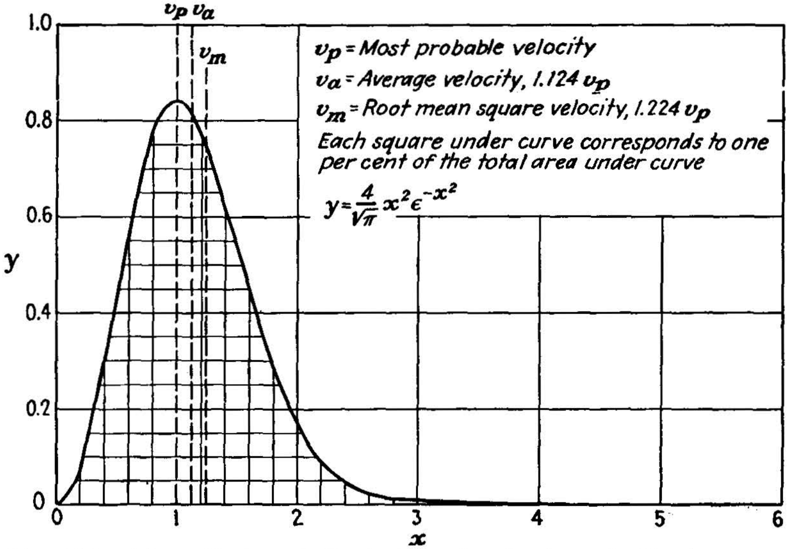


FIG. 21.2.—Maxwell's law of distribution of molecular velocities in a gas.

The rms velocity is involved in energy calculations and has the value of 1.224 times the most probable velocity,

$$v_m = 1.224v_p \tag{21.8}$$

where v_m is the rms velocity and v_p is the most probable velocity. All the velocities cited above are independent of the pressure involved. A curve showing average velocities at room temperature of various gases as determined by their molecular weight is given in Fig. 21.3.

Mean Free Path of a Gas Molecule. Although it is true that the molecules of a gas have rather high velocities, it is a common observation that gases diffuse together very slowly. This is undoubtedly due to the fact that the molecules collide frequently with one another and so move in zigzag paths made up of rather short straight-line segments. The term *mean free path* is used to indicate the average length of path of a molecule between collisions. The collisions themselves are of an elastic nature and tend to leave the magnitude of the velocities involved unchanged on the average.

An estimate of the mean free path of a gas molecule among other gas molecules can be had by considering the total area of molecules in a volume of cross-sectional area A and thickness t. The number of molecules in such a volume is nAt, where n is the molecular density. Let the volume be viewed from the surface of area A, and assume that the molecules are uniformly distributed throughout the volume in such a way that their projections upon the surface of area A are also uniformly distributed. When the number and arrangement of the projections of the molecules upon the surface of area A have the form indicated in Fig. 21.4, then it will be impossible for a molecule to travel a distance t perpendicular to the surface of area A without making contact with another molecule. The equilateral triangle shown in Fig. 21.4 has an altitude of 1.5 molecular diameters and so has an area of $\frac{3\sqrt{3}\,d_m^2}{4}$ and contains effectively one-half of a molecule. The density of the molecules as projected on the surface of area A is therefore $\frac{2}{3\sqrt{3}\,d_m^2}$, where d_m is the molecular diameter. Equating the number of molecules projected

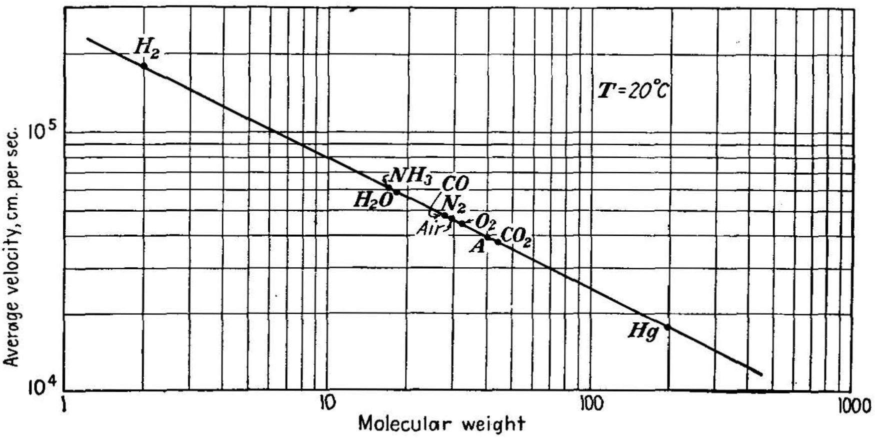


FIG. 21.3.—Average velocity of gases at room temperature.

on the surface for this density to the number of molecules contained in the volume,

$$\frac{2A}{3\sqrt{3}\,d_m{}^2} = nAt \qquad (21.9)$$

from which

$$t = \frac{1}{2.60nd_m{}^2} \qquad (21.10)$$

There is a 100 per cent probability that a molecule will collide with another molecule in a distance t, or roughly a 50 per cent probability that a molecule will collide with another in a distance $\frac{t}{2}$ which will be called the "mean free path" of a molecule. Accordingly,

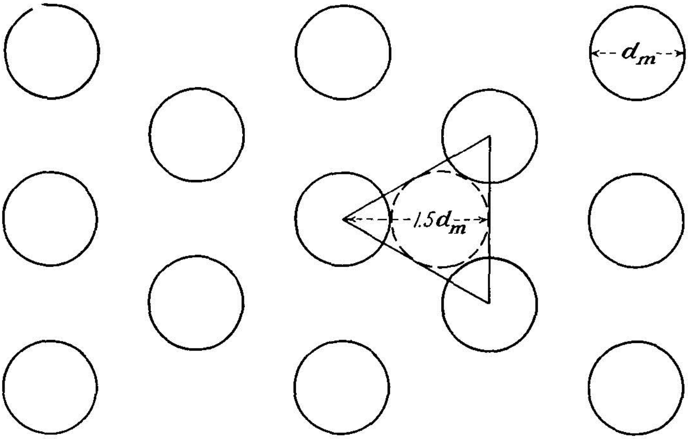


FIG. 21.4.—Arrangement of molecules in a gas to illustrate the concept of the mean free path of a gas molecule.

$$\text{Approximate mean free path of gas molecule} = \frac{0.1923}{nd_m{}^2} \qquad (21.11)$$

This formula is only approximate, for it does not consider the random distribution of molecular velocities. Maxwell has considered this problem and proposed the following formula:

$$\text{Mean free path of gas molecule} = \frac{1}{\sqrt{2}\pi nd_m{}^2} = \frac{0.225}{nd_m{}^2} \qquad (21.12)$$

Some further refinements on this formula give the coefficient as 0.315, but for ordinary purposes Maxwell's form as given in Eq. (21.12) is generally used. The mean free path of a molecule of a gas is seen to be an inverse function of the molecular density and also an inverse function of the equivalent molecular cross-sectional area or diameter squared.

The molecular diameter is a convenient fiction useful in many gas relations. The agreement between the values of molecular diameter as determined from various considerations is fairly good.[1] Values of the molecular diameter of various gases are given in the following table. Values are given only to two significant figures since the agreement between various determinations is no greater than this.

TABLE XIII
MOLECULAR DIAMETER OF THE GASES

Gas	Molecular Diameter, Cm
A	2.9×10^{-8}
CO	3.2×10^{-8}
CO_2	3.3×10^{-8}
H_2	2.4×10^{-8}
He	1.9×10^{-8}
Kr	3.2×10^{-8}
N_2	3.1×10^{-8}
Ne	2.35×10^{-8}
NH_3	3.0×10^{-8}
O_2	3.0×10^{-8}
Xe	3.5×10^{-8}

Since the molecular density as given by Eq. (21.4) is directly proportional to pressure, it follows that the mean free path of a molecule is inversely proportional to pressure. Upon combining Eqs. (21.4) and (21.12) the mean free path is given by

$$\text{Mean free path of gas molecule} = \frac{3.107T}{Pd_m^2} \times 10^{-17} \quad \text{cm} \qquad (21.13)$$

where T is temperature in degrees absolute, P is pressure in bars, and d_m is molecular diameter in centimeters. A curve of the mean free path of nitrogen, the principal ingredient of air, as a function of pressure is given in Fig. 21.5. It is convenient to remember that the mean free path of nitrogen at room temperature and a pressure of 1 bar (about 10^{-3} mm of mercury) is approximately 10 cm and varies inversely with the pressure.

The average number of collisions of a gas molecule per centimeter of travel is the reciprocal of the mean free path and is given by

$$\text{Collisions per cm} = 3.219 \times 10^{16} \frac{Pd_m^2}{T} \qquad (21.14)$$

in which the units are the same as for Eq. (21 13).

---

[1] Dushman, *op. cit.*, p. 27.

Mean Free Path of an Electron among Gas Molecules. The reasoning that was used in the previous subsection to get an estimate of the mean free path of an electron may be used to get an estimate of the mean free path of an electron moving among gas molecules. Since the electron has a negligible cross-sectional area compared with a molecule, if we again consider a volume of surface area A and thickness t, then the molecules must be sufficient in number and must arrange themselves as shown in Fig. 21.6 to ensure that there will be a collision of an electron with a gas molecule in a distance t. For the arrangement of molecules shown, each equilateral triangle of altitude $\frac{3d_m}{4}$ has an area of $\frac{3\sqrt{3}\,d_m{}^2}{16}$ and contains effectively half a molecule. Accordingly, the density of the molecules as projected upon the surface of area A is $\frac{8}{3\sqrt{3}\,d_m{}^2}$, or just four times the density required to ensure a molecule collision in a distance t. One may therefore expect that the mean free path of an electron among gas molecules is four times that of the gas molecules themselves. Needless to say, this is a very rough estimate and applies only to electrons having velocities corresponding

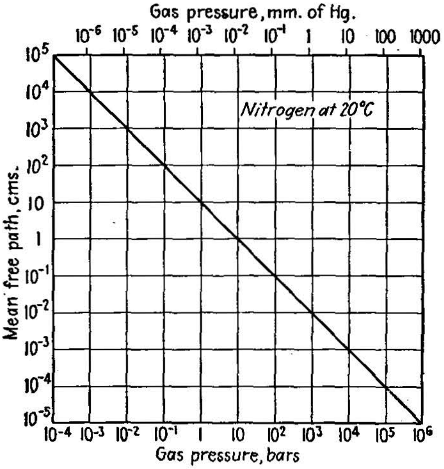


FIG. 21.5.—Mean free path of a nitrogen molecule as a function of pressure.

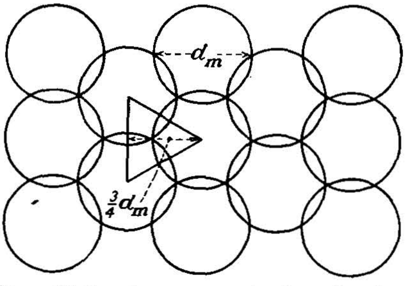


FIG. 21.6.—Arrangement of molecules in a gas to illustrate the concept of mean free path of an electron.

to a few volts. Even at velocities corresponding to about 10 volts the above estimate will not hold, for the exchange of energy between electrons and molecules is extremely complex in the vicinity of the ionization potential of the gas.[1]

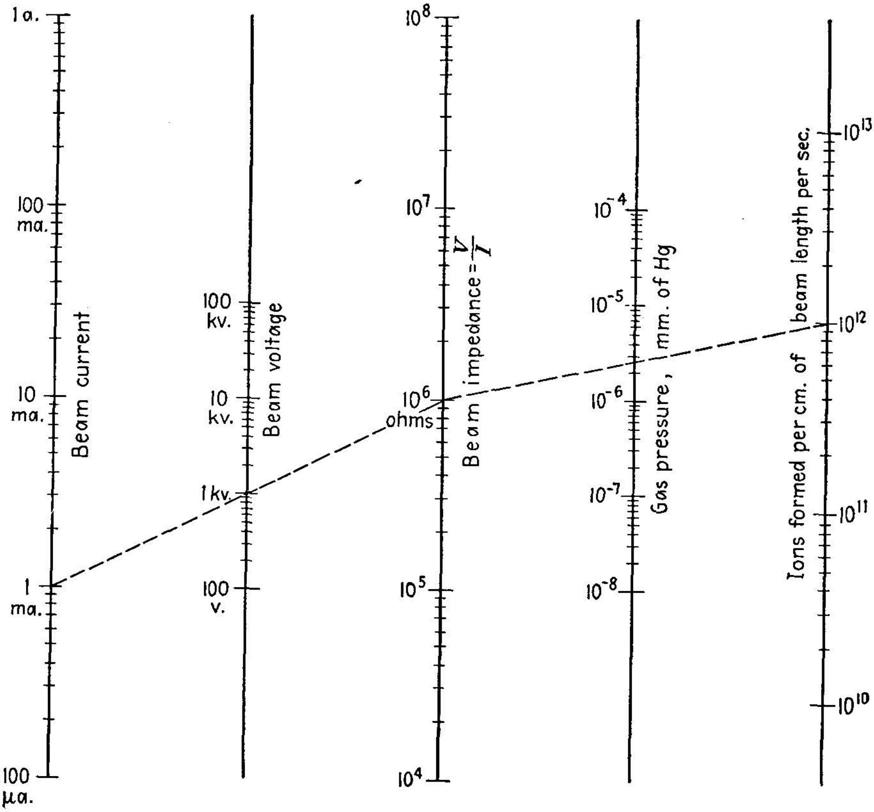


FIG. 21.7.—Nomographic chart of the number of ions formed per centimeter of electron beam per second as a function of current, voltage, and pressure.

The concept of the mean free path of an electron among gas molecules needs revision when the electron velocity becomes appreciable. As the electron acquires a higher voltage, it may approach closer to a molecule before it will be deflected or before it will produce ionization. This is because there is less time for the electrostatic forces to effect a transfer of energy as the electron velocity increases. Of more interest than the mean

---

[1] BRODE, R. B., Quantitative Study of the Collisions of Electrons with Atoms, *Rev. Modern Phys.*, vol. 5, pp. 257–279, October, 1933.

free path of an electron, which is difficult to define, is the distance between ionizing collisions. Above a velocity corresponding to about 500 volts for heavy molecules such as mercury and above a few hundred volts for the molecular constituents of air, it is found that the number of ions formed per centimeter per second by an electron beam is proportional to the current and the gas pressure and inversely proportional to the beam voltage. A satisfactory empirical form[1] of this relation for air is

$$N = \frac{3.75 \times 10^{23} Ip}{V} \qquad \begin{array}{c}\text{ions formed per sec}\\ \text{per cm of length}\end{array} \tag{21.15}$$

where I is current in amperes, V is voltage through which the electrons have been accelerated in volts, and p is gas pressure in millimeters of mercury. A nomographic chart of the number of ions formed per centimeter of path per second is given in Fig. 21.7. Since current is the product of the number of electrons passing a reference plane per second multiplied by the charge, the number of ions formed by each electron per centimeter of travel is

$$\begin{array}{c}\text{Number of ions formed by 1}\\ \text{electron per cm of travel}\end{array} = \frac{6 \times 10^4 p}{V} \tag{21.16}$$

where p is in millimeters of mercury and V is in volts. The distance between ionizing collisions of an electron is the reciprocal of Eq. (21.16), or

$$\begin{array}{c}\text{Distance between ionizing}\\ \text{collisions of an electron}\end{array} = \frac{V_{kv}}{60p} \qquad \text{cm} \tag{21.17}$$

where V_{kv} is the potential through which the electrons have been accelerated, in kilovolts, and p is pressure in millimeters of mercury. Thus an electron with a velocity corresponding to 6,000 volts, moving in a vacuum of 10^{-6} mm of mercury, experiences an ionizing collision every 100,000 cm. At 10^{-4} mm of mercury the distance between ionizing collisions is only 1,000 cm. A nomographic chart showing the distance between ionizing collisions, as a function of gas pressure and voltage through which the electron has been accelerated, is given in Fig. 21.8.

21.3. Measurement of Vacuum. The range of pressures over which vacuum devices operate is so large that no one pressure-measuring device can cover it. Accordingly, it is necessary to use a number of devices to handle the entire range of pressures from atmospheric down to the best vacuums producible. The number of types of vacuum gauges runs into the dozens, but of these there are about half a dozen that have shown more

---

[1] BENNETT, W. H., Magnetically Self-focusing Streams, *Phys. Rev.*, vol. 45, pp. 890–897, June 15, 1934. Equation (21.15) is given originally as $N = 200pI/Ve$, with p in millimeters of mercury and electrical quantities in esu.

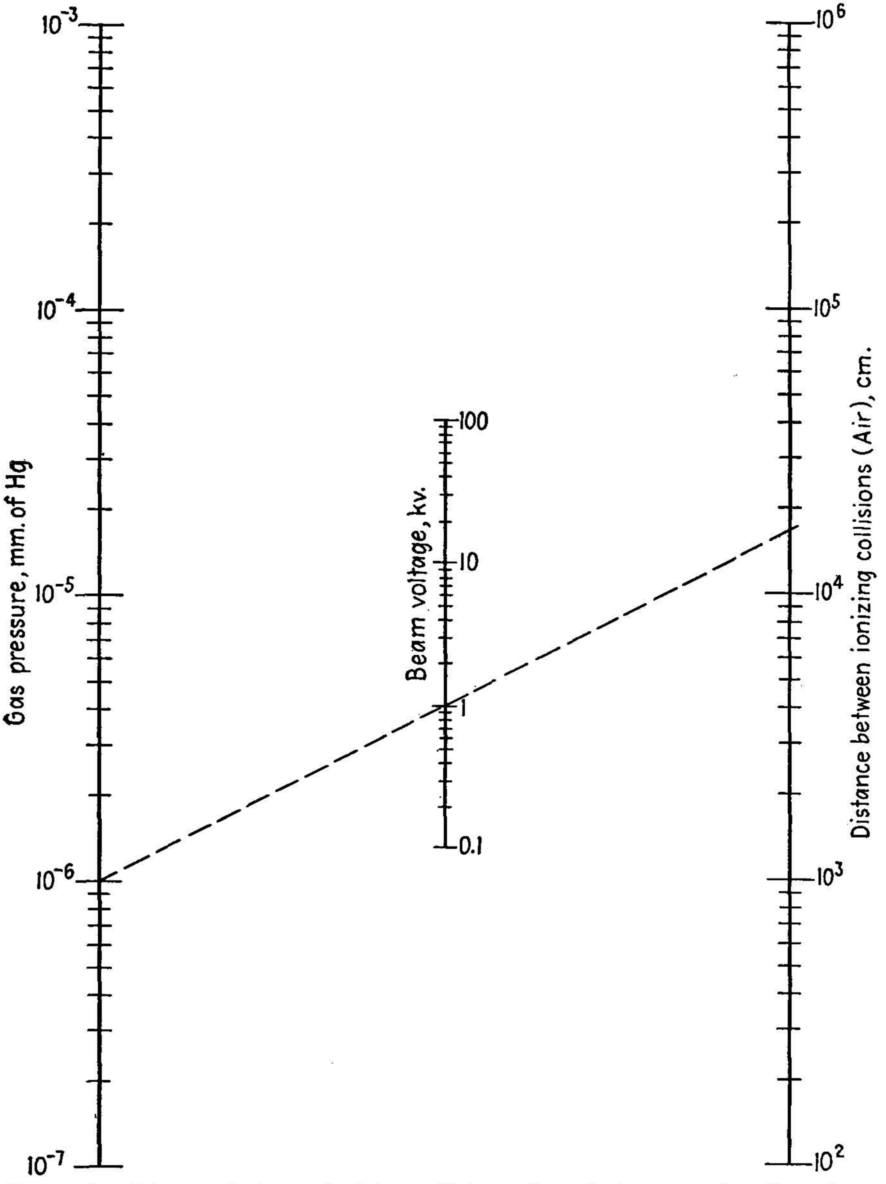


FIG. 21.8.—Distance between ionizing collisions of an electron as a function of gas pressure and voltage.

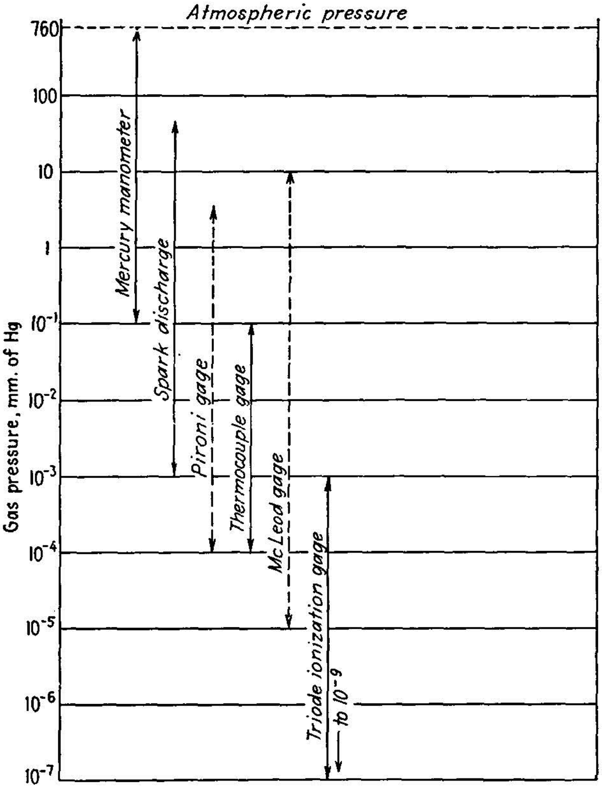


FIG. 21.9.—Range of various pressure-measuring devices.

ruggedness and versatility than the others. These will now be described briefly. In Fig. 21.9 is shown the range of the most commonly used types of vacuum gauges. Where the range line is *solid,* the range of pressures indicated can be covered with a single instrument. Where the range line is *dotted,* several instruments of the same type are required to cover the range indicated.

Manometers. The simplest type of vacuum gauge is the mercury manometer, or U tube, shown in Fig. 21.10. One surface of the mercury column is exposed to atmospheric pressure, and the other is exposed to the low pressure to be measured. The mercury column thus experiences a difference of pressure on the two

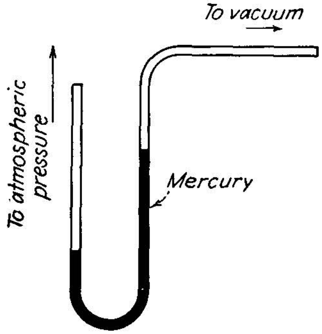


FIG. 21.10.—Mercury manometer.

surfaces and adjusts the height of these surfaces until the forces are in equilibrium around the column. The difference in height of the two surfaces is a measure of the vacuum relative to atmospheric in that atmospheric pressure alone will support a mercury column of height 760 mm. The U-tube manometer is suitable for measuring only relatively poor vacuums of the order of 10^{-1} mm of mercury or less.

The McLeod Gauge. The McLeod gauge is a special type of mercury manometer. It works on the principle of compressing a sample of the gas of which the pressure is to be measured by a known volume ratio and thus increasing the pressure in inverse ratio to an amount which is large enough to measure by direct observation. The McLeod gauge is one of the few gauges that give an absolute pressure indication. Most of the other types of vacuum gauge have to be calibrated against the McLeod gauge, which serves as a standard of measurement.

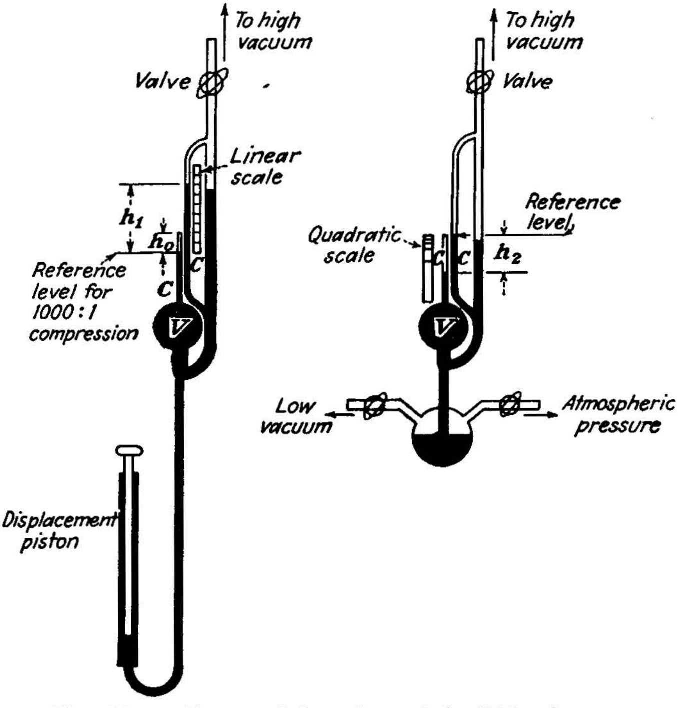


FIG. 21.11.—Long and short form of the McLeod gauge.

The general form of the McLeod gauge and the means by which a sample of gas is trapped and compressed are shown in Fig. 21.11. All McLeod gauges have in common a large volume V in which a sample of gas can be trapped by raising a column of mercury. The volume V has a sealed-off capillary tube C sealed into its top. By raising the mercury until its surface is well up in the capillary the volume of gas trapped can be compressed by a factor of many hundred times. The pressure of the compressed sample of gas can be measured by comparing the height of the mercury column in a parallel capillary of the same diameter—preferably from the same specimen of tubing as the compression capillary—in which the gas is uncompressed. The structure of the McLeod gauge requires that it always be made of glass.

The various McLeod gauges differ only in the means of raising the mercury column and of reading the gauge. In Fig. 21.11 are shown two of the commonest methods of raising the mercury column. The arrangement at the left shows a design, known as the "long form," in which the mercury column is raised by means of a displacement piston. The piston can be arranged with a clamp and screw thread so that fine adjustments of level can be obtained by turning the piston when it is clamped in one position. This arrangement has the advantage that the use of rubber tubing can be avoided since the displacement piston can be made of metal. The structure is of necessity quite high since a difference of elevation of 760 mm must exist between the level of mercury in the gauge and that in the displacement-piston reservoir. This form of the McLeod gauge requires mounting on a vertical rack rising from floor level to a height of about 5 ft. A somewhat shorter vertical height can be achieved with the arrangement at the right, known as the "short form." With this arrangement the mercury level may be raised and lowered by gas pressure through a combination of valves. The mercury is raised by opening the valve to the high vacuum to be measured. Too rapid a rise is offset by partly evacuating the mercury reservoir by opening the valve connected to a source of low vacuum such as a mechanical pump. If the pressure in the mercury reservoir is made too low, the mercury will fall but this can be offset by admitting air through another valve attached to the reservoir. This form of the McLeod gauge requires extremely careful handling. Another form of the McLeod gauge, not illustrated, carries the compression volume, the compression capillary, and the comparison capillary on a framework that can be tilted to achiéve an effective raising or lowering of the mercury relative to the measuring tubes.

The McLeod gauge can be used in two ways as a pressure indicator. In the first method there is determined a point on the compression capillary such that when the mercury is raised to this level the compression volume is compressed by some convenient factor such as 100 or 1,000. The height of the mercury in the free capillary above this reference point will be a *linear function* of the gas pressure. Specifically, the difference in height will be the original pressure in millimeters of mercury times the compression ratio. By using capillaries of equal diameter, surface tension forces are equal in the two tubes. Thus

$$P = h_1 \frac{V_0}{V} \tag{21.18}$$

where P is the pressure of the gas in millimeters of mercury; h_1 is the difference in level between the mercury in the free and in the compression capillaries, in millimeters, when the mercury is raised to a point a distance h_0 from the top of the compression capillary at which point the compressed volume of gas is V_0; and V is the original trapped volume of gas. Since

$$V_0 = \frac{\pi d^2 h_0}{4} \tag{21.19}$$

where d is the diameter of the capillary

$$P = \left(\frac{\pi d^2 h_0}{4V}\right) h_1 \tag{21.20}$$

The other method of using the McLeod gauge raises the mercury level in the free capillary to a height the same as that of the top of the compression capillary each time a measurement is made and reads the difference in height h_2 between the mercury levels in the free and compression capillary. With this method there is a different compression ratio with each gas pressure, but the basic relation of Eq. (21.19) holds in the form

$$P = h_2 \frac{V_2}{V} \tag{21.21}$$

where V_2 is the volume of the gas in the compression capillary to a height h_2. Accordingly,

$$V_2 = \frac{\pi d^2 h_2}{4} \tag{21.22}$$

Hence

$$P = \frac{\pi d^2}{4V} h_2^2 \tag{21.23}$$

and it is seen that the difference in mercury levels by this method is a *quadratic function* of the gas pressure. The gas pressure is conveniently read by attaching a suitable quadratic scale upside down to the compression capillary, lined up so that the zero of the quadratic scale corresponds to the top of the compression capillary. The quadratic and linear methods of measuring pressure may be used with either the long or the short form of the gauge.

The sensitivity of the McLeod gauge may be increased by increasing the size of the compressible volume or by decreasing the diameter of the compression capillary. There are definite limits beyond which neither of these quantities can be carried. If the compression volume is made too large, the weight of the mercury needed to fill it becomes a limitation. The weight of 330 cm^3 of mercury is 10 lb, and the problems involved in making the gauge strong enough to support this weight are considerable. If the compression capillary is made too small, there is trouble with the mercury sticking. This imposes a practical limit of about ½ mm at the smallest diameter capillary. The reference heights used on the calibrated scales can be reduced to increase the sensitivity, but there are limits here, too. The reference heights have to be more than a few diameters of the capillary because of the difficulty of estimating the volume of the rounded end of the capillary.

As a rough guide to the design of McLeod gauges let it be assumed that for linear-scale operation the reference height h_0 is nine times the diameter of the capillary and that the smallest measurable difference in height h_1 is equal to the diameter of the capillary. With these assumptions, Eq. (21.20) reduces to

$$P_m = \frac{7.06d^4}{V} \tag{21.24}$$

where P_m is the minimum pressure that can be read with ease and accuracy where all quantities are expressed in terms of millimeters. Likewise, for quadratic-scale operation let it be assumed that the lowest practical height difference h_2 is three times the diameter of the capillary. Substitution of this value into Eq. (21.23) again yields Eq. (21.24), which may be used as a design equation. A nomographic chart showing the relations between the variables in Eq. (21.24) is given in Fig. 21.12. The sample construction line drawn shows that a sensitivity of 10^{-4} mm of mercury can be realized with a compression volume of 100 cm^3 and a capillary of diameter 1.091 mm.

McLeod gauges are frequently made with two capillaries attached in series on top of the compression volume. The large-diameter capillary is sealed directly to the volume, and the small capillary is connected to the large one and sealed off at its end. Two free parallel capillaries of the same diameter are used. With this arrangement the use of linear scales over a wide range of pressures is facilitated.

The McLeod gauge does not indicate the presence of water vapor, carbon dioxide, ammonia, pump oil vapors, and condensable vapors in general. When used with an oil-diffusion pump a cold trap should be placed between the pump and the gauge; otherwise, the latter will simply indicate the vapor pressure of mercury, which is 2.777×10^{-3} mm at room temperatures. The tendency of the mercury to stick in the compression capillary can be reduced by warming and thus degassing the capillary with a soft flame. Mercury that has stuck can be evaporated by heating. The McLeod gauge is considered reliable to 10^{-4} mm of mercury and is useful to 10^{-5}. Qualitative indications may be had for pressures as low as 10^{-6} mm of mercury.

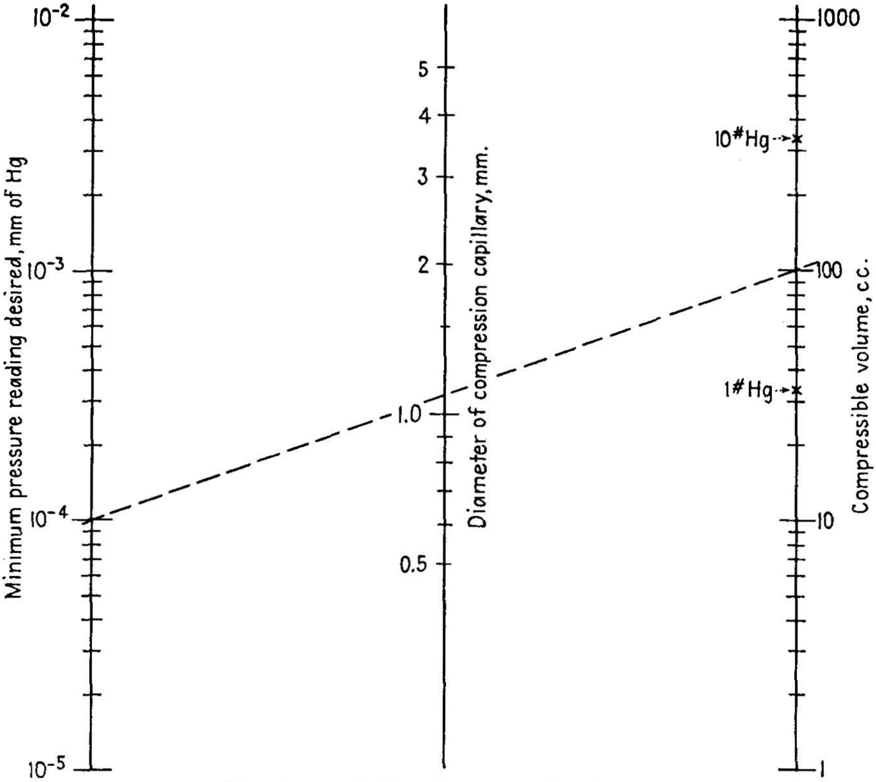


Fig. 21.12.—McLeod-gauge design chart.

The Spark-discharge Tube. A convenient device for monitoring low pressures is a spark-discharge tube about 1 in. in diameter and about 8 in. in length, with disk electrodes supported on tungsten wires sealed through the glass, as shown in Fig. 21.13. The tube has a T joint to the vacuum system and has a d-c potential of about 15 kv applied in series, with a resistance large enough to limit the current to about 2 ma. The series resistance is necessary because the resistance of the discharge tube between electrodes changes greatly with pressure. The nature of the discharge serves as a rather good index of pressure in the range of 50 to 10^{-3} mm. The glow generally has these distinctive parts: Immediately surrounding the cathode, or negative electrode, and assuming its contour is the cathode glow. Beyond the cathode glow is the *Crookes dark space*. Beyond the Crookes dark space is the *negative glow*. Extending from the positive anode, or electrode, is the *positive glow*, which will be continuous or striated depending upon the pressure. The characteristics of the discharge are roughly as follows:

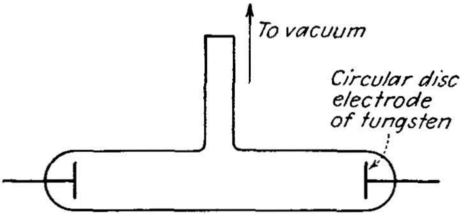


FIG. 21.13.—Structure of the gas-discharge tube.

Mm of Mercury	Appearance
20–50	Narrow streamers
10	Broad streamers
6	Cathode and negative glow forms, positive glow is small tuft at positive electrode
4	Positive glow elongates
0.4	Elongated positive glow breaks into a row of tufts (very pretty)
0.2	Number of tufts in positive glow decreases, and tufts become larger and more widely separated
0.15	Limit of tuft structure of positive glow—two large tufts close to anode
0.10	Negative glow, which has been a small tuft at all higher pressures, elongates. Positive glow is a single tuft
0.05	Negative glow extends nearly to anode, positive glow disappears
0.03	Glow diffuses the whole tube, no definite structure
0.03–0.001	Glow disappears and glass fluoresces from electron bombardment

In addition to the glow discharge changing its structure, it also changes its color. At high pressures the predominating color for air is pink. At lower pressures the pink changes to a blue as the oxygen and nitrogen, which have higher molecular mobility, are removed and carbon dioxide remains. The color of the glass fluorescence depends upon the glass being a yellowish blue in all cases, but more yellow than blue for the soft glasses and more blue than yellow for the hard glasses. The presence of water vapor is indicated by a whitish glow.

In ordinary vacuum setups a spark coil that can be applied to an insulated electrode of the system can be used as a rough pressure indicator at low vacuums. The voltage will set up a glow in the entire system. The nature of the glow is roughly as follows for different pressures:

Mm of Mercury	Appearance
100–40	Bluish-white filamentlike discharge
40–4	Purple filament
4–0.4	Wide, stringy pink discharge
0.4–0.04	Full glow—pink changing to gray to pale gray
0.04–0.004	Discharge disappears, glass fluorescence appears. Predominant color is pale gray. Glass fluorescence disappears at lower limit

A high-frequency Tesla coil can be used instead of a spark coil. This has the advantage that it is safer and will not puncture the glass. Leaks in *glass* can be detected with a spark coil since a spark will tend to jump from the coil electrode to any leak in the glass.

The Pirani Gauge. The Pirani gauge is simply a temperature-sensitive resistance element to which a small amount of power is supplied and which is cooled by the conduction away of energy by molecules of the gas which have collided with it. Thus if the power to the temperature-sensitive resistance element is kept constant, the cooling of the element will be a function of the pressure and will produce a variation in the temperature of the element that can be detected as a change in resistance.

A fine tungsten wire can be used as the temperature-sensitive element. In fact, the filament of a 10-watt light bulb works quite well; a gauge may be made by sealing a piece of glass tubing to such a bulb and attaching it to the vacuum system. The resistance of the filament increases rather rapidly with the power consumed by it. If the filament structure is not coiled but consists of straight wire, it will make a better gauge. In general, the higher the thermal efficiency of the filament as a light-producing element, the lower its effectiveness as a vacuum-measuring gauge, and vice versa. This is because, the higher the thermal efficiency, the less effective the cooling by molecular impact. In operation, the Pirani gauge is conveniently used at a temperature of 100 to 500°C above room temperature, just below the temperature of appreciable radiation. This is the temperature range of greatest sensitivity, for the cooling of the filament is then mostly by conduction rather than by radiation.

In principle, the Pirani gauge can be operated in three fundamental ways.

1. Maintain the voltage across the filament constant, and measure the resistance as a function of pressure.
2. Maintain the current through the filament constant, and measure the resistance as a function of pressure.
3. Maintain the resistance of the filament constant, and measure the power supplied as a function of pressure.

Method 3 has been tested by Campbell, who found that the power required to keep the filament resistance constant was a linear function of pressure, becoming less as the pressure became less. This is in accordance with the expectation that the conduction cooling by molecules of the gas is proportional to the number striking the filament per second, which is proportional to the pressure. The direct proportionality has an upper limit at the pressure at which the mean free path of the molecules is of the order of the bulb dimensions. Below this pressure the heat is conducted by the molecules directly from the filament to the bulb walls. Above this pressure the heat is conducted from layer to layer of gas surrounding the filament; and since the heat conductivity of a gas according to the predictions of the kinetic theory of gases is constant under these conditions, there is no further change in heat loss with pressure. Since the mean free path of nitrogen is about 1 cm at 10^{-2} mm of mercury pressure, this will ordinarily be the upper limit of linearity between power and pressure for a constant filament resistance, though indications can be obtained up to 10^{-1} mm of mercury. It is convenient in obtaining pressure data by method 3 to plot the ratio $\frac{V^2 - V_0^2}{V_0^2}$ as a function of pressure. In this ratio, V is the voltage required to produce a given resistance at a pressure p, and V_0 is the voltage required to produce the same resistance at pressures less than 10^{-4} mm of mercury. The low-pressure voltage V_0 across the filament represents heat loss primarily by metallic conduction, though there will be some by radiation. For pressures less than this value there are so few molecules present that there is virtually no molecular cooling action. Accordingly, the range of the Pirani gauge is about 10^{-1} to 10^{-4} mm of mercury when an ordinary light-bulb filament is used. It is possible to extend the upper limit of pressure with specially designed tubes of small dimensions.

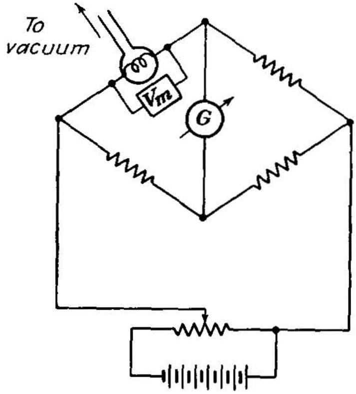


FIG. 21.14.—Constant-resistance method of using the Pirani gauge.

The simplest circuit by which pressure may be measured by the constant resistance method is shown in Fig. 21.14. Filament resistance is determined by the bridge balance as shown by the galvanometer. The bridge is first balanced at very low pressures, and the bridge resistances are then left unchanged. Bridge voltage is chosen to impart a suitable temperature, well below color, to the filament. As pressure rises, the cooling of the filament will increase and lower the temperature, in turr lowering the resistance. Bridge balance is restored by increasing the bridge voltage.

The Pirani gauge can be incorporated into an automatically self-balancing bridge by means of the circuit of Fig. 21.15. The gauge-bridge circuit is in the feedback circuit of an audio amplifier tuned to about 1,000 cycles. The amplifier will oscillate at an output-power level that nearly balances the bridge. Thus, if the amplifier gain between input and output terminals is 100, the output power will rise until the ratio of the input and output voltage of the bridge is $\frac{1}{100}$, which is nearly a condition of balance. Under the conditions stated the gauge resistance will be maintained constant within about 1 per cent, which is close enough for pressure measurements. As pressure rises and the gauge filament resistance tends to fall, the amplifier will supply more power to keep the resistance constant. The power supplied by the amplifier is conveniently indicated by a thermocouple milliammeter, properly shunted, in the output circuit. Since the deflection of a thermocouple meter is proportional to current squared, the deflection is directly proportional to power. Hence the meter can be engraved with a scale that

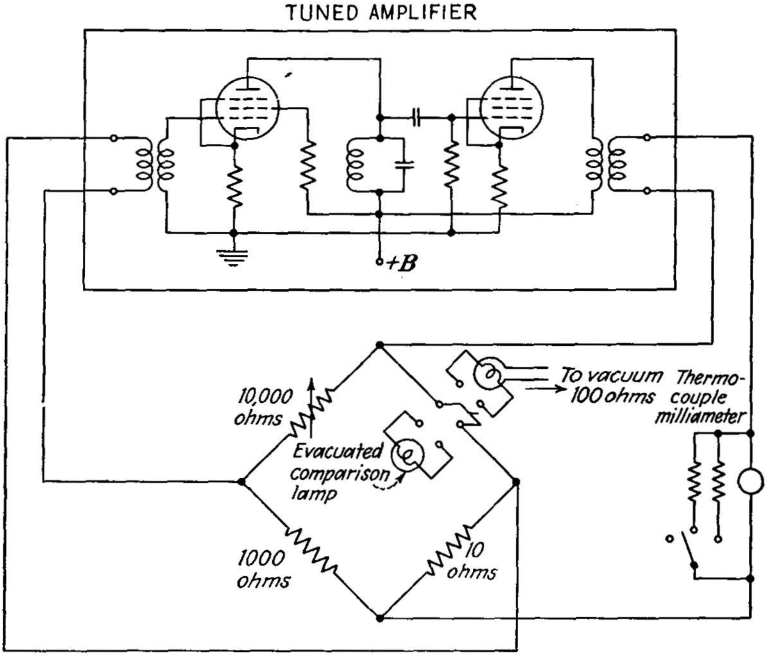


FIG. 21.15.—Automatically self-balancing bridge circuit for use with the Pirani gauge.

will be linear with pressure but have zero pressure slightly upscale. This may be corrected by adjusting the zero setting of the needle so that it is negative by the proper amount. When this is done, the indication of the meter will be linear with pressure. The shunt may be adjusted to give different ranges of pressure. The power supplied by the amplifier to the bridge will divide in constant ratio between the bridge arms. If the ratio of the bridge resistances is as indicated in the figure, the power consumed by the gauge filament will be nearly ten-elevenths of the output power.

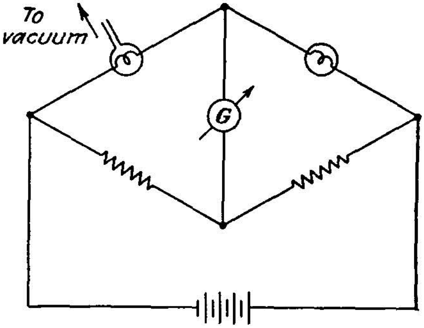


Fig. 21.16.—Pirani gauge with bridge unbalance indicator.

The simplest of all possible methods of using the Pirani gauge consists in putting the filament in a bridge circuit to which a constant direct voltage is applied and calibrating the unbalance current against pressure. The circuit of Fig. 21.16 shows how this may be done. The use of identical filaments one of which is sealed off at high vacuum compensates for external-temperature variations. A typical curve of bridge unbalance current as a function of pressure is shown in Fig. 21.17.

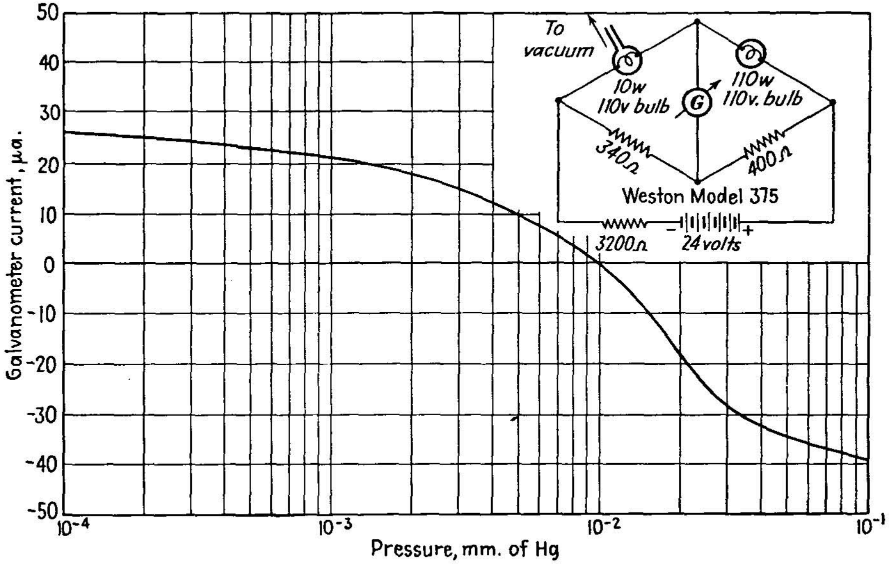


Fig. 21.17.—Bridge unbalance current of a Pirani gauge as a function of pressure.

The Pirani gauge will respond to compressible gases and thus may be used to check the presence of components not revealed by the McLeod gauge. The calibration curves of different gases are slightly different. In general, the more mobile gases will conduct heat away from the filament more readily. In calibrating the Pirani gauge against the McLeod gauge a cold trap should be placed between the gauges to keep mercury vapor out of the former.

The Thermocouple Gauge. The thermocouple gauge works on the same principle as the Pirani gauge. A thermojunction is attached to a heater wire as shown in Fig. 21.18. The heater wire is usually of tungsten and is heated with a current of 10 to 100 ma. The thermojunction can be made of any of the standard combinations such as platinum-platinum rhodium; chromel P–alumel; copper-constantan (advance); iron-constantan (advance); nichrome-constantan (advance). It is connected directly to a sensitive d-c microammeter.[1,2] The cooling of the resistance wire is a function of the pressure, which is recorded by the microammeter, which is activated by the thermal emf generated by the junction on the wire. The range of the thermocouple is from 10^{-1} to 10^{-4} mm of mercury, and it must be calibrated against some standard pressure gauge such as the McLeod gauge.

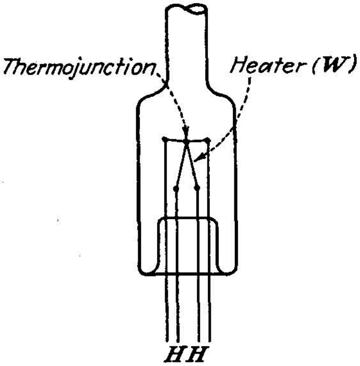


FIG. 21.18.—Thermocouple vacuum gauge.

The specific dimensions of a thermocouple that is excellent for routine pressure indications are as follows: Couple of 3-mil nichrome and 4-mil advance wire, each 1⅜ in. long. Heater of 4-mil platinum wire 2¾ in. long. The wire lengths are long enough to eliminate thermal end effects. Heating current is 150 ma to give a junction current of 200 microamperes in a perfect vacuum.[3]

Triode Ionization Gauge. All the gauges mentioned thus far have been limited in their range to relatively low vacuums. The triode ionization gauge is the most extensively used high-vacuum gauge. The

---

[1] "Handbook of Chemistry and Physics," 26th ed., pp. 1876–1878, Chemical Rubber Co., Cleveland, Ohio, 1942.

[2] WEBER, R. L., "Temperature Measurement and Control," Chap. IV, Blakiston, Philadelphia, 1941.

[3] DUNLAP, F. C., and J. G. TRUMP, Thermocouple Gage for Vacuum Measurements, *Rev. Sci. Instr.*, vol. 8, pp. 37–38, January, 1937.

gauge itself has the form of (or may actually be) an ordinary triode tube sealed into the vacuum system. The most sensitive electrode connection is, however, not the usual triode connection. The grid is operated at a relatively high *positive* voltage, while the plate is operated at a relatively low *negative* voltage. With this arrangement the filament has to be operated at considerably below its normal rating, *i.e.*, the temperature of the filament must be low enough so that the emission is temperature-limited. The basic circuit arrangement involved is shown in Fig. 21.19.

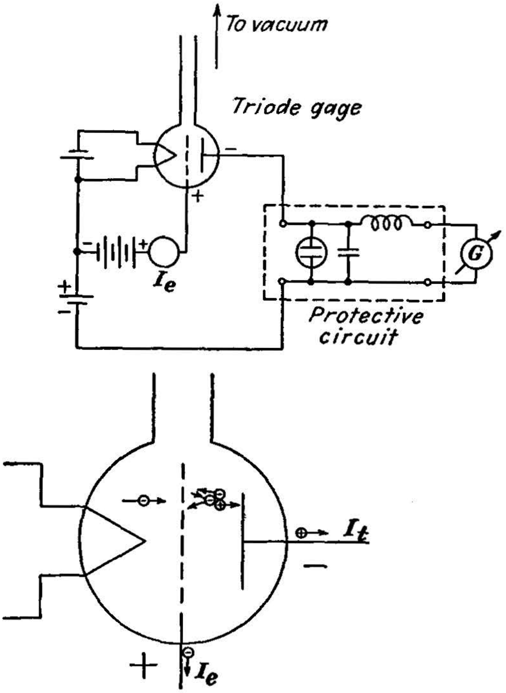


FIG. 21.19.—Basic circuit of the triode ionization gauge.

The function of the positive grid is to attract a stream of electrons into the space between the positive grid and negative plate. In their initial flight from the filament most of the electrons will miss the grid, and their momentum will carry them toward the plate, where the negative potential will repel them and return them to the grid. Some of the electrons will make several oscillations about the grid before they fall into it. While in flight the electrons may ionize some of the gas molecules present, by impact. When this occurs the positive ions created in the grid-plate space will be attracted to the negative plate. The positive-ion current in the plate circuit is therefore a measure of the number of ionizing collisions, which, in turn, is a measure of the pressure. Thus for any emission current I_e the positive-ion current in the plate circuit, I_+, is a linear function of the pressure.

The range of the triode ionization gauge is about 10^{-3} to 10^{-9} mm of mercury. The upper limit of pressure occurs when a glow discharge exists. The lower limit of pressure is fixed by the smallest positive-ion current that can be measured, which in turn depends upon the leakage resistance of the gauge between electrodes. The positive-ion plate current of a type 45 triode used with an emission current of 5 ma, a grid voltage of +120 volts, and a plate voltage of −15 volts is of the order of 3 microamperes at a pressure of 10^{-4} mm of mercury. The positive-ion current is linear with pressure within the range indicated, *i.e.*, the positive-ion current will be 0.3 microampere at 10^{-5} mm of mercury and 0.03 microampere at 10^{-6} mm of mercury for the operating conditions given. Positive-ion current is linear with electron-emission current up to about 20 ma for the type 45. The positive-ion current will also increase with positive grid voltage, but not in a linear fashion. The increase of positive-ion current to the plate with positive grid voltage is most rapid at first and then relatively less rapid. The grid may be operated as high as +200 volts relative to filament. The limit to which the emission current and positive grid voltage can be raised is the dissipation capacity of the grid, which is of the order of 1 watt. For reliable readings both the grid and plate should be degassed by heating to a dull red heat either by r-f induction coil or by direct electron bombardment. To keep the emission down to the level of 5 to 10 ma it is necessary to keep the filament voltage quite low, for the type 45 used as a gauge about 1 volt instead of the rated 2.5. Oxide-coated filaments are fairly satisfactory for triode ionization gauges. They have the advantage that they will not burn out if the vacuum system accidentally springs a leak. On the other hand, the emission is easily poisoned by pressures lower than 10^{-3} mm. When this occurs, it is frequently possible to restore emission by heating the filament to emission temperatures in the presence of a glow discharge at pressures of the order of 3×10^{-2} mm of mercury. The restoration of emission results from positive-ion bombardment of the filament. Some gauges use tungsten or tantalum emitters. These are very rugged but may give rise to false readings at low pressures, for the filament itself will collect the molecules that strike it and so tend to reduce the pressure in the gauge. The characteristics of such a triode ionization gauge are given in Fig. 21.20.

Since the positive-ion current is so small, it is necessary to have either a sensitive galvanometer or a vacuum-tube amplifier. When a galvanometer is used, it is well to protect it against gas bursts or leaks by placing an inductance of ½ henry or so in series with it and shunting the galvanometer with a neon bulb and a condenser of about ½ microfarad. Any pulse of positive-ion current will tend to be by-passed around the meter by the neon tube and condenser. Since high-sensitivity galvanometers are expensive, it is common practice to use some sort of amplifier that will give an indication on a low-sensitivity instrument. One simple arrangement is shown in Fig. 21.21. This circuit makes use of a cathode follower to measure the voltage across large resistors placed in series with the plate of the ion gauge. The resistors are large, ranging from 10 megohms on down in steps of 10. One-tenth microampere through the 10-megohm resistor produces a voltage drop of 1 volt, which the cathode follower triode reproduces almost exactly in its cathode circuit in the form of 1 ma through 1,000 ohms. A zero adjustment of the output meter is provided in the form of a potentiometer. With the plate load resistance of the triode gauge set to zero the potentiometer is adjusted to give zero current in the cathode circuit of the cathode follower tube. Resistance is now switched into the plate circuit of the triode gauge;

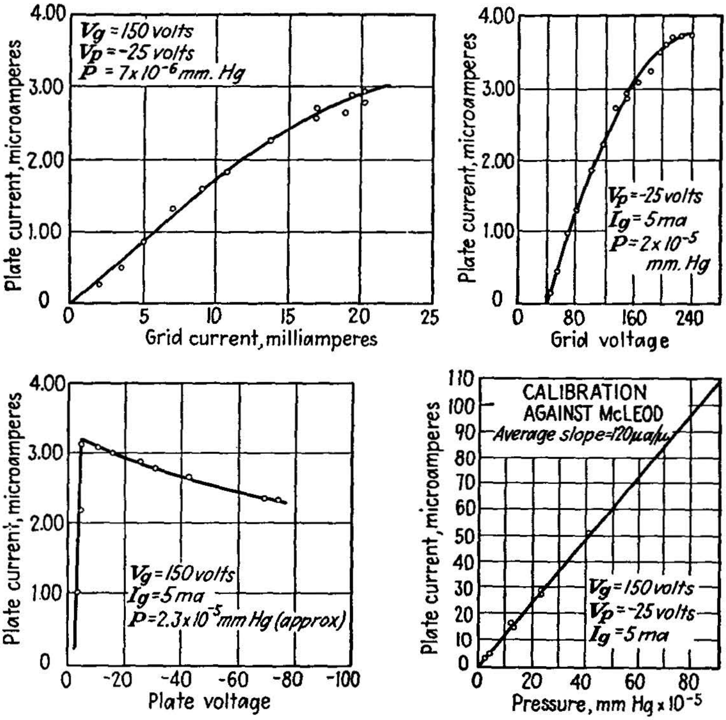


FIG. 21.20.—Characteristics of a typical triode ionization gauge.

and as positive-ion current flows, a positive voltage will appear on the grid of the cathode follower tube and current will flow in its cathode lead. The cathode voltage "follows" the grid voltage almost linearly and so gives suitable indication. Almost any high-mu high-current triode can be used in the cathode follower circuit. The tube characteristics should be such that about 5 ma of plate current will flow when the grid is at cathode potential.

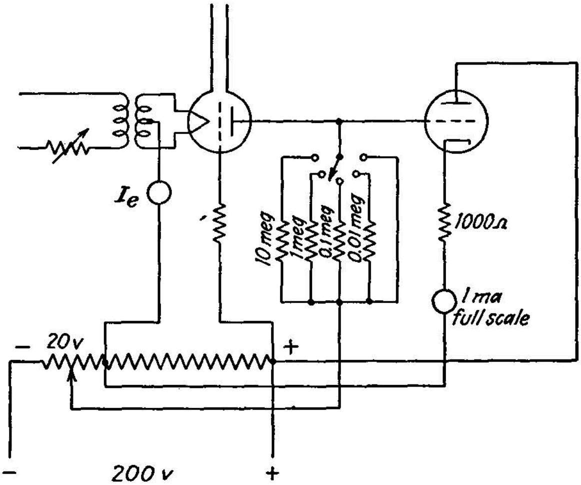


FIG. 21.21.—Cathode-follower metering circuit for use with triode ionization gauge.

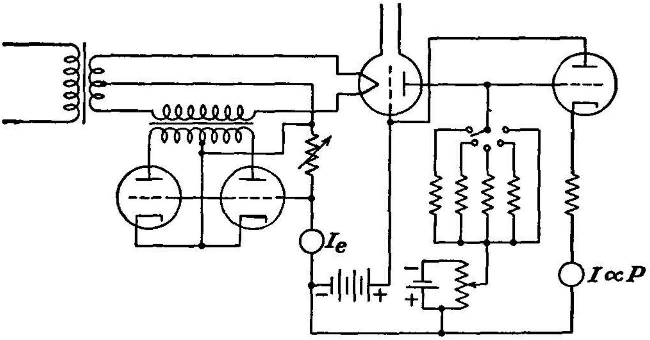


FIG. 21.22.—Circuit for regulating the emission of a triode ionization gauge.

Variations and refinements of the basic circuit shown in Fig. 21.21 are numerous. If extensive vacuum work is done, it is sometimes convenient to have a circuit that will maintain the emission current at a fixed value. Such a circuit is shown in Fig. 21.22. This circuit inserts a variable resistance in series with the filament in the form of the transformed plate resistance of a pair of triodes, one of which conducts for each half of the alternating-voltage cycle. The magnitude of the plate resistance is controlled by a grid voltage derived from a resistor through which the emission current flows. If the emission current tends to increase, the triodes are biased negatively, with the result that their plate resistance and hence the resistance in series with the filament circuit is increased, thus decreasing the filament current and offsetting the increase in emission.[1] It is also possible to construct a circuit using a "magic-eye" electron-ray tube as an indicator and thus save the cost of sensitive meters.[2,3]

21.4. Pumping Speed. Before talking about means of producing low pressures it is well to define the terms in which the characteristics of such devices will be described. In talking about vacuum pumps we are concerned with the laws related to the movement of gases through tubes and orifices.

Speed of an Aperture. Consider the case of a large volume of gas in a chamber closed except for a small aperture opening into a perfect vacuum. Gas will move out of the volume at a rate given by

$$\frac{dV}{dt} = 10.08 \sqrt{\frac{T}{293}\,\frac{29}{M}}\,A \qquad \text{liters per sec} \qquad (21.25)$$

where A is the area of the aperture in square centimeters, T is temperature in degrees absolute, and M is molecular weight of the gas involved (29 for air). It is seen that at room temperature of 20°C the flow through an aperture of area 1 cm^2 is 10.08 liters per sec. At a temperature of 27°C it is 10.2 liters per sec. The volume flow is independent of the pressure! This occurs because, although the number of molecules passing through the aperture is proportional to pressure, the volume of gas corresponding to a given number of molecules is inversely proportional to pressure.

Definition of Pump Speed. By analogy with an aperture the speed of a pump is measured in volume flow, usually in units of liters per second. Pumps are similar to apertures in that their speeds are nearly constant over a wide range of pressure and that their speeds are comparable with those of pump output apertures. Pumps will, however, have a limiting pressure, subsequently referred to as P_0. Accordingly, the

---

[1] RIDENOUR, L. N., and C. W. LAMSON, Thermionic Control of an Ionization Gage, *Rev. Sci. Instr.*, vol. 8, pp. 162–164, May, 1937.

[2] RIDENOUR, L. N., Magic Eye Ionization Gage, *Rev. Sci. Instr.*, vol. 12, pp. 134–136, March, 1941.

[3] PERKINS, W. E., and H. A. HIGGINBOTHAM, An Ionization Gage Circuit, *Rev. Sci. Instr.*, vol. 12, pp. 366–367, July, 1941.

removable volume of gas of any nominal volume in terms of any pressure we wish to use as reference is proportional to the difference between the existing and ultimate pressure,

$$V = k(P - P_0) \tag{21.26}$$

Speed is defined as the volume flow

$$S = \frac{dV}{dt} = k\frac{dP}{dt} = \frac{V}{P - P_0}\frac{dP}{dt} \tag{21.27}$$

Integration of this equation from a pressure P_1 at time zero to pressure P_2 at time t yields

$$t = \frac{V}{S}\ln\left(\frac{P_2 - P_0}{P_1 - P_0}\right) \tag{21.28}$$

The above is useful in estimating the time required to reduce pressure to a given level. When the ultimate pressure is low compared with the other pressures concerned, then Eq. (21.28) reduces to

$$t = \frac{V}{S}\ln\left(\frac{P_2}{P_1}\right) \tag{21.29}$$

Speed of Tubing. At low pressures the flow of gases through tubing is molecular rather than hydrodynamic in nature. The flow involves frequent collisions with the walls and relatively few collisions between molecules. On the assumption that the mean free path of the molecules is large compared with the diameter of the tubing involved, that Lambert's cosine law holds for reflection from any impact with the walls, that the velocity distribution is Maxwellian, and that the number of molecules striking any area is proportional to the pressure, the flow through a piece of tubing has been calculated to be

$$G = \frac{r^3}{l}\frac{1}{\left(1 + \frac{8r}{3l}\right)}\sqrt{\frac{T}{300}\frac{29}{M}} \quad \text{liters per sec} \tag{21.30}$$

where r and l are radius and length of the tubing in millimeters, respectively, T is temperature in degrees absolute, and M is the molecular weight of the gas involved.[1] For air at 27°C the radical has the value of unity. The symbol G is used because the quantity is analogous to electrical conductance. A nomographic chart of tube conductance as a function of radius and length is given in Fig. 21. 23. The above equation is accurate to within a few per cent, provided that the diameter of the tub-

---

[1] KNUDSEN, M., Die Molekularströmung der Gase durch Offnungen und die Effusion, *Ann. Physik*, vol. 28, pp. 999–1016, 1908.

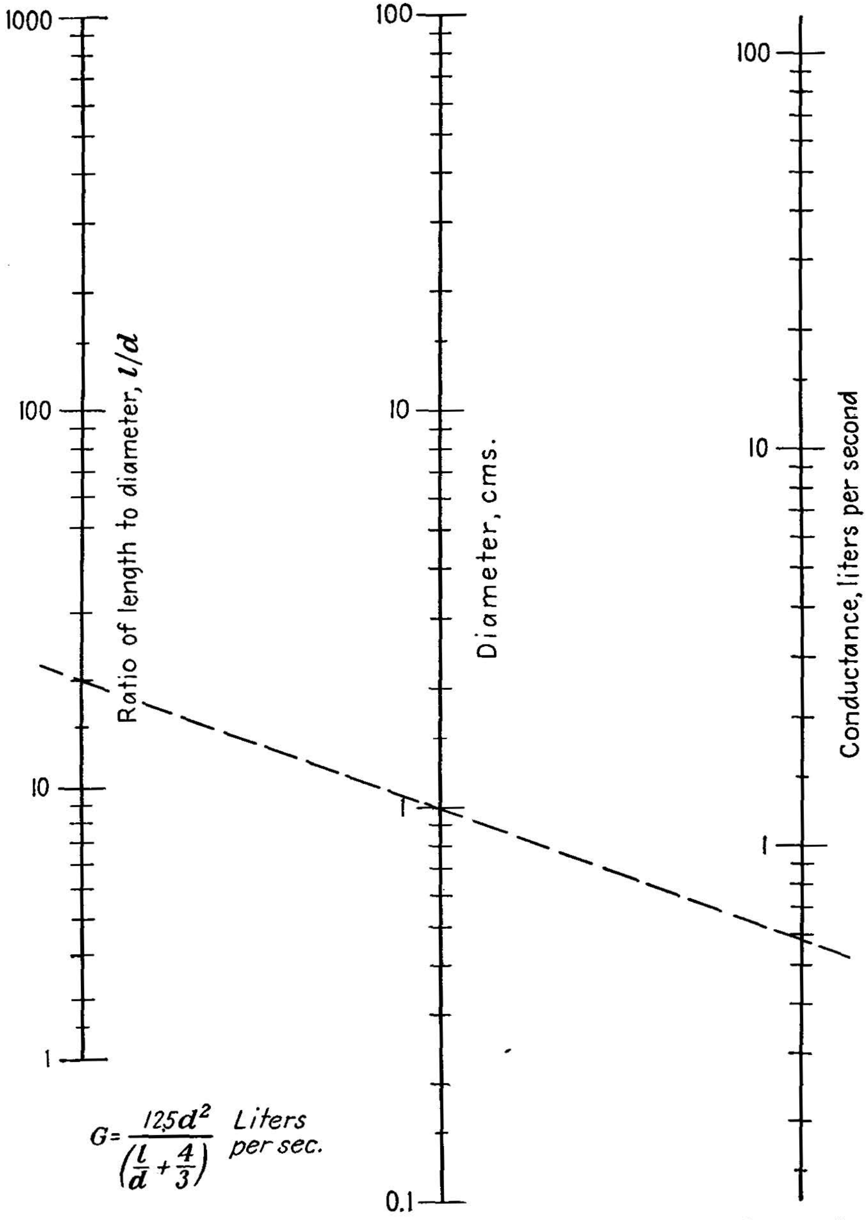


FIG. 21.23.—Nomographic chart of the speed of air flow through tubing at low pressures.

ing is less than the mean free path of the gas molecules. The quantity in parentheses in Eq. (21.30) is a correction factor for shortness of tubing. When the tubing is long compared with the radius, this quantity approaches unity very closely. *Hence for long tubing containing air at room temperature the conductance in liters per second is given approximately by the radius in millimeters cubed divided by the length in millimeters.* The importance of using large-diameter tubing is evident from the dependence of the conductance upon the cube of the diameter. Reduction of diameter by a factor of 3 reduces conductance by a factor of 27. The speed of flow of gases in a vacuum system cannot be greater than that given by the lowest tubing conductance in the system.

When a number of pieces of tubing are connected in series, the reciprocal of the resultant conductance is equal to the sum of the reciprocals of the individual conductances.

$$\frac{1}{G_{eq}} = \frac{1}{G_1} + \frac{1}{G_2} + \frac{1}{G_3} + \cdots + \frac{1}{G_n} \tag{21.31}$$

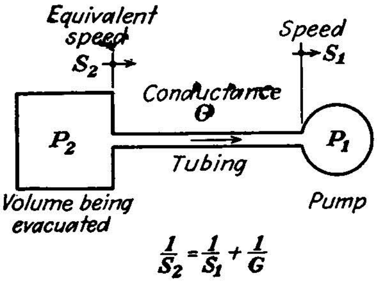


FIG. 21 24.—Diagram illustrating equivalent pumping speed.

where G_{eq} is the equivalent conductance and $G_1, G_2 \cdots G_n$ are the conductances of the different portions of tubing. If the resistance of a long piece of tubing is defined as the reciprocal of the conductance, then, for air at 27°C,

$$R = \frac{1}{G} = \frac{l}{r^3} \qquad \text{sec per liter} \tag{21.32}$$

where l and r are the length and radius of the tubing in millimeters, respectively. Then the resultant resistance is simply the sum of the individual resistances,

$$R_{eq} = R_1 + R_2 + \cdots + R_n \tag{21.33}$$

Only the length and diameter of a tubing are of importance in calculating the gas flow. Bends and corners have little effect. It must always be remembered that such computations as are indicated above are restricted to the range where the diameter of the tubing is less than the mean free path of the gas molecules.

Effect of Tubing upon Pumping Speed. The speed of a pump has the same units of conductance, *i.e* , liters per second. Hence a pump may be considered as a piece of tubing, of conductance equal to its speed, feeding into an infinite reservoir of gas at the ultimate pressure of the pump.

The reciprocal of the equivalent pumping speed of a pump with associated tubing is found by adding the reciprocal speed of the pump and the reciprocal conductance of the tubing. Thus for the arrangement of Fig. 21.24, where the speed of the pump is S_1 and the conductance of the tubing is G, the equivalent speed at the volume being evacuated, S_2, is given by

$$\frac{1}{S_2} = \frac{1}{G} + \frac{1}{S_1}^{*} \tag{21.34}$$

It is seen that the resultant speed is lower than both the tubing and the pump speed. The importance of using large-diameter high-speed tubing is again evident. A nomographic chart of Eq. (21.34) giving the resultant speed of a pump and tube in terms of the pump speed and tube conductance is shown in Fig. 21.25.

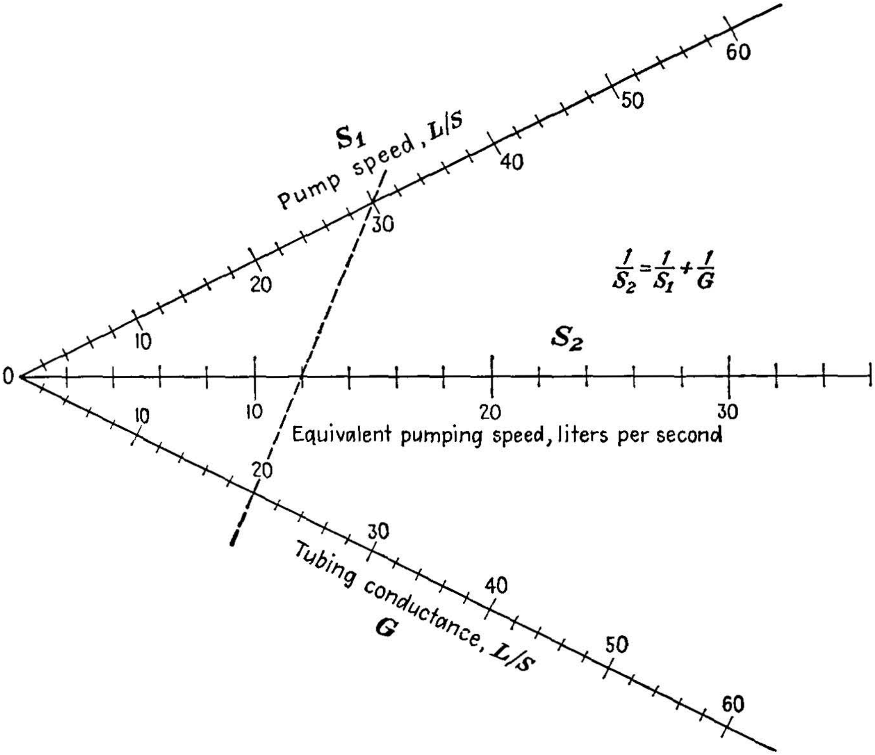


FIG. 21.25.—Nomographic chart of equivalent pump speed.

---

* Proof of Eq. (21.34) may be found by equating the mass of flow at different points in the system as $Q = G(P_2 - P_1) = S_1P_1 = S_2P_2$. When pressures are eliminated from these relations, Eq. (21.34) for the equivalent speed results.

21.5. Production of Low Vacuum. Numerous applications require the production of vacuums of 10^{-4} mm of mercury or less. In addition, high-vacuum pumps will not operate if required to exhaust directly into air but must exhaust into a low vacuum to be efficient. As a result, the subject of production of low vacuum falls into a class by itself.

Low vacuums are most easily obtained by means of a mechanical pump. Numerous designs for such pumps have been suggested, but the successful pumps that are used in large quantities all embody the same principle. An example of a widely used mechanical pump is the Cenco Hyvac. A diagram of the internal structure of this pump is given in Fig. 21.26. Essential features of this pump are an eccentric rotor *A*, a valve *K*, which divides the space between the rotor and stator into two volumes, and an output valve *L*. As the rotor turns in the direction indicated in Fig. 21.26, there is presented to the space being evacuated a volume *H*, which expands, allowing gas to enter, and is then sealed off by the rotor surface. The trapped volume of gas is then compressed

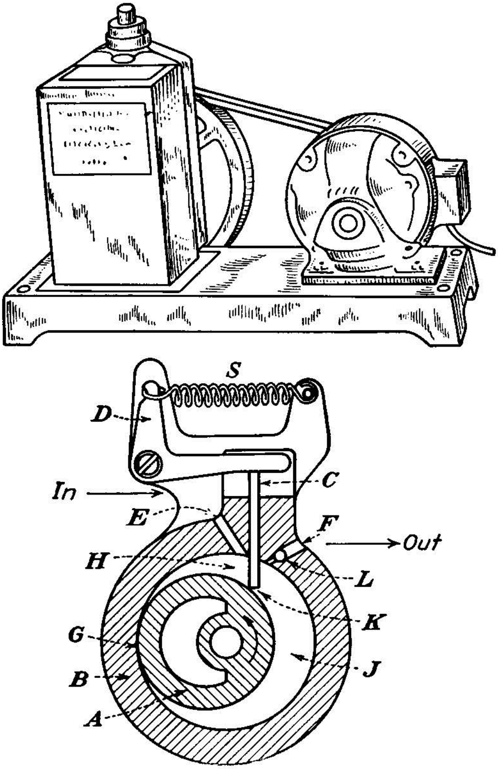


FIG. 21.26.—Diagram and picture of Cenco Hyvac pump.

against the output valve *L* through which it is expelled. By ganging such rotors and running up the speed, very good evacuation properties may be had. In Fig. 21.27 are shown the speed-pressure curves of some well-known mechanical pumps. Ultimate pressures are of the order of 10^{-3} to 10^{-5} mm of mercury, the average being 10^{-4} mm of mercury. Speeds of mechanical pumps are usually from three- to five-tenths of the speed of the input aperture except in the immediate vicinity of the ultimate pressure, where the speed is much lower. Limiting pressures are determined largely by the excellence of the mechanical tolerance in the rotor and valve construction. Pumps are usually immersed in oil to improve the valve action.

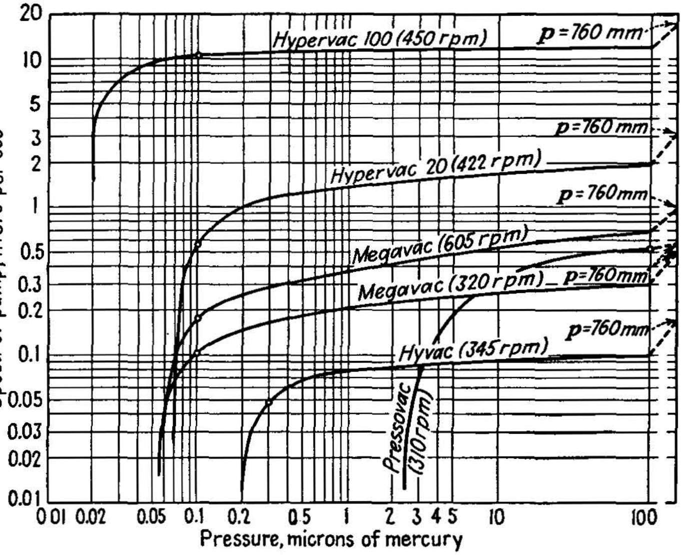


FIG. 21.27.—Speed-pressure curves of mechanical vacuum pumps.

21.6. Production of High Vacuum. For the production of pressures lower than those which can be obtained with mechanical pumps, vapor-diffusion pumps are invariably used.[1] Roughly speaking, the diffusion

---

[1] Some information on this subject is given in the general references cited earlier (p. 747). For more recent information see HICKMAN, K. C. D., and C. R. SANFORD, A Study of Condensation Pumps, *Rev. Sci. Instr.*, vol. 1, pp. 140–163, March, 1930; Ho, T. L., Multiple Nozzle Diffusion Pumps, *Rev. Sci. Instr.*, vol. 3, pp. 133–135, March, 1932; Ho, T. L., Speed, Speed Factor and Power Input of Different Designs of Diffusion Pumps, and Remarks on the Measurement of Speed, *Physics*, vol. 2,, pp. 386–395, May, 1932. See also the excellent summary given in STRONG, J., and others, "Procedures in Experimental Physics," pp. 111–124, Prentice-Hall, New York, 1941.

pump works on the principle of creating a gas flow of a condensable vapor that draws along with it all the molecules from the system being evacuated that get into the flow. The vapor wind so formed is terminated by condensation, and the liquid so formed is returned to a vaporizing unit, where it is again used to form part of a vapor flow. The operation of this principle is best seen by considering specific vapor-diffusion pumps.

The Mercury-diffusion Pump.—One commonly used type of vapor-diffusion pump makes use of mercury as the circulating vapor. A diagram of an early metal pump used by Langmuir is shown in Fig. 21.28. Mercury is heated in the bottom of the unit at D, and the resultant vapor rises up the chimney F, where it is deflected downward by the umbrella-shaped cup E placed over the end of the chimney. The mercury vapor moves down between the outside of the chimney and the outer wall of the pump A, which is cooled by a water jacket J. The principle of counterflow in cooling is purposely avoided, for the back vapor pressure will be least if the condenser temperature is lowest at the high-vacuum end of the pump. As the mercury vapor moves down, it is cooled to the point of condensation and then runs back down into the reservoir at the bottom, where it is again vaporized and recirculated. The probability that any gas molecule that gets into the mercury-vapor stream will experience a collision that will force it to move in the direction of the exhaust is extremely strong. As long as the input pressure exceeds the exhaust pressure by a factor of 100, the forces driving molecules toward the exhaust will predominate over those acting in the opposite direction.

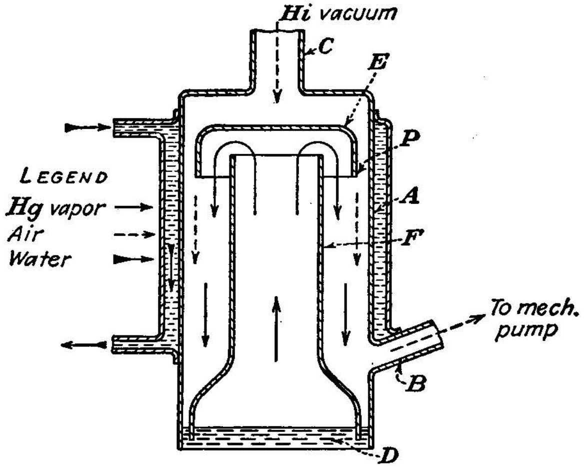


FIG. 21.28.—Langmuir's mercury-diffusion pump.

Diffusion pumps of the type described require a low output pressure to give an effective condensation action. As a result, vapor-diffusion pumps are always operated into a mechanical pump. Diffusion pumps will operate into exhaust pressures, or "fore pressures," as high as 10^{-1} mm of mercury but in general should not be operated into fore pressures of more than 10^{-2} mm of mercury for any length of time. This is about the pressure at which a spark coil will fail to produce a discharge through a gas, and therefore in practice the diffusion pump is not turned on until the mechanical pump has reduced the pressure to the point where a glow discharge can no longer be observed upon application of a spark coil.

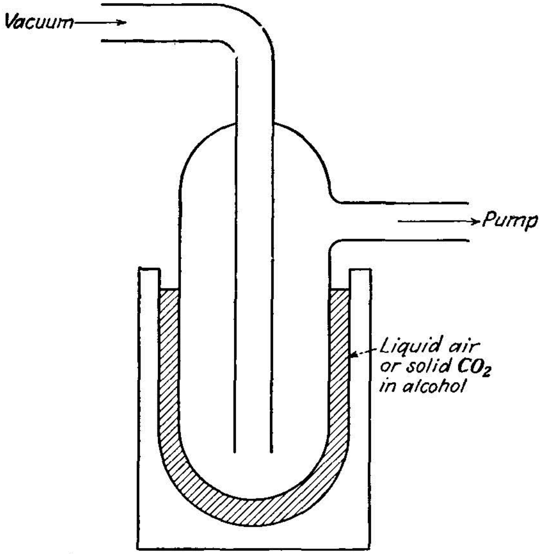


FIG. 21.29.—Freezing trap for collecting condensable vapors.

When mercury is used as the pump vapor, it is necessary to place a freezing trap between the pump and the system being evacuated to catch such mercury molecules as diffuse out from the pump. If this is not done, the minimum pressure that can be obtained with the system is the vapor pressure of mercury, which is about 10^{-3} mm of mercury at room temperature. A commonly used type of cold trap is shown in Fig. 21.29. Cold traps may be cooled with liquid air or with a slush formed by adding alcohol to carbon dioxide snow. The temperature of the trap must be such that the vapor pressure of the mercury is reduced to a value below a level corresponding to the lowest pressure desired. A curve of the vapor pressure of mercury as a function of temperature is given in Fig. 21.30.

With vapor pumps, in general, it is necessary to use two pumps in series to get a very good vacuum. This is because there is a maximum ratio of input and output pressures of about 100 that can be achieved by a single stage.

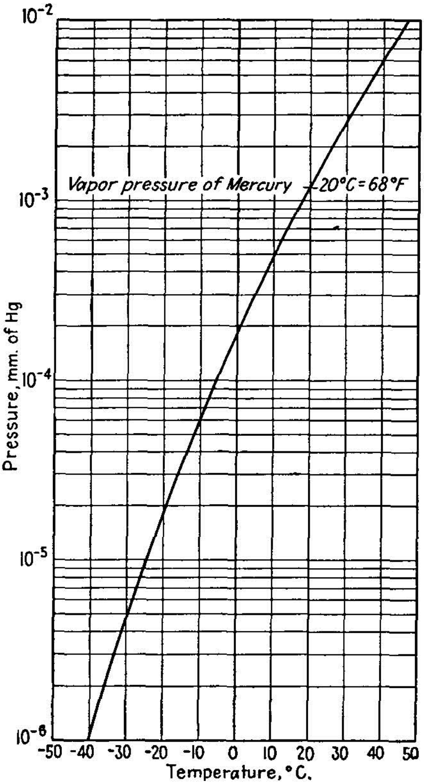


FIG. 21.30.—Vapor pressure of mercury as a function of temperature.

Oil Pumps. Pumps using oil as a diffusion-pump liquid have become more popular than mercury-vapor pumps as oils were developed that had successively lower and lower vapor pressures.[1–4] The various oils now in use and their corresponding vapor pressures are shown in Fig. 21.31. Other properties of the principal oils are given in Table XIV. Oils have the advantage over mercury vapor that a freezing trap is not needed. Furthermore, the speed factors of pumps using oils will tend to be about ten times as great as of those using mercury vapor. Relative to the equivalent aperture oil-vapor pumps are about 50 per cent effective. In oil pumps a baffle or charcoal trap must be used to prevent oil vapor from diffusing into the chamber being evacuated, with some resulting reduction in speed. Care must also be taken not to

---

[1] BURCH, C. R., Oils, Greases and High Vacua, *Nature*, vol. 122, p. 729, Nov. 10, 1928.

[2] VON BRANDENSTEIN, M., and H. KLUMP, Ueber die Verwendun organischer Substanzen in der Hochvakuumtechnik, inbesondere bei dem Betrieb von Hochvakuum Pumpen, *Physik. Zeit.*, vol. 33, pp. 88–93, Jan. 15, 1932.

[3] KLUMB, H., and H. D. GLIMM, Ueber die Sauggeschwindigkeit von Diffusionpumpen die mit organischen Substanzen betrieben werden, *Physik. Zeit.*, vol. 34, pp. 64–65, Jan. 15, 1933.

[4] HICKMAN, K. C. D., Vacuum Pumps and Pump Oils, *Jour. Franklin Inst.*, vol. 221, Part I, Some Fractionating Pumps, pp. 215–235, February, 1936; Part II, A Comparison of Oils, pp. 383–402, March, 1936.

expose the oil while hot to air at pressures greater than 10^{-2} mm of mercury. Likewise, the oil must not be overheated even at extremely low pressure, for then decomposition will be accelerated. All the oils listed in Table XIV are formed by fractional distillation, ending with the component having the lowest product of vapor pressure and rate of chemical breakdown.

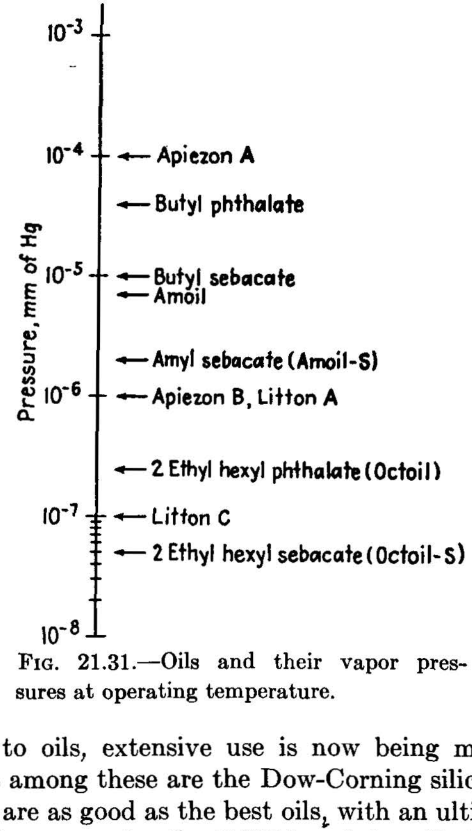


FIG. 21.31.—Oils and their vapor pressures at operating temperature.

In addition to oils, extensive use is now being made of silicones. Most prominent among these are the Dow-Corning silicones DC702 and DC703. These are as good as the best oils, with an ultimate pressure of 5×10^{-7} mm of mercury for the DC703 and 1×10^{-6} mm of mercury for the DC702. The great advantage of the silicones over oils is their *resistance to oxidation;* the silicones do not burn. However, there is a loss of ultimate vacuum due to absorbed gases, but the recovery time to ultimate vacuum is about the same as for the best oils. The disadvantage of the silicones lies in their present high cost. They also exhibit the same undesirable back diffusion as the high-grade oils. Poisoning of oxide cathodes is about the same as for oils without traps or baffles. Both silicones and the highly refined natural oils from which sulphur

TABLE XIV
CHARACTERISTICS OF PUMP OILS*

Name	Formula	Mol. wt.	Sp. gr. at 25°C	Refractive index† N, at 40°C	Pour pt., °F	Ult. vacuum‡ at 25°C	Boiling pt., °C, at 10^{-2} mm	Conductivity G, mhos, at 27°C	Viscosity, sec§			Temp.,‖ °C
									80°F	100°F	130°F	
Butyl phthalate	$C_6H_4(COOC_4H_9)_2$	278.1	1.0465	1.4851	−96	4×10^{-5}	80	1.7×10^{-10}	79	58	45	94
Butyl sebacate	$C_8H_{16}(COOC_4H_9)_2$	314.3	0.933	1.4362	4	2×10^{-5}	90	1.6×10^{-11}	65	46	43	103
Amoil	$C_6H_4(COOC_5H_{11})_2$	306.2	1.0190	1.4802	−61	7×10^{-6}	100	1.7×10^{-11}	100	81	52	96
Amoil-S	$C_8H_{16}(COOC_5H_{11})_2$	343.3	0.9251	1.4395	24	2×10^{-6}	111	2.9×10^{-13}	71	65	50	118
Octoil	$C_6H_4(COOC_8H_{17})_2$	390.3	0.9796	1.4795	−61	2×10^{-7}	122	2.6×10^{-10}	178	171	75	125
Octoil-S	$C_8H_{16}(COOC_8H_{17})_2$	426.3	0.9103	1.4440	−69	5×10^{-8}	143	6.7×10^{-12}	83	81	57	140

* From data supplied by Distillation Products, Inc.

† For white light.

‡ With fractionating pump.

§ Saybolt viscosimeter (approximate).

‖ Maximum safe temperature for exposure to atmospheric pressure.

compounds have been removed have a chemical stability higher than the synthetic oils, and fractionation is less important.[1]

A typical diffusion pump using oil as the liquid is shown in Fig. 21.32. The arrangement of the parts is evident. From the boiler, vapor rises into the nozzle, where it is blown down a water-cooled section of tubing, at the bottom of which the condensed oil is collected and returned to the boiler. In general, the action of oil pumps is more positive when water-cooled than when not, though air-cooled pumps are quite common. Two such pumps in series using a good oil can achieve a pumping speed of 30 liters per sec at 10^{-5} mm of mercury and have an ultimate pressure of better than 10^{-7} mm of mercury if a charcoal trap is used. The fore pressure required is generally of the order of 10^{-3} mm of mercury for positive action. The charcoal trap serves to collect molecules of oil vapor that tend to stray into the chamber being evacuated.[2] One common form of charcoal trap is shown in Fig. 21.33. This trap consists simply of a pan of charcoal powder located so that no oil-vapor molecules may move directly into the chamber being evacuated without coming in contact

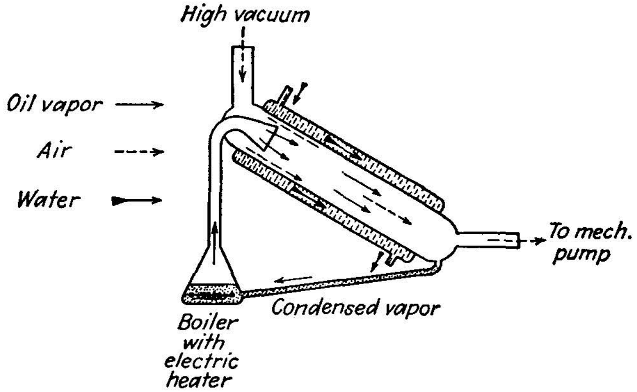


FIG. 21.32.—Typical glass, water-cooled, oil-diffusion pump.

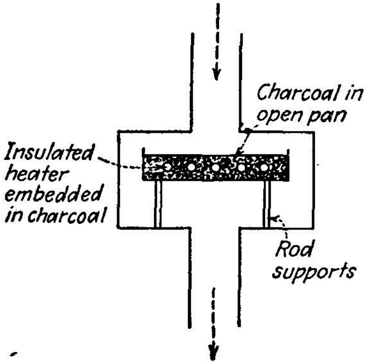


FIG. 21.33.—Charcoal trap with electric heater.

---

[1] The above information on silicones was privately communicated to the author by C. V. Litton.

[2] BECKER, J. A., and E. N. JAYCOX, A New High Vacuum System, *Rev. Sci. Instr.*, vol. 2, pp. 773–784, December, 1931.

with the charcoal. Oil molecules will stick to the charcoal. Provision must be made for degassing the charcoal. This usually takes the form of an electric heater embedded in the charcoal. When the charcoal is heated to a temperature of several hundred degrees, the absorbed oil molecules are decomposed into gases that may be removed by the pump.

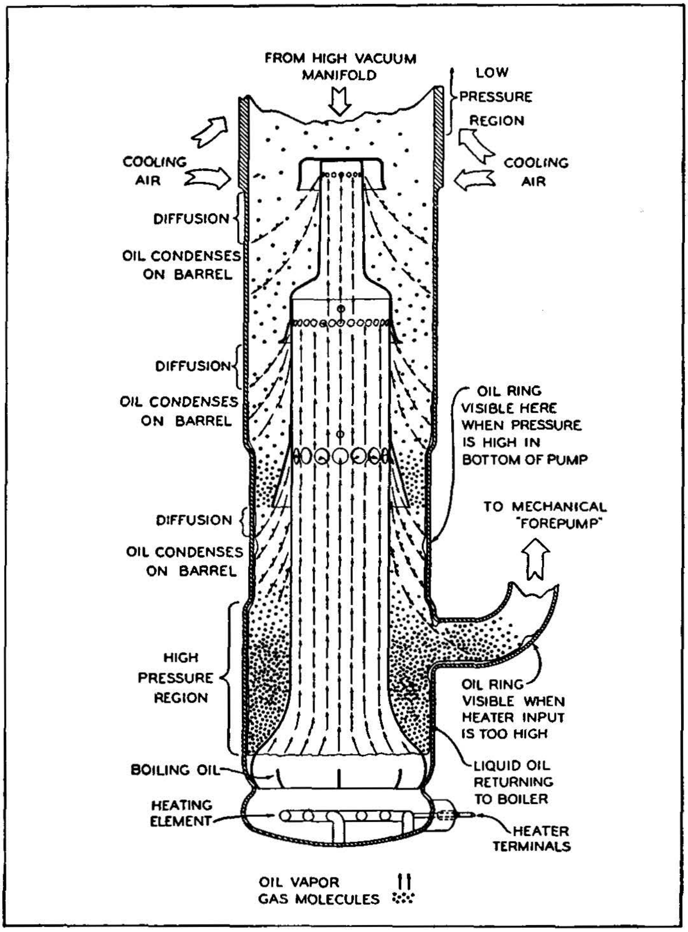
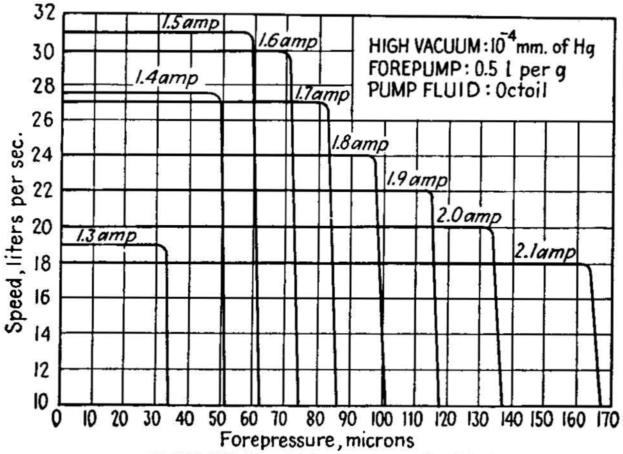


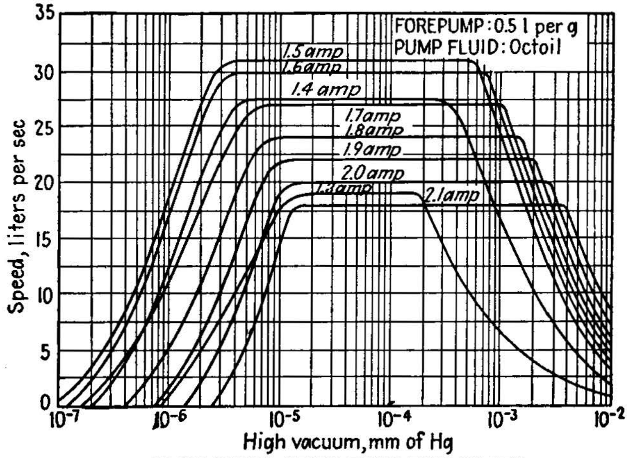
FIG. 21.34.—Air-cooled oil-vapor pump. (Eimac HV1.)

A charcoal trap can absorb several thousand times its own volume of oil vapor. Ultimate pressures of 10^{-8} mm of mercury have been recorded with charcoal traps and a good pump oil. Other means of keeping oil vapors out of the vacuum system are baffles of some metal, such as aluminum, that will not react with the oil and yet that has a good heat conductivity so that oil vapor will condense on it.

A popular form of air-cooled oil-vapor pump is shown schematically in Fig. 21.34. This pump achieves a three-stage action by suitable bleeding of vapor from a chimney over a single boiler. Throat areas and vapor speeds are adjusted to give maximum effectiveness at the different pressures encountered in the system. Operating characteristics typical of such pumps are shown in Fig. 21.35. Pumping speeds in the range of



GLASS THREE-STAGE PUMP, TYPE GF-25A
Speed vs Forepressure



GLASS THREE-STAGE PUMP, TYPE GF-25A
Speed vs. High vacuum

FIG. 21.35.—Operating characteristics of a triple-jet air-cooled oil-vapor pump.

20 to 30 liters per sec at 10^{-4} mm of mercury and ultimate pressures of 10^{-5} to 10^{-6} mm of mercury may be obtained with this type of pump. Pumping speed drops quite sharply if the fore pressure becomes too low. The operating range of such a pump may be shifted toward low pressure by decreasing heater power. In some pumps of this type the output tubing leading to the mechanical pump contains a series of trapping ridges or alembics that prevent substances of high volatility from returning to the pumping-fluid reservoir.

Fractionating Pumps. In order to obtain extremely low pressures it is necessary that the pump oils be uncontaminated with materials of lower vapor pressure. One means of ensuring this is to use a type of pump incorporating a fractional distillation still that continuously refines the oils used.[1] A diagram of a two-stage fractionating pump is given in Fig. 21.36. In operation the alembics in the output chimney collect the extreme volatiles, which if left in the system cause turbulence in the vapor flow. The boiler *A* at the low-pressure end of the pump operates at the highest temperature and utilizes mainly the more volatile low-vapor-pressure components of the oil. Less volatile components flow through the connecting tube to the middle boiler *B*, where they are more effective at the lower pressure because of their lower vapor pressure. The third boiler *C* serves to collect relatively nonvolatile residue and redistill the volatile components back into the other two boilers. Operating characteristics of a three-stage fractionating pump are shown in Fig. 21.37. Ultimate pressures of 10^{-9} mm of mercury may be obtained with pumps of this design, though pumping speeds are only of the order

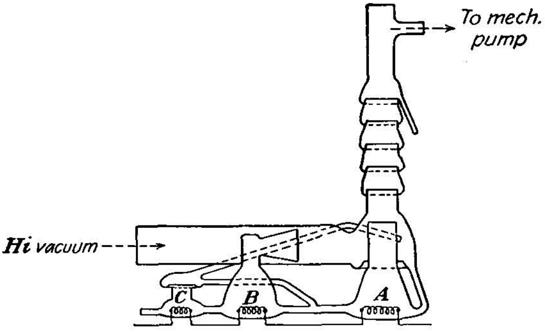
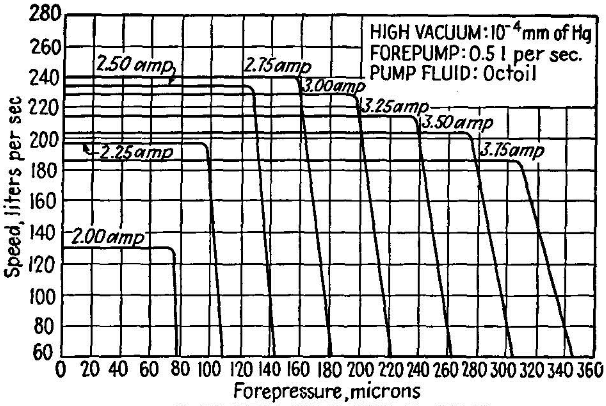


FIG. 21.36.—Diagram of two-stage fractionating pump.

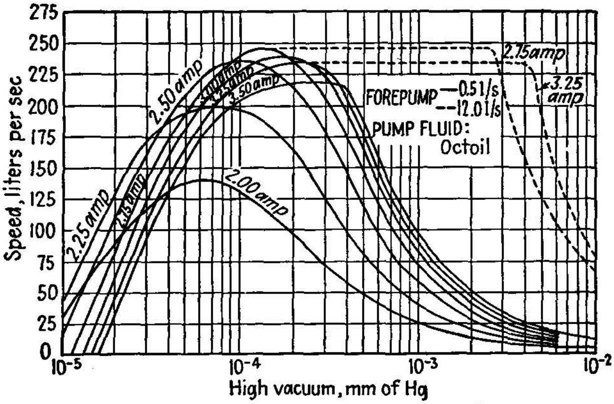
---

[1] HICKMAN, K. C. D., Trends in the Design of Fractionating Pumps, *Jour. Appl. Phys.*, vol. 11, pp. 303–313, May, 1940.

of 30 liters per sec. The operation of this type of pump is quite critical with respect to temperature.



GLASS-METAL PUMP, TYPE GM-220 AB



GLASS-METAL PUMP, TYPE GM-220 AB

FIG. 21.37.—Operating characteristics of a three-stage fractionating pump.

21.7. Glass and Its Properties. Almost every vacuum system or vacuum tube contains some glass in it. Early systems and tubes were entirely of glass, though the trend at present is to use more metal and less glass. Nevertheless, glass is still an indispensable item in vacuum-tube research and construction.

The usefulness of glass is derived from its excellent working characteristics. It can be shaped or molded into almost any form. The varieties of glass which are available are so numerous that a glass can be found suitable for almost any purpose. The greatest disadvantage of glass is the ease with which it breaks, but even this can be minimized with proper design.

Composition of Glass. Glass is a fused mixture of silica, SiO_2, and various metallic oxides. The silica is the predominant component, being from 60 to 80 per cent of the total weight. The characteristics of the glasses are determined by the percentages of the metallic oxides. Pyrex glass is made up of 81 per cent SiO_2, 12 per cent B_2O_3, 4 per cent Na_2O, and a few per cent of other oxides. Lead glass is made up of 61.5 per cent SiO_2, 23 per cent PbO, and sodium and potassium oxide. These two glasses lie near the extremes of a scale of glasses, pyrex being at the so-called "hard" end and lead glass being at the so-called "soft" end. Other glasses lie between these extremes both in composition and in physical characteristics. Of considerable interest in transmitting-tube manufacture is nonex glass, which is a hard glass but not as hard as pyrex. Nonex glass contains 73 per cent SiO_2, 16.5 per cent B_2O_3, and 6 per cent PbO. Nonex is not inactive enough chemically to make it useful for chemical glassware though it is extensively used in transmitting-tube manufacture.

Physical Properties of Glass. Loosely speaking, any material that is hard, brittle, and transparent is referred to as a glass. More properly, glass is an amorphous material that is hard and transparent at room temperatures. As it is heated it softens gradually, becoming softer and softer. Because of this gradual change, it has no definite melting temperature. The transition from a solid to a viscous state is usually defined in terms of the following arbitrary reference temperatures:

Strain point. An arbitrary point on the temperature-viscosity curve, representing a viscosity of $10^{14.6}$ poises[1] where rapid cooling will not produce permanent strain.

Anneal point. Arbitrary point; viscosity $10^{13.4}$ poises, corresponding to relief of strain in ¼ in. plate in 15 min.

Softening point. Arbitrary point; viscosity $10^{7.65}$ poises, corresponding to unit elongation of glass rod in given time interval.

Working temperature. Arbitrary point; viscosity 10^4 poises. Close to maximum temperatures for glassworking—in general, higher than temperatures used for metal seals by 150 to 200°C.

The transition between the various physical states of glass is shown in Fig. 21.38. The temperature scale will be different for each kind of

---

[1] The poise unit of viscosity is the force in dynes required to impart a relative velocity of 1 cm per sec to two parallel surfaces each having an area of 1 cm^2 and spaced 1 cm apart with the viscous material between them. The viscosity of pitch at 15°C is 10^{10} poises. The viscosity of castor oil at room temperature is 2 poises.

glass, but the general characteristics of the curve will be the same. The softening temperatures of the various glasses depend upon the composition, being higher for the higher percentages of silica. Glass must be worked above the softening point. If the glass is maintained at a temperature near the softening point too long, it will be devitrified and possibly oxidized, with the result that its physical characteristics will be impaired. The critical temperatures for the commonest types of glasses are listed in Table XV.

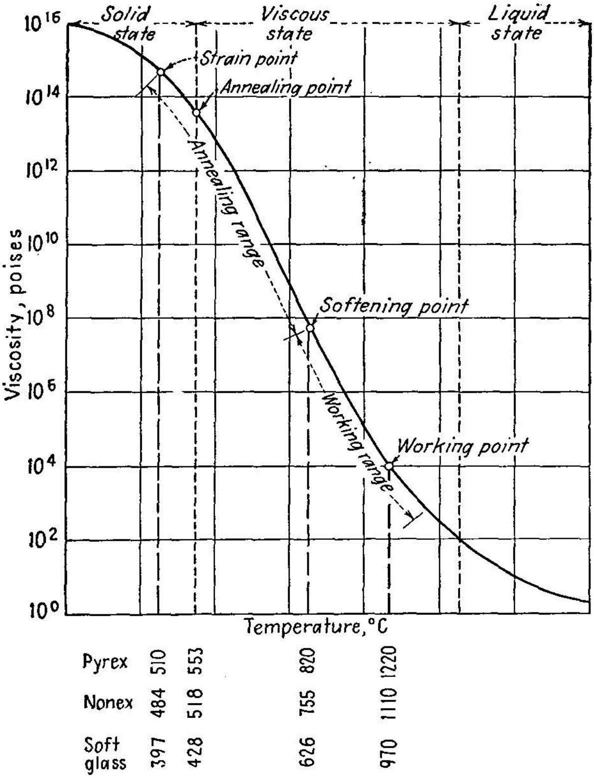


FIG. 21.38.—Viscosity of glass as a function of temperature.

The expansion characteristics of glasses are different for the different grades of glass and are nonuniform with temperature, unlike those of the pure metals.[1] Hence different types of glass cannot be joined together without cracking upon cooling unless their expansion coefficients are

---

[1] PETERS, C. G., and C. H. CRAGOE, Measurement of the Thermal Dilatation of Glass at High Temperatures, *U.S. Bur. Standards Sci. Paper* 393.

TABLE XV
PROPERTIES OF THE PRINCIPAL GLASSES*

Code	Corning Lab. No.	Density	Composition	Exp × 10^{-6}/°C	Strain pt., °C	Ann. pt., °C	Soft pt., °C	Work pt., °C	Use
.001	G-1	2.85	Soda potash lead (22 per cent PbO)	9.1	397	428	626	970	Lamp and rectifier-tube stems
.005	G-5	2.92	Soda lead	9.1	404	429	619	970	Vacuum-tube bulbs
.008	G-8	2.47	Lime	9.2	475	510	696	1000	Lamp and rectifier-tube bulbs
.012	G-12	3.05	Soda potash lead (30 per cent PbO)	8.9	400	433	630	970	Vacuum-tube stems
3320	G-371-BN	2.29	Soda-alum-borosilicate plus uranium	4.0	497	535	780		Intermediate nonex to pyrex. Also tungsten sealing
704	G-705-BA	2.24	Potash borosilicate with alumina	4.75	450	484	702	1080	Kovar sealing. Not much used on account of weathering
705	G-705-AJ	2.23	Soda-borosilicate with alumina	4.6	461	496	703	1100	Kovar-tungsten-molybdenum sealing
7052	G-705-FN	2.29	Potash-barium borosilicate	4.6	442	480	708	1115	Standard Kovar sealing (poor glass quality)
706	G-705A0	2.25	Borosilicate	5.0	463	495	690	1050	Kovar sealing. Best at present, but expensive
707	G-707-OG	2.13	Lithia borosilicate	3.2	455	490	746	1100	Low loss for ultra-high frequency
772	G-702-P	2.35	Soda-lead borosilicate	3.6	484	518	755	1110	Nonex–power-tube–tungsten sealing
774	G-726-MX	2.23	Soda-alum-borosilicate	3.25	510	553	820	1220	Pyrex chemical ware
775	G-705-R	2.19	Soda-alum-borosilicate	4.05	431	467	704	1100	Practically obsolete
776	G-720-GO	2.23	Borosilicate	3.4	480	520	780	1213	Pyrex bulbs
790		2.18	96 per cent silica (Vycor)	8	820	910	1500		Quartz substitute
9700	G-970-G	2.26	Borosilicate (Corexo)	3.7	517	558	804	1195	Ultraviolet transmitting
9740	G-970-HW	2.15	Borosilicate	4.5	405	450	775	1250	Ultraviolet transmitting

* Prepared by C. V. Litton and reproduced with his permission.

nearly alike. In general, all glasses have a lower rate of expansion at the low temperatures than at the high. Soft glasses have higher coefficients of expansion than hard glasses. The expansion-temperature characteristics of the principal glasses and metals used in vacuum tubes are shown in Fig. 21.39.

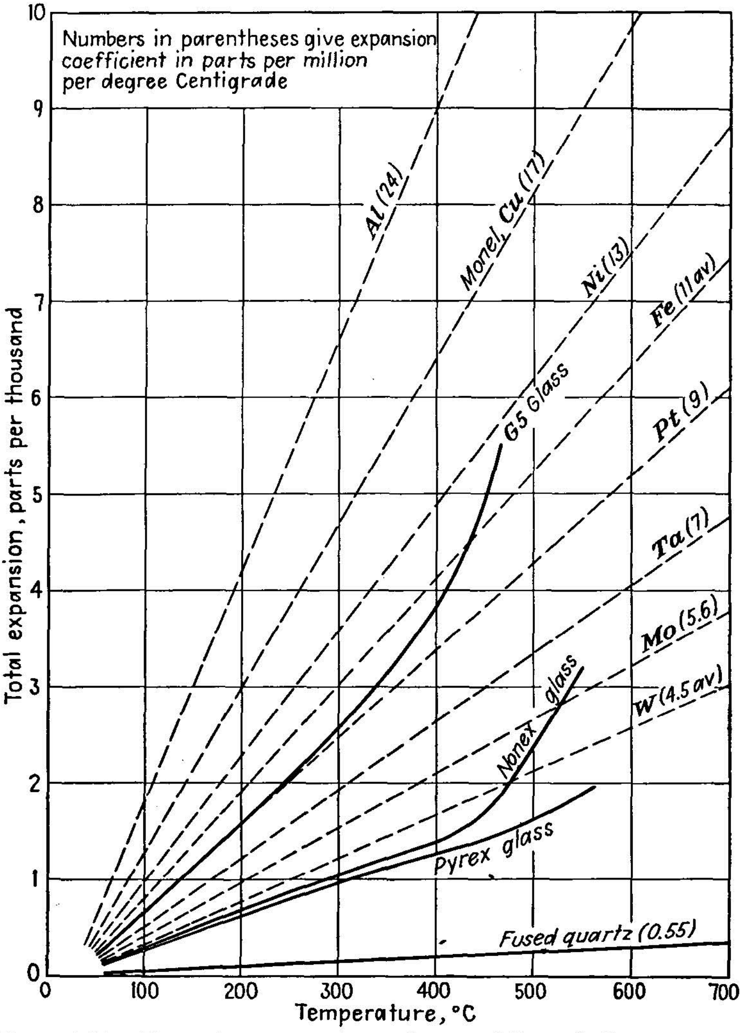


FIG. 21.39.—Expansion-temperature characteristics of the common metals and glasses used in vacuum-tube construction.

Working of Glass. The working of glass requires a high degree of physical coordination and skill. Simple operations can be learned in a short time, but a professional touch is slowly acquired![1]

21.8. Sealing of Glass to Other Materials. Very few vacuum tubes have been built with no glass in them. Even in the so-called "metal tubes" the leads are brought into the tube through a glass bead sealed into an eyelet. In experimental and developmental work, glass is used even more extensively and glass-to-metal sealing assumes even greater importance.

Sealing of Small Leads into Glass. The problem of bringing leads into vacuum tubes is ever present. The principal problem involved is that of finding a metal of which the expansion coefficient matches that of the glass quite closely. Since the expansion coefficient of metals is nearly constant with temperature while that of glass generally increases with temperature, the perfect combination is seldom found. However, if the diameter of the lead is small, a considerable mismatch in expansion can be tolerated. Thus with tungsten, of which the expansion coefficient is 4 ppm (parts per million) per °C, leads of diameter 0.020 in. or less can be sealed into pyrex glass, of which the expansion coefficient is 3.3 ppm per °C, whereas leads of diameter as great as 0.125 in. can be sealed into nonex glass, of which the expansion coefficient is 3.6 ppm per °C, without cracking. Because the coefficient of expansion of platinum, 9 ppm per °C, is very nearly the same as that of G-12 soft lime glass, 8.7 ppm per °C, lead size of this glass-metal combination is limited only by the budget. Platinum leads can also be sealed into G-12 soft cobalt lead glass, of which the coefficient of expansion is 8.7 ppm per °C. In all cases the glass and metal must be heated to a red heat together, bringing the glass to a soft state so that it will wet the metal. This generally requires that the metal be coated with an adherent coating of oxide and that the glass and metal be heated together so that the oxide partly dissolves in the glass, though perfect seals can be made *with no oxide* on copper, tungsten, or Kovar.

Metal-glass combinations other than those mentioned above may also

---

[1] For further information the reader is referred to FRARY, F. C., C. S. TAYLOR, and J. D. EDWARDS, "Laboratory Glass Blowing," 2d ed., McGraw-Hill, New York, 1928, and also the excellent illustrated treatment of STRONG, J., and others, "Procedures in Experimental Physics," Chap. I, Prentice-Hall, New York, 1941; PERCIVAL, G. A., The Technique of Glass Manipulation, *Electronic Eng.*, April, 1944, pp. 453–457; BREODUER, R. L., and C. H. SIMMS, Planning a Glassworking Department, *Jour. Sci. Instr.*, vol. 21, pp. 169–173, October, 1944; HOLDMAN, J. D., "Techniques of Glass Manipulation," Prentice-Hall, New York, 1946.

be used for sealing small leads into glass. Dumet, which is a copper-clad iron alloy, is extensively used in receiving-tube stems of soft glass.[1] The expansion coefficient of dumet is close enough to that of the soft glasses so that it can be used in diameters under 0.040 in. Molybdenum can be sealed into pyrex and nonex in small diameters. Corning G-71, softest of the hard glasses, matches the expansion of molybdenum very closely and can be used to fairly large sizes. Some of the stainless steels have expansion coefficients low enough to be used with this same glass. Chrome-iron alloys containing 26 to 28 per cent chromium match G-6 glass quite well at low temperatures.

Copper-to-glass Seals. Copper may be joined to almost any type of glass if the edge of the metal that is being joined to the glass is made extremely thin. This is possible in spite of the fact that the coefficient of expansion of copper is much greater than that of any of the glasses. A thin piece of copper will give to high stresses because of its high ductility and its low yield point. The technique of joining copper to glass was perfected by Housekeeper, and such seals are often referred to as "Housekeeper seals."[2] Copper-to-glass seals are invariably used in transmitting tubes for any seals requiring conductors larger than ⅜ in. in diameter.

Copper is prepared for sealing by cutting or rolling the edge of copper tubing so that the edge is 1.5 ± 0.5 thousandths of an inch thick and tapered back at about a 2.5-deg angle to about 40 thousandths thickness. The joining of glass to the copper requires a high degree of skill and is probably the most difficult of all the glassworking operations to perform. Small seals, up to ½ in. in diameter, are commonly made with the glass applied only to the inside of the copper edge. This is done because the expansion of copper is greater than that of glass and the differential expansion is therefore in the right direction to maintain the bond. For seals larger than ½ in. in diameter it is common to coat both the inside and the outside of the copper edge with glass. This is done primarily to prevent overoxidation of the thin copper at the seal. Copper must be heated with an *oxidizing* flame. The black oxide is formed, and seal temperature must be maintained constant throughout the operation. The glass is bound to the black oxide. On further heating the excess oxygen of the black oxide combines with more copper, changing it to the red oxide. Simultaneously some of the black oxide dissolves in the glass.

---

[1] Dumet cores are 42 per cent nickel, and the copper coating is 20 to 25 per cent of the total volume.

[2] HOUSEKEEPER, W. G., Glass to Metal Seals, *Jour. Amer. Inst. Elec. Eng.*, vol. 42, pp. 954–960, September, 1923. Earliest seals were made by Kruh and Kraus. Housekeeper introduced the featheredge.

As a result, the final interface between the red oxide and the copper lies at a new depth, created after the glass was fused to the outside. Mechanically a rim of glass is first attached to the outside of the copper edge from a piece of glass tubing and then detached from the glass tubing, leaving a little glass projecting over the edge of the copper, which is then folded over to cover the inner side. The glass tubing is then joined to this "bead" of glass on the copper edge. The color of a properly fashioned copper-glass seal is a bright red and will stand heating up to the softening temperature of the glass. Seals are commonly made up to 6 in. in diameter, and some as large as 10 in. in diameter have been made. The same procedure is used in joining copper to all types of glass. Joining of copper to pyrex is the most difficult, for there is a temperature interval of only a couple of hundred degrees between the temperature at which the glass softens and that at which the copper melts. The only disadvantages of copper-glass seals are that they are relatively difficult to make and that they have a relatively low mechanical strength because of the thinness of the copper next to the glass.

In addition to the Housekeeper seal it is possible to make disk seals to copper. This was anticipated by Housekeeper but only recently put into extensive commercial use. Disk seals are used in tubes of the lighthouse type and in reflex-klystron-oscillator tubes designed to work with external cavities.[1,2] The general method of construction consists in stacking a circular copper disk with a circular hole in its center between two equal-diameter pieces of glass and then heating the metal by torch or preferably by r-f eddy currents until the copper disk becomes hot enough to melt the glass, which forms a bond with the metal. The glass does not cover the edges of the copper. If the copper disk is thin enough, 15 thousandths of an inch or less, then no intermediate materials are needed between the glass and copper. As with the Housekeeper featheredge seal, the difference between the expansions of the glass and copper is taken up by the copper. Copper disks are frequently given a circular crimp to weaken them to radial forces and allow radial contraction without having to stretch the whole metal area. The surfaces of thick disks are often coated with a layer of copper borate, which ensures maximum bonding strength.

Kovar and Fernico. Kovar and Fernico are trade names used by the Westinghouse and General Electric Company, respectively, for some nickel-cobalt alloys of iron having nonuniform expansion characteristics

---

[1] Disc Seal Tubes, *Gen. Elec. Rev.*, vol. 48, pp. 50–51, January, 1945.

[2] McArthur, E. D., Disc Seal Tubes, *Electronics*, vol. 18, pp. 98–102, February 1945.

that match very closely the expansion of some commercial glasses.[1–3] Pure metals have expansion coefficients that are virtually independent of temperature. Ferromagnetic alloys, however, experience an increase in their expansion characteristics at the temperature at which the alloy becomes hot enough to lose its magnetic properties. The action is perfectly reversible, *i.e.*, the magnetism is restored and the coefficient of expansion reduced as the alloy is cooled. The composition of Kovar and Fernico is as follows:

Alloy	Iron, per cent	Nickel, per cent	Cobalt, per cent	Matching glass
Kovar A	53.8	29	17	705A0, 705FN
Fernico	54	28	18	705A0, 705FN
Fernichrome	30	25	8	G8

The difference in contraction of the principal sealing glasses and metals when cooled at a slow rate is shown in Fig. 21.40. It is seen that the iron alloys match the glass characteristics quite closely over the entire temperature range. As a result, the sealing of these metals to their corresponding glasses is a relatively simple matter. No featheredges are needed; in fact, edges as thick as ⅛ in. can be joined directly. Seals as large as 4 in. in diameter can be made. Leads of Fernico wire in matching glass set in a Fernico eyelet that is welded to a metal base are used in the mass production of metal receiving tubes. As may also be seen from Fig. 21.40, the reason why nonex seals fairly successfully to tungsten is that the differential expansion is nearly zero in the annealing range. Uranium nonex gives a better match and is sometimes used as an intermediary between tungsten and pyrex glass.

Glass-to-porcelain Seals. The expansion characteristics of nonex glass and some porcelains are close enough so that nonex can be sealed directly to porcelain. Where a porcelain-pyrex joint is desired, nonex should be used as an intermediary material.

Glass-to-mica Seals. Mica can be sealed to a special high-expansion lead borosilicate glass having an expansion of coefficient of about 9.8

---

[1] Burger, E. E., Expansion Characteristics of Some Common Glasses and Metals, *Gen. Elec. Rev.*, vol. 37, pp. 93–99, February, 1934.

[2] Hull, A. W., and E. E. Burger, Glass to Metal Seals, Part I, *Physics*, vol. 5, pp. 384–405, December, 1934.

[3] Hull, A. W., E. E. Burger, and L. Navais, Glass to Metal Seals, Part II, *Jour. Appl. Phys.*, vol. 12, pp. 698–707, September, 1941.

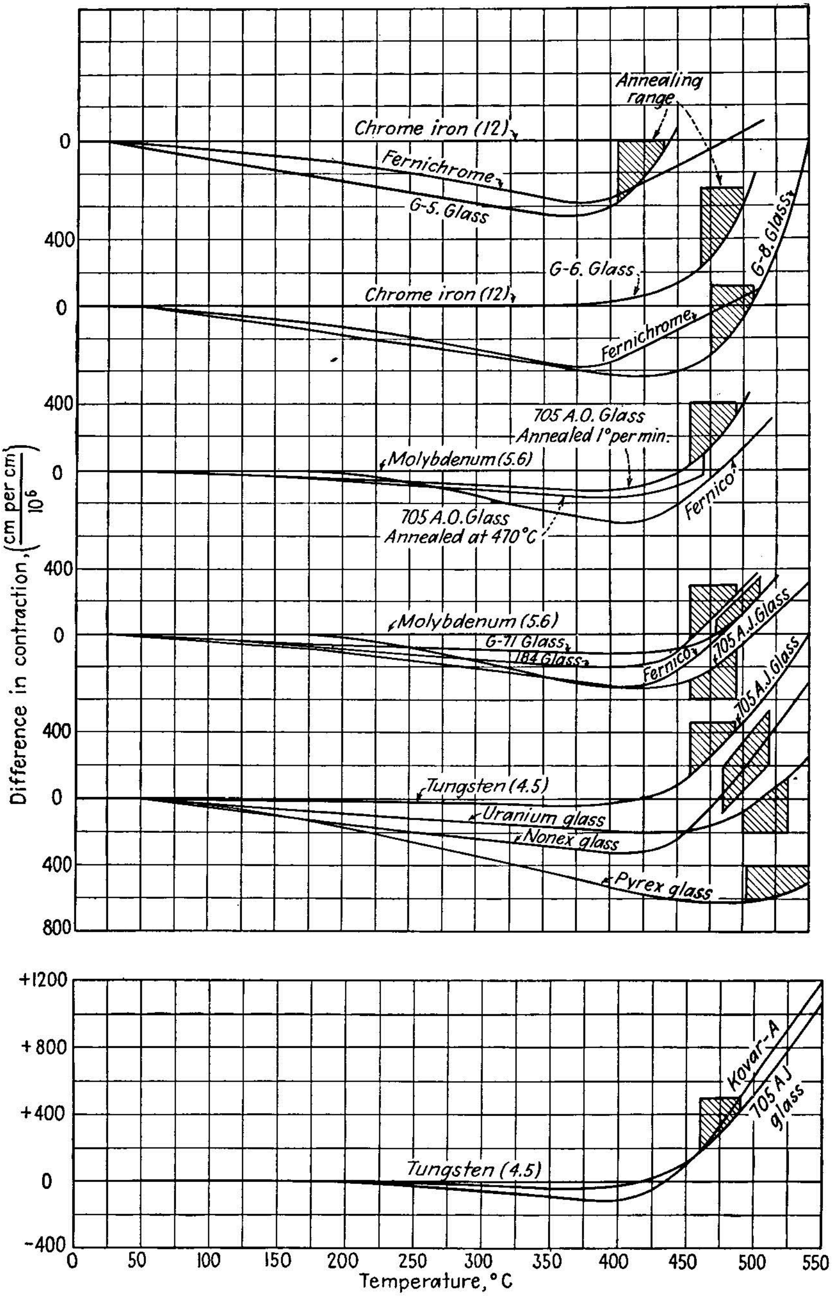


FIG. 21.40.—Difference in contraction of the principal sealing glasses and metals.

ppm per °C.[1] Mica-to-metal joints are more difficult to form, for the metal invariably has a higher expansion coefficient than the mica, which results in the mica becoming bowed on cooling. With proper intermediary oxides and glass, mica can be sealed to copper with very little resultant bowing. Window thicknesses may range from five to twenty-thousandths of an inch in thickness and be as large as two inches in diameter.

Metal-to-metal Sealing. Metals can be joined by suitable solders. For demountable systems kept on a pump and not involving high temperatures, brass can be used, joined by ordinary soft solder, the eutectic proportions (lowest melting temperature) of tin and lead giving the best results. Brass and soft solder cannot be used in tubes that are to be sealed off, for brass is somewhat porous, evolves great quantities of gas, and tends to vaporize its zinc at high temperatures. For tubes that are to be sealed off, oxygen-free copper, most iron alloys, aluminum, and beryllium can be used as vacuum-tight containers. Joining is most satisfactorily effected by means of high-melting-temperature silver-copper alloys melted in a hydrogen (reducing) atmosphere by means of a tungsten filament or induction heater. This involves heating the metals to a red heat, at which temperature the silver-copper alloys flow freely and wet clean metal surfaces. For high-temperature work, gold-copper alloys are also used. Gold has a lower vapor pressure than silver.

21.9. Metals Useful in Tube Construction. The properties required of metals for use in vacuum-tube construction are rather numerous. In general, no one metal meets all the requirements, but each metal in turn has its distinctive advantages.[2,3]

Mechanically, a metal to be useful in vacuum-tube construction should have a strength and ductility that permit easy forming of electrode shapes. The strength must be retained at high temperature without excessive crystallization to avoid deformation during degassing and subsequent use. The stiffness and damping factor of the metal should be high, to reduce vibration effects.

Thermally, the coefficient of expansion should be relatively low and,

---

[1] DONAL, J. S., JR., Sealing Mica to Glass or Metal to Form a Vacuum Tight Joint, *Rev. Sci. Instr.*, vol. 13, pp. 266–267, June, 1942.

[2] See WISE, R. M., Nickel in the Radio Industry, *Proc. I.R.E.*, vol. 25, pp. 714–752, June, 1937, for a detailed treatment of this subject with special reference to nickel. This paper contains an extensive bibliography on the general subject of metals in tubes.

[3] ESPE and KNOLL, "Werkstoffkunde der Hochvakuumtechnik," *op. cit.*, pp. 1–110. Obtainable from Edwards Bros., Ann Arbor, Mich. A classic source containing the most extensive information available in book form.

except for special applications, quite constant. Good thermal conductivity is generally sought. Depending upon the application, metals should have either a high reflectivity or a high thermal emissivity. The vapor pressure at degassing temperatures should be low, while the melting temperature itself should be well above the highest degassing or operating temperature.

Electrically, a moderate conductivity is desired. Too low a conductivity introduces appreciable resistance and attendant losses, while too high a conductivity makes spot welding difficult. Except for cathodes, the primary and secondary emission should be low. Except for shielding applications, the magnetic permeability should be low, and the metal should be one that is readily demagnetized by a magnetic field.

Chemical freedom from oxidation at high temperatures simplifies construction processes immensely. Resistance to corrosion by various cleaning agents should be low. Most important of all, the metal should absorb only a small amount of gas and give this up easily when heated in vacuum.

In addition, materials should be relatively inexpensive and generally available. Alloys having a wide range of physical characteristics as determined by their chemical content are especially useful.

Nickel. Nickel is the metal that is most extensively used in forming receiving-tube electrodes. It is easily drawn and formed. It stretches easily and does not exhibit any sharp break at its yield point. Its hardness and strength at high temperatures are good. It has thirteen times the mechanical damping factor of iron and molybdenum. It spot-welds well to almost all metals. Its expansion coefficient is nearly constant with temperature, and its thermal and electrical conductivity are good. When polished, nickel has an emissivity which ranges from 5 to 20 per cent of that of a black body, *i.e.*, it makes a good reflector. When carbon-coated, the thermal emissivity ranges from 80 to 94 per cent of that of a black body, *i.e.*, it makes a good radiator. Anodes formed of nickel are usually carbon-coated to increase their radiation. Vapor pressure is low at all but very high temperatures, 10^{-6} mm of mercury at a red heat. The work function of nickel is high, 5 volts, but commercial nickel may have appreciable thermionic emission due to barium contamination. Alloying about 4.5 per cent manganese reduces both primary and secondary emission. Others of the desirable properties are likewise present. As a result, nickel is an ideal metal for tube construction in all applications except those where a high temperature is involved.

Copper. The outstanding physical characteristics of copper are its high thermal and electrical conductivity. As has also been mentioned, it can be sealed to all glasses by the Housekeeper technique. It is extensively used as an anode material in water- and air-cooled tubes. It is moderately porous and requires a thick wall to withstand atmospheric pressure when hot. Likewise, it oxidizes readily and so cannot be allowed to assume temperatures above a few hundred degrees centigrade. It must be used in the oxygen-free form in all applications that involve heating for red heat. Even within a vacuum, copper must be protected from high temperatures, for it softens and vaporizes at relatively low temperatures. Its high ductility and low yield point make it easy to draw, form, and spin.

Aluminum. Aluminum is easy to work and is fairly noncorrosive to other materials encountered in vacuum-tube construction. One valuable property is that it does not sputter easily. However, it melts at too low a temperature and absorbs too much gas to be very useful in sealed-off tubes.

Molybdenum. Molybdenum has most of the excellent properties of nickel except that it is somewhat harder to work and is more expensive. Its relatively high melting temperature and low vapor pressure make it useful in low-power transmitting tubes. It is readily spot-welded to iron or nickel but not to tungsten. It absorbs oxygen when heated to a dull red heat. Molybdenum is used in applications that involve temperatures in the range of 200 to 500°C.

Tantalum. Next to tungsten, tantalum has the highest melting temperature of all the metals. Its vapor pressure is very low. It is easily formed and drawn. The metal is expensive as a result of the relatively complicated vacuum processing required to put it into form suitable for vacuum-tube construction. It is extensively used in radiation-cooled transmitting tubes, where the electrodes are often run at a red heat. It has a getter action that causes it to absorb gases, particularly hydrogen, the maximum absorption occurring at 1000°C (cherry red). The gases that have been absorbed are given off again at temperatures of 1300°C and higher. Minimum temperature for getter action is approximately 800°C. Tantalum is also used as an emitter in applications requiring specially shaped cathodes. Its work function is lower than that of tungsten, with the result that its emission is greater at the same temperature. Tungsten can, of course, achieve higher emission because it can be heated to higher temperatures without melting.

Tungsten. Reference has already been made to some of the numerous applications of tungsten in vacuum-tube construction; as an emitter and filament wire and in some lead-sealing applications it has virtually no substitute. Its high melting temperature makes it especially useful in some vacuum-tube construction processes. It is used as a filament wire for silver-soldering operations. It is likewise used as a filament heater in numerous metal-evaporation processes. It is one of the few metals that can be used as a target in X-ray tubes. Numerous gauges

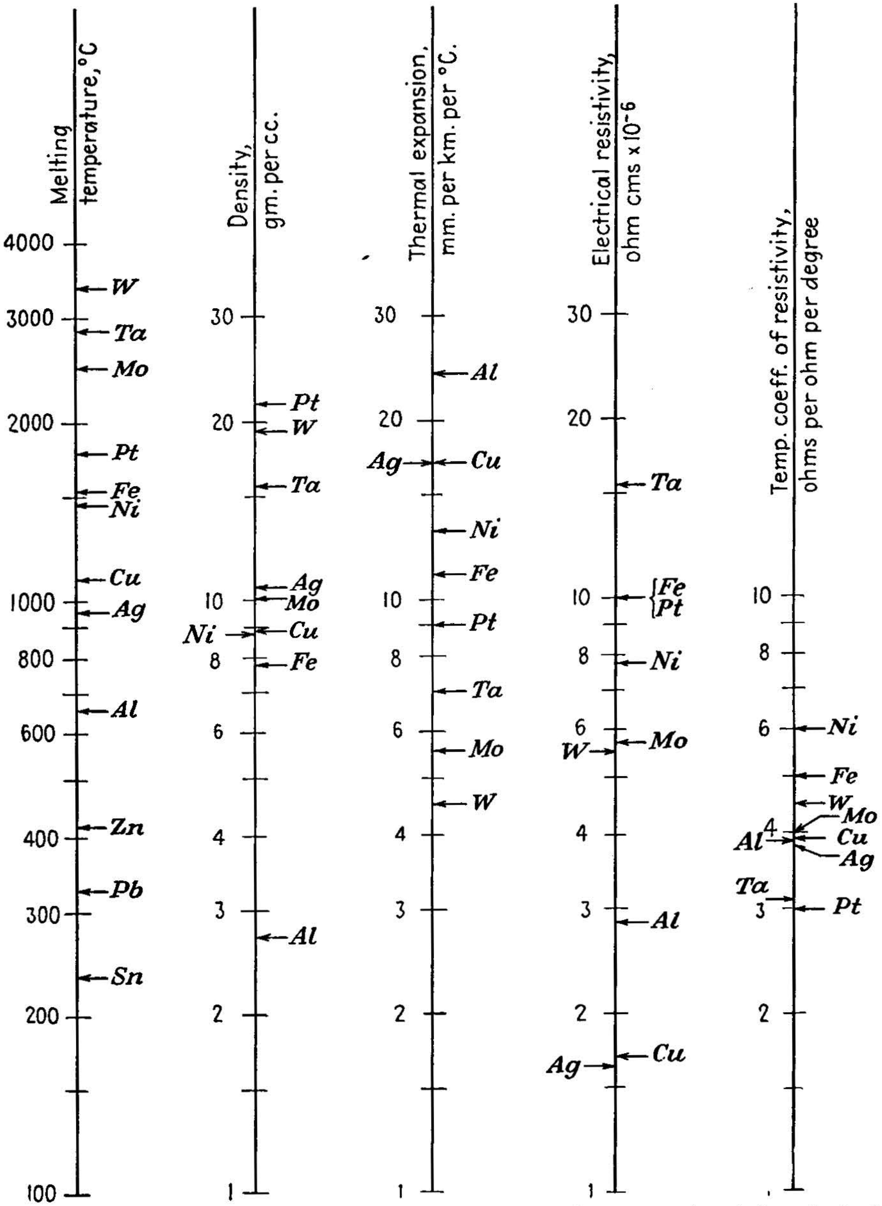
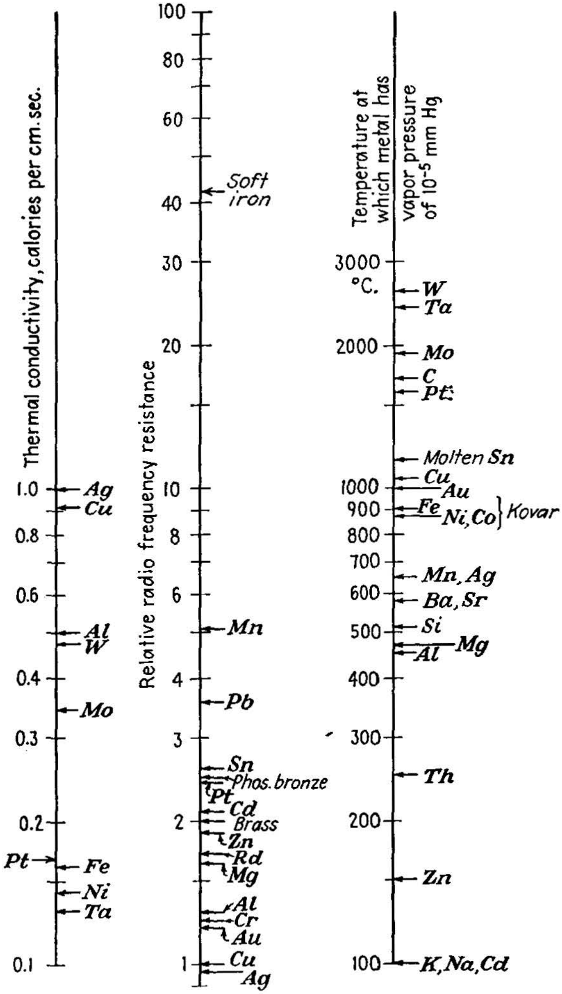


FIG. 21.41.—Relative properties of the principal

and control devices make use of its large change of resistance with temperature.

Tungsten is not readily drawn or formed. It must be hammered or swaged into shape. As a result, it is principally available in wire or rod form. Tungsten has a pronounced crystalline structure, which is accentuated by heating. Tungsten filaments therefore become brittle if overheated for appreciable periods of time. Tungsten is relatively inactive chemically, which reduces contamination problems. It is sometimes alloyed with molybdenum (W/Mo = $^{49}\!/_{51}$) to give a material



metals used in vacuum-tube construction.

that is more workable than tungsten itself and yet retains a high melting temperature.

Relative Properties of the Metals. Metals other than those listed separately above find many special applications in tube construction, but those listed take care of the majority of the applications. The relative properties of the principal metals used in vacuum-tube construction are shown on the scale lines of Fig. 21.41.[1] One of the most important properties of a metal is its vapor pressure, which is an increasing function of temperature. The last scale line gives the temperature of which a metal has a vapor pressure of 10^{-5} mm of mercury. This determines the highest temperature to which a metal in a tube can be raised during the exhaust process.

Spot Welding. In the construction of vacuum tubes the majority of small metal-to-metal joints are formed by spot welding. Basically the process of spot welding consists in passing a large current through the joint to be welded. The joint is heated by the large current density, of the order of thousands of amperes per square inch, to the point where the metals melt and dissolve into one another, forming a weld.

Spot-welding machines consist of a set of pointed jaws supported by a mechanical arrangement that brings the jaws together by the operation of a foot pedal. The materials to be welded are placed between the jaws, and pressure is applied by the foot pedal. Care must be taken in supporting the work between the jaws to see that current will flow from the jaws through the work and through the point to be welded. The jaws are connected to a step-down transformer that gives a large current through a closed circuit when the primary is closed by means of another foot pedal. For most operations the jaws are made of copper and because of their resulting high conductivity will have relatively little heat developed at their point of contact with the work. Where welding operations are at all critical, an electronic circuit should be used to control the amount of current and the time duration of current flow. Many welding operations require a current flow of hundreds of amperes for a fraction of a second.

Not all metal combinations will spot-weld readily. Difficulties are encountered with metals of high conductivity, high melting temperature, and high oxidation tendencies. In Table XVI there is indicated the relative ease with which different metals can be spot-welded to one another.[2]

Spot welding forms only a part of the art of joining metals. In the

---

[1] For more complete data than are given here the reader is referred to ESPE and KNOLL, *loc. cit.*

[2] ESPE and KNOLL, *op. cit.*, pp. 135–139

newer tube designs, extensive use is made of r-f brazing both in hydrogen atmosphere and in air. This brazing process makes use of single shots of r-f power (about 100 kc) where the pulses are of the duration of 0.001 to 0.1 sec for the whole weld. Arc welding is also employed in both hydrogen and argon atmospheres. This is essentially atomic welding. The gas is dissociated by the arc and then recombines on the work, where it liberates energy in very concentrated form.

TABLE XVI
SPOT-WELDING PROPERTIES OF THE METALS

	Monel	Constantan	Ni/Cr (8/2)	Invar	Fe/Cr (7/3)	Al	Cu	Fe	Ni	Ta	Mo	W
Tungsten	..	..	..	..	..	4	4	C3	2	A4	C4	4
Mo	B	..	..	B	..	4	4	C3	2	4	C4	
Ta	..	..	..	..	..	4	A4	C3	2	3		
Ni	B	3	B	B	2	3–4	3	1	1			
Fe (pure)	B	2	B	..	..	3	2–3	1				
Cu	..	..	..	..	..	4	4	B3-4				
Al	4	2	4	..	..	3						
Fe/Cr (7/3)	..	..	..	3								
Invar	B	..	B	B								
Ni/Cr (8/2)	B	..	B									
Constantan	..	B										
Monel	B											

1. Very good.
2. Good.
3. Difficult.
4. Bad or impossible.

A. Good with suitable flux.
B. Good with controlled current impulses.
C. Good for small wires with short controlled current impulses.

21.10. Insulators. In addition to glass, which is a good insulator at low temperatures, mica and various ceramics are the principal insulators used in vacuum-tube construction.

Mica is extensively and almost exclusively used as an insulator and electrode spacer in receiving tubes. It is a potassium-aluminum silicate, which in its natural form is known as "muscovite." Mica as used in tubes is a dehydrated muscovite. It has a crystalline structure that permits it to be split into thin sheets. Sheets as thin as 0.5 thousandths of an inch can be had. For receiving-tube use, the sheets are usually of the order of 20 thousandths of an inch thick. Mica has one of the highest specific resistances of all known insulators. Its dielectric constant of 5 to 8 makes it useful in electrical condensers. Its breakdown voltage lies in the range of 60 to 200 kv per mm. It can be used at temperatures up to 500°C.

Various ceramics are used as insulators in transmitting tubes that involve higher temperatures and strength than mica can withstand and furnish. Most useful are the various silicates, principally those of magnesium. These materials are not machinable but can be formed to almost any shape desired before they are fired. A compromise on machinability has been achieved in some special materials, such as Alsimag 222, which can be machined with a stellite or other hard tool. For experimental work, soapstone, which is very soft, is often machined to shape and then hardened by heating in hydrogen to a red heat.

Porcelain, as has been mentioned, finds some applications where it is necessary to get a glass-ceramic seal. It is not machinable but must be formed in the desired shape before firing.

Aluminum oxide is often used as an insulating coating on filament wires. The coating is obtained by either dipping or spraying from a suspension of amyl acetate and then drying at about 600°C. A hard vitreous coating is formed by flashing at 1500°C. Such insulating coatings are most effective if made of a succession of thin layers each baked individually. The resulting insulation has a sufficiently high mechanical strength to make it useful for filament wires used in indirectly heated cathodes. The electrical strength is likewise adequate for low-voltage applications.

21.11. Degassing of Glass and Metals. Materials used in vacuum tubes must be heated to drive off gases during the evacuation process. Some of the gas is merely condensed on the surface, in which it is said to be *adsorbed*. Other gases are in chemical combination with the material, in which case they are said to be *absorbed*. With metals there will generally be considerable quantities of gas trapped in crevices, seams, and flaws. Such gases are said to be *occluded*.

In general, tubes should be degassed by heating at temperatures appreciably greater than the temperatures the tube will encounter in practice. The time required for outgassing may range from 15 min for receiving tubes to hours or days for high-power transmitting tubes.

The gases encountered with glass are mostly adsorbed. A 40-watt lamp bulb will evolve about 500 cm^3 of gas (measured at room temperature and pressure) when heated at 500°C. About 90 per cent of this gas is in the form of water vapor. Glasses should be heated at about 90 per cent of their annealing temperature to drive off adsorbed gases. At higher temperatures the glass may soften, and some gases will be given off by decomposition of the glass. The time required for outgassing of glass is about 15 min at top temperature. Heating may be done with either soft gas flames or with a baking oven that surrounds the entire tube. Heating with a baking oven allows a better control of temperatures, though receiving tubes are frequently degassed with gas flames. Degassing of a tube should not be begun until the tube has been evacuated to a pressure of 10^{-3} mm of mercury or less.

The gases encountered with metals are mostly in the form of occluded gases. Metal electrodes and parts may be degassed by heating to about 50 per cent of the melting temperatures of the metals. The amount of gas evolved from a metal will depend upon the area multiplied by a depth of a few thousandths of an inch, except for tungsten and molybdenum, which have a laminar structure. The principal component of the gases involved is generally carbon monoxide, which is present to the extent of about 30 to 90 per cent of the total gases. The remainder of the gas is mostly nitrogen, which comes off at a higher temperature than carbon monoxide. Interestingly enough, when a metal has been degassed by heating in a vacuum it will pick up very little gas upon subsequent exposure to air at atmospheric pressure, if carefully handled.[1] Degassing of metals is commonly achieved by r-f induction heating. Radiation-cooled transmitting tubes may be degassed by direct electronic bombardment of the elements.

21.12. Getters. Getters are materials used in vacuum tubes to clean up residual gases by chemical combination. The alkali metals are most extensively used. Barium seems to be most effective in cleanup action though magnesium, calcium, sodium, and phosphorus have also been used.[2,3] The getter material is usually enclosed in the pure metal form in a small cup or wire cage of base metal and then reduced and vaporized, after the tube is sealed off, by heating to a temperature of about 700°C by r-f induction currents. Sometimes the getter material is contained in a tube formed of a rolled nickel sheet, in which case the vaporized metal escapes through the crack in the tube. The vaporized metal deposits on the wall of the tube, care always being taken that it does not deposit on any of the insulators. When gas molecules come in contact with this layer, they will combine (except for the noble gases), with the result that the vacuum gets progressively better with time. A

---

[1] NORTON, E. J., and A. L. MARSHALL, The Degassing of Metals, *Gen. Elec. Co. Research Lab. Rept.* 613, March, 1932.

[2] LEDERER, E. A., and D. H. WAMSLEY, Batalum, a Barium Getter for Metal Tubes, *RCA Rev.*, vol. 11, pp. 117–123, July, 1937.

[3] LEDERER, E. A., Recent Advances in Barium Getter Techniques, *RCA Rev.*, vol. 14, pp. 310–318, January, 1940.

getter in a receiving tube will usually be sufficient to improve the vacuum obtained from a mechanical pump to 10^{-6} mm of mercury in about 10 min. Previous treatment of the getter to remove gases seems to be more important then the material of the getter itself.[1,2]

The absorption properties of other metals may also be used in the form of an auxiliary filament. Tungsten, molybdenum, and tantalum can be used for this purpose if heated to 1000°C or higher. Most interesting of all the metals in its cleanup action is zirconium.[3,4] Zirconium will absorb 5 times its own volume of hydrogen at 400°C, while at 1400°C it will absorb carbon monoxide and carbon dioxide as well as 40 times its own volume of oxygen and 20 times its own volume of nitrogen. At temperatures below 200°C, protective oxides and nitrides form. For complete getter action, two filaments, one to work at 400°C and one to work at 1400°C, are necessary. Zirconium-filament getters are seldom used in commercial tubes but are useful in experimental tubes. Zirconium is often used in the form of a sprayed powdered coating applied to metal anodes. This gives increased thermal emissivity and also a continuous getter action during operation.

---

[1] ANDREWS, M. R., and J. S. BACON, The Comparison of Certain Commercial Getters, *Gen. Elec. Research Paper* 574, June, 1931, also published in *Jour. Amer. Chem. Soc.*, pp. 1674–1681, May, 1931.

[2] DUSHMAN, "The Production and Measurement of High Vacuum," *op. cit.* The last half of this book is devoted to the subject gas sorption and degassing of materials.

[3] FAST, J. D., Zirkon und seine hochschmelzenden Verbindungen, *Philips Tech. Rev.*, vol. 3, pp. 353–360, December, 1938.

[4] FAST, J. D., Metals as Getters, *Philips Tech. Rev.*, vol. 5, pp. 217–221, August, 1940.

APPENDIX I

PROPERTIES OF THE ELEMENTS

A. Atomic Weights and Numbers

	Symbol	Atomic number	Atomic weight		Symbol	Atomic number	Atomic weight
Aluminum	Al	13	26.97	Molybdenum	Mo	42	95.95
Antimony	Sb	51	121.76	Neodymium	Nd	60	144.27
Argon	A	18	39.944	Neon	Ne	10	20.183
Arsenic	As	33	74.91	Nickel	Ni	28	58.69
Barium	Ba	56	137.36	Nitrogen	N	7	14.008
Beryllium	Be	4	9.02	Osmium	Os	76	190.2
Bismuth	Bi	83	209.00	Oxygen	O	8	16.0000
Boron	B	5	10.82	Palladium	Pd	46	106.7
Bromine	Br	35	79.916	Phosphorus	P	15	30.98
Cadmium	Cd	48	112.41	Platinum	Pt	78	195.23
Calcium	Ca	20	40.08	Potassium	K	19	39.096
Carbon	C	6	12.010	Praseodymium	Pr	59	140.92
Cerium	Ce	58	140.13	Protoactinium	Pa	91	231.
Caesium	Cs	55	132.91	Radium	Ra	88	226.05
Chlorine	Cl	17	35.457	Radon	Rn	86	222.
Chromium	Cr	24	52.01	Rhenium	Re	75	186.31
Cobalt	Co	27	58.94	Rhodium	Rh	45	102.91
Columbium	Cb	41	92.91	Rubidium	Rb	37	85.48
Copper	Cu	29	63.57	Ruthenium	Ru	44	101.7
Dysprosium	Dy	66	162.46	Samarium	Sm	62	150.43
Erbium	Er	68	167.2	Scandium	Sc	21	45.10
Europium	Eu	63	152.0	Selenium	Se	34	78.96
Fluorine	F	9	19.00	Silicon	Si	14	28.06
Gadolinium	Gd	64	156.9	Silver	Ag	47	107.880
Gallium	Ga	31	69.72	Sodium	Na	11	22.997
Germanium	Ge	32	72.60	Strontium	Sr	38	87.63
Gold	Au	79	197.2	Sulphur	S	16	32.06
Hafnium	Hf	72	178.6	Tantalum	Ta	73	180.88
Helium	He	2	4.003	Tellurium	Te	52	127.61
Holmium	Ho	67	164.94	Terbium	Tb	65	159.2
Hydrogen	H	1	1.0080	Thallium	Tl	81	204.39
Indium	In	49	114.76	Thorium	Th	90	232.12
Iodine	I	53	126.92	Thulium	Tm	69	169.4
Iridium	Ir	77	193.1	Tin	Sn	50	118.70
Iron	Fe	26	55.85	Titanium	Ti	22	47.90
Krypton	Kr	36	83.7	Tungsten	W	74	183.92
Lanthanum	La	57	138.92	Uranium	U	92	238.07
Lead	Pb	82	207.21	Vanadium	V	23	50.95
Lithium	Li	3	6.940	Xenon	Xe	54	131.3
Lutecium	Lu	71	174.99	Ytterbium	Yb	70	173.04
Magnesium	Mg	12	24.32	Yttrium	Y	39	88.92
Manganese	Mn	25	54.93	Zinc	Zn	30	65.38
Mercury	Hg	80	200.61	Zirconium	Zr	40	91.22

B. Periodic Table

1 **H** 1.0081	Atomic number → Atomic weight →	2 **He** 4.003

Group I	Group II	Group III	Group IV	Group V	Group VI	Group VII	Group 0
3 **Li** 6.940	4 **Be** 9.02	5 **B** 10.82	6 **C** 12.010	7 **N** 14.008	8 **O** 16.0000	9 **F** 19.000	10 **Ne** 20.183
11 **Na** 22.997	12 **Mg** 24.32	13 **Al** 26.97	14 **Si** 28.06	15 **P** 30.98	16 **S** 32.06	17 **Cl** 35.457	18 **A** 39.944

I a	II a	III b	IV b	V b	VI b	VII b	VIII			I b	II b	III a	IV a	V a	VI a	VII a	0
19 **K** 39.096	20 **Ca** 40.08	21 **Sc** 45.10	22 **Ti** 47.90	23 **V** 50.95	24 **Cr** 52.01	25 **Mn** 54.93	26 **Fe** 55.84	27 **Co** 58.94	28 **Ni** 58.69	29 **Cu** 63.57	30 **Zn** 65.38	31 **Ga** 69.72	32 **Ge** 72.60	33 **As** 74.91	34 **Se** 78.96	35 **Br** 79.916	36 **Kr** 83.7
37 **Rb** 85.48	38 **Sr** 87.63	39 **Y** 88.92	40 **Zr** 91.22	41 **Cb** 92.91	42 **Mo** 95.95	43 **Ma**	44 **Ru** 101.7	45 **Rh** 102.91	46 **Pd** 106.7	47 **Ag** 107.88	48 **Cd** 112.41	49 **In** 114.76	50 **Sn** 118.70	51 **Sb** 121.76	52 **Te** 127.61	53 **I** 126.92	54 **Xe** 131.3
55 **Cs** 132.91	56 **Ba** 137.36	57 to 71	72 **Hf** 178.6	73 **Ta** 180.88	74 **W** 183.92	75 **Re** 186.31	76 **Os** 190.2	77 **Ir** 193.1	78 **Pt** 195.23	79 **Au** 197.2	80 **Hg** 200.61	81 **Tl** 204.39	82 **Pb** 207.21	83 **Bi** 209.00	84 **Po** 210	85	86 **Rn** 222
87	88 **Ra** 226.05	89 to 96	**Th?**	**Pa?**	**U?**												

57 to 71 Rare earth elements	57 **La** 138.92	58 **Ce** 140.13	59 **Pr** 140.92	60 **Nd** 144.27	61 **Il**	62 **Sm** 150.43	63 **Eu** 152.0	64 **Gd** 156.9	65 **Tb** 159.2	66 **Dy** 162.46	67 **Ho** 164.94	68 **Er** 167.2	69 **Tm** 169.4	70 **Yb** 173.04	71 **Lu** 174.99	
89 to 96 Actinium series	89 **Ac** 227.05	90 **Th** 232.12	91 **Pa** 231	92 **U** 238.07	93 **Np** 237	94 **Pu** 239	95 **Am*** 241	96 **Cm*** 242								

*** Americium (Am) and curium (Cm) have been prepared artificially, but are not known to occur naturally.**

Reprinted by permission from "Fundamentals of Physical Science," 2d ed., by Konrad B. Krauskopf, McGraw-Hill, New York, 1948.

APPENDIX II

DIFFERENTIAL OPERATORS AND VECTOR NOTATION[1]

1. Differential operators for rectangular coordinates (mutually perpendicular unit vectors $\boldsymbol{a}_x$, $\boldsymbol{a}_y$, $\boldsymbol{a}_z$).

Gradient:

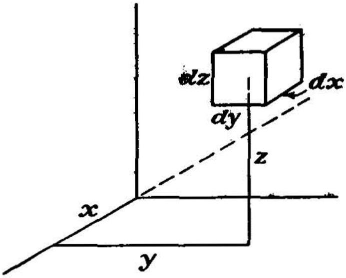
$$-\boldsymbol{E} = \text{gradient } V = \nabla V = \left(\boldsymbol{a}_x \frac{\partial V}{\partial x} + \boldsymbol{a}_y \frac{\partial V}{\partial y} + \boldsymbol{a}_z \frac{\partial V}{\partial_z}\right)$$

Components:

$$\text{grad}_x\, V = \frac{\partial V}{\partial x}$$

$$\text{grad}_y\, V = \frac{\partial V}{\partial y}$$

$$\text{grad}_z\, V = \frac{\partial V}{\partial z}$$



The gradient of any scalar quantity is always a vector quantity.

Divergence:

$$\text{Divergence } \boldsymbol{E} = \text{div } \boldsymbol{E} = \nabla \cdot \boldsymbol{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}$$

The divergence of any vector quantity is always a scalar quantity.

Curl:

$$\text{curl } E = \nabla \times E = \begin{vmatrix} \boldsymbol{a}_x & \boldsymbol{a}_y & \boldsymbol{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}$$

---

[1] For the development of the relations of this appendix and further information on vector notation and relations see

Skilling, H. H., "Fundamentals of Electric Waves," Wiley, New York, 1942.

Harnwell, G. P., "Principles of Electricity and Electromagnetism," McGraw-Hill, New York, 1938.

Stratton, J. A., "Elektromagnetic Theory," McGraw-Hill, New York, 1941.

Components:

$$\operatorname{curl}_x \boldsymbol{E} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}$$

$$\operatorname{curl}_y \boldsymbol{E} = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}$$

$$\operatorname{curl}_z \boldsymbol{E} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}$$

The curl of a vector quantity is always a vector quantity.

Laplacian:

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}$$

The Laplacian of a scalar quantity is always a scalar quantity.

2. Differential operators for cylindrical coordinates (mutually perpendicular unit vectors $\boldsymbol{a}_r$, $\boldsymbol{a}_\theta$, $\boldsymbol{a}_z$).

Gradient:

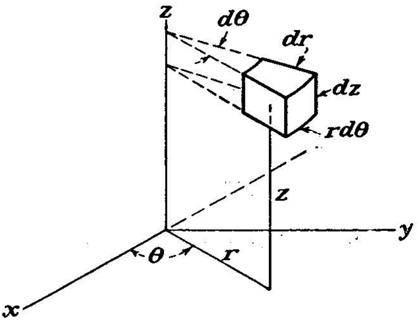
$$-\boldsymbol{E} = \text{gradient } V = \nabla V$$

Components:

$$\operatorname{grad}_r V = \frac{\partial V}{\partial r}$$

$$\operatorname{grad}_\theta V = \frac{1}{r}\frac{\partial V}{\partial \theta}$$

$$\operatorname{grad}_z V = \frac{\partial V}{\partial z}$$



Divergence:

$$\operatorname{div} \boldsymbol{E} = \nabla \cdot \boldsymbol{E} = \frac{1}{r}\frac{\partial}{\partial r}(rE_r) + \frac{1}{r}\frac{\partial E_\theta}{\partial \theta} + \frac{\partial E_z}{\partial z}$$

Curl:

$$\operatorname{curl} \boldsymbol{E} = \nabla \times \boldsymbol{E} = \frac{1}{r}\begin{vmatrix} \boldsymbol{a}_r & r\boldsymbol{a}_\theta & \boldsymbol{a}_z \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial z} \\ E_r & rE_\theta & E_z \end{vmatrix}$$

Components:

$$\operatorname{curl}_r \boldsymbol{E} = \frac{1}{r}\frac{\partial E_z}{\partial \theta} - \frac{\partial E_\theta}{\partial z}$$

$$\operatorname{curl}_\theta \boldsymbol{E} = \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r}$$

$$\operatorname{curl}_z \boldsymbol{E} = \frac{1}{r}\frac{\partial (rE_\theta)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \theta}$$

Laplacian:

$$\nabla^2 V = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} + \frac{\partial^2 V}{\partial z^2}$$

3. Differential operators for spherical coordinates (mutually perpendicular unit vectors $\boldsymbol{a}_r$, $\boldsymbol{a}_\theta$, $\boldsymbol{a}_\varphi$).

Gradient:

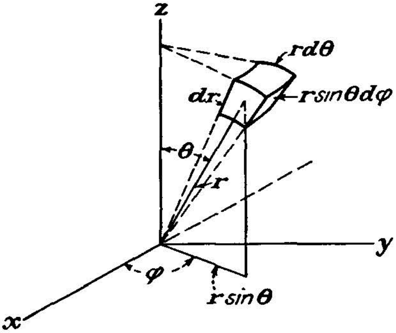
$$-\boldsymbol{E} = \text{gradient } V = \nabla V$$

Components:

$$\operatorname{grad}_r V = \frac{\partial V}{\partial r}$$

$$\operatorname{grad}_\theta V = \frac{1}{r}\frac{\partial V}{\partial \theta}$$

$$\operatorname{grad}_\varphi = \frac{1}{r \sin \theta}\frac{\partial V}{\partial \varphi}$$



Divergence:

$$\operatorname{div} \boldsymbol{E} = \nabla \cdot \boldsymbol{E} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 E_r) + \frac{1}{r \sin \theta}\frac{\partial}{\partial \theta}(\sin \theta \, E_\theta) + \frac{1}{r \sin \theta}\frac{\partial}{\partial \varphi}\left(\frac{\partial E_\varphi}{\partial \varphi}\right)$$

Curl:

$$\operatorname{curl} \boldsymbol{E} = \nabla \times \boldsymbol{E} = \frac{1}{r^2 \sin \theta}\begin{vmatrix} \boldsymbol{a}_r & r\boldsymbol{a}_\theta & r \sin \theta \, \boldsymbol{a}_\varphi \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \varphi} \\ E_r & rE_\theta & r \sin \theta \, E_\varphi \end{vmatrix}$$

Components:

$$\operatorname{curl}_r E = \frac{1}{r \sin \theta}\left[\frac{\partial(\sin \theta \, E_\varphi)}{\partial \theta} - \frac{\partial E_\theta}{\partial \varphi}\right]$$

$$\operatorname{curl}_\theta E = \frac{1}{r \sin \theta}\frac{\partial E_r}{\partial \varphi} - \frac{1}{r}\frac{\partial (rE_\varphi)}{\partial r}$$

$$\operatorname{curl}_\varphi E = \frac{1}{r}\left[\frac{\partial (rE_\theta)}{\partial r} - \frac{\partial E_r}{\partial \theta}\right]$$

Laplacian:

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin\theta}\frac{\partial^2 V}{\partial\varphi^2}$$

4. Differential operators for general orthogonal coordinates (mutually perpendicular unit vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$).

h_1, h_2, and h_3 are scale factors such that an element of length is given by

$$ds^2 = h_1^2\,du_1^2 + h_2^2\,du_2^2 + h_3^2\,du_3^2$$

Gradient:

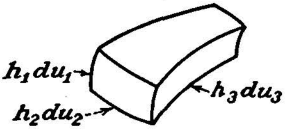
$$-\boldsymbol{E} = \text{gradient } V = \nabla V$$

Components:

$$\text{grad}_{u_1} V = \frac{1}{h_1}\frac{\partial V}{\partial u_1}$$

$$\text{grad}_{u_2} V = \frac{1}{h_2}\frac{\partial V}{\partial u_2}$$

$$\text{grad}_{u_3} V = \frac{1}{h_3}\frac{\partial V}{\partial u_3}$$



Divergence:

$$\text{div } \boldsymbol{E} = \nabla\cdot\boldsymbol{E} = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(h_2h_3E_1) + \frac{\partial}{\partial u_2}(h_3h_1E_2) + \frac{\partial}{\partial u_3}(h_1h_2E_3)\right]$$

where

$$\boldsymbol{E} = \boldsymbol{a}_1E_1 + \boldsymbol{a}_2E_2 + \boldsymbol{a}_3E_3$$

Curl:

$$\text{curl } \boldsymbol{E} = \nabla\times\mathbf{E} = \frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\boldsymbol{a}_1 & h_2\boldsymbol{a}_2 & h_3\boldsymbol{a}_3 \\ \dfrac{\partial}{\partial u_1} & \dfrac{\partial}{\partial u_2} & \dfrac{\partial}{\partial u_3} \\ h_1E_1 & h_2E_2 & h_3E_3 \end{vmatrix}$$

Laplacian:

$$\nabla^2 V = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}\left(\frac{h_2h_3}{h_1}\frac{\partial V}{\partial u_1}\right) + \frac{\partial}{\partial u_2}\left(\frac{h_3h_1}{h_2}\frac{\partial V}{\partial u_2}\right) + \frac{\partial}{\partial u_3}\left(\frac{h_1h_2}{h_3}\frac{\partial V}{\partial u_3}\right)\right]$$

APPENDIX III

A NOTE ON MKS UNITS

In this book there are used rationalized practical mks units. Much has been written on the subject of units. This section is intended to be not an exposition of the topic but rather a group of comments that will aid the student in using mks units.

In the mks system of units, distance is measured in meters, mass in kilograms, and time in seconds. The term "rationalized" means that the defining constants of the system have had the factor 4π included in them in such a way that Maxwell's equations have the simplest possible form. The term "practical" indicates that the common electrical quantities such as potential, current, power, charge, and resistance are expressed in the practical units of volts, amperes, watts, coulombs, and ohms. This latter simplifies things greatly, for no conversion factors need be applied for the common electrical quantities. It may be argued that it would be more appropriate to use rationalized practical cgs units in a book on vacuum tubes than the corresponding mks units because it is easier to think in terms of coulombs per cubic centimeter than in terms of coulombs per cubic meter, etc. The mks units have been used, however, because they are so extensively employed in books and papers on electromagnetic theory and so will ordinarily be reasonably familiar to the student. It is probably a simpler matter to shift a decimal point than to remember two sets of constants.

The basic constants of the rationalized practical mks units are the permeability and dielectric constant of free space, which have values of

$$\mu_0 = 4\pi \times 10^{-7} = 1.2576 \times 10^{-6} \qquad \text{henry per meter}$$

and

$$\varepsilon_0 = \frac{1}{36\pi} \times 10^{-9} = 8.8485 \times 10^{-12} \qquad \text{farad per meter}$$

The dimensions of these units become apparent if one works out the expression for the inductance of a long solenoid and the capacity of a parallel-plate condenser in these units. The resulting expressions are

$$L = \mu N^2 \frac{\text{area}}{\text{length}} \qquad \text{henrys}$$

where N is the number of turns, and

$$C = \varepsilon \frac{\text{area}}{\text{spacing}} \qquad \text{farads}$$

Although the numerical values of the two fundamental constants given above look very awkward, they may be remembered or quickly derived by virtue of their relation to two other well-known physical constants. One of these is the velocity of light, which has the value

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 3 \times 10^8 \qquad \text{meters per sec}$$

The other is the so-called "intrinsic impedance of free space," which is the ratio of the electric- to the magnetic-field strength in a plane-polarized wave,

$$\eta = \frac{E}{H} = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi = 377 \qquad \text{ohms}$$

which by coincidence is the same as the angular frequency of a 60-cycle wave. From the above it is seen that

$$\mu_0 = \frac{\eta}{c}$$

and

$$\varepsilon_0 = \frac{1}{\eta c}$$

In rationalized practical mks units, Maxwell's equations have the form

$$\begin{aligned} \text{div } D &= \rho \\ \text{div } B &= 0 \\ \text{curl } E &= -\dot{B} \\ \text{curl } H &= \dot{D} + J \end{aligned}$$

This set of equations differs notably from the corresponding equations written in Gaussian units by the fact that all the numerical coefficients are unity. In particular, the factors c and 4π do not appear. This means that the factor 4π and c have been absorbed into the constants μ and ε. Unfortunately, if the factor 4π is suppressed in one place it will necessarily crop out in another. In any rationalized system of units the factor 4π will not appear in any relations involving rectangular coordinates, but it will appear in relations involving spherical coordinates. This is just the reverse of the situation encountered with unrationalized units, of which the Gaussian units are an example. Since rectangular components are used more frequently than spherical components in vacuum-tube problems and for that matter in virtually all except antenna-radiation problems, the rationalization seems justifiable.

So far it seems that the rationalized practical mks units achieve some simplifications of formulas in return for some other slight disadvantages. Another factor to be considered is that in the system of units used here magnetic-flux density does not equal magnetic field but rather

TABLE XVII
RELATIONS BETWEEN THE PRINCIPAL PHYSICAL, MAGNETIC, AND ELECTRIC QUANTITIES IN THE PRINCIPAL SYSTEMS OF UNITS

	Rationalized practical mks	Electrostatic	Magnetic
Length, l	1 meter	100 centimeters	100 centimeters
Mass, m	1 kilogram	1,000 grams	1,000 grams
Time, t	1 second	1 second	1 second
Force, F	1 newton	10^5 dynes	10^5 dynes
Work, energy, $\mathcal{E}$	1 watt-second 1 joule 1 meter-kilogram	10^7 ergs	10^7 ergs
Power, W	1 watt	10^7 ergs per second	10^7 ergs per second
Charge, q	1 coulomb	3×10^9 statcoulombs	10^{-1} abcoulomb
Current, I	1 ampere	3×10^9 statamperes	10^{-1} abampere
Electric field, E	1 volt per meter	$\frac{1}{3 \times 10^4}$ statvolt per centimeter	10^6 abvolts per centimeter
Potential difference, or emf, V	1 volt	$\frac{1}{300}$ statvolt	10^6 abvolts
Electric-flux density, D	1 coulomb per square meter	3×10^5	10^{-5}
Magnetic field, H	1 ampere turn per meter	$12\pi \times 10^7$	$4\pi \times 10^{-3}$ oersted
Magnetic-flux density, B	1 weber per square meter	$\frac{1}{3 \times 10^6}$	10^4 gausses
Resistance, R	1 ohm	$\frac{1}{9 \times 10^{11}}$ statohm	10^9 abohms
Inductance, L	1 henry	$\frac{1}{9 \times 10^{11}}$ stathenry	10^9 abhenry
Capacity, C	1 farad	9×10^{11} statfarads	10^{-9} abfarad
Permeability of free space, μ_0	$4\pi \times 10^{-7}$ henry per meter	$\frac{1}{9 \times 10^{20}}$ (seconds per centimeter)^2	Unity
Dielectric constant of free space, ε_0	$\frac{1}{36\pi \times 10^9}$ farads per meter	Unity	$\frac{1}{9 \times 10^{20}}$ (seconds per centimeter)^2

$$B = \mu H$$

It is important to distinguish between B and H. Likewise, electric-flux density does not equal electric field but rather

$$D = \varepsilon E$$

In both the above relations there is a big difference in the numerical values of the flux-density and field factors even for free space.

The price that is paid for reducing the common electrical quantities to practical units is that some other quantities appear in relatively unfamiliar units. Thus the unit of force becomes the newton, which is equal to 10^5 dynes and is sometimes known as the "dyne-five." The magnetic units are a little strange, too. The magnetic field H appears in units of ampere turns per-meter, which, however, makes good physical sense. The magnetic-flux density appears in units of webers per square meter, each one of which is equal to 10^4 gausses. These are not too difficult to remember, however.

The relation between the most commonly used quantities in electrostatic, electromagnetic, and rationalized practical units are given in Table XVII. Quantities in any row are equal.

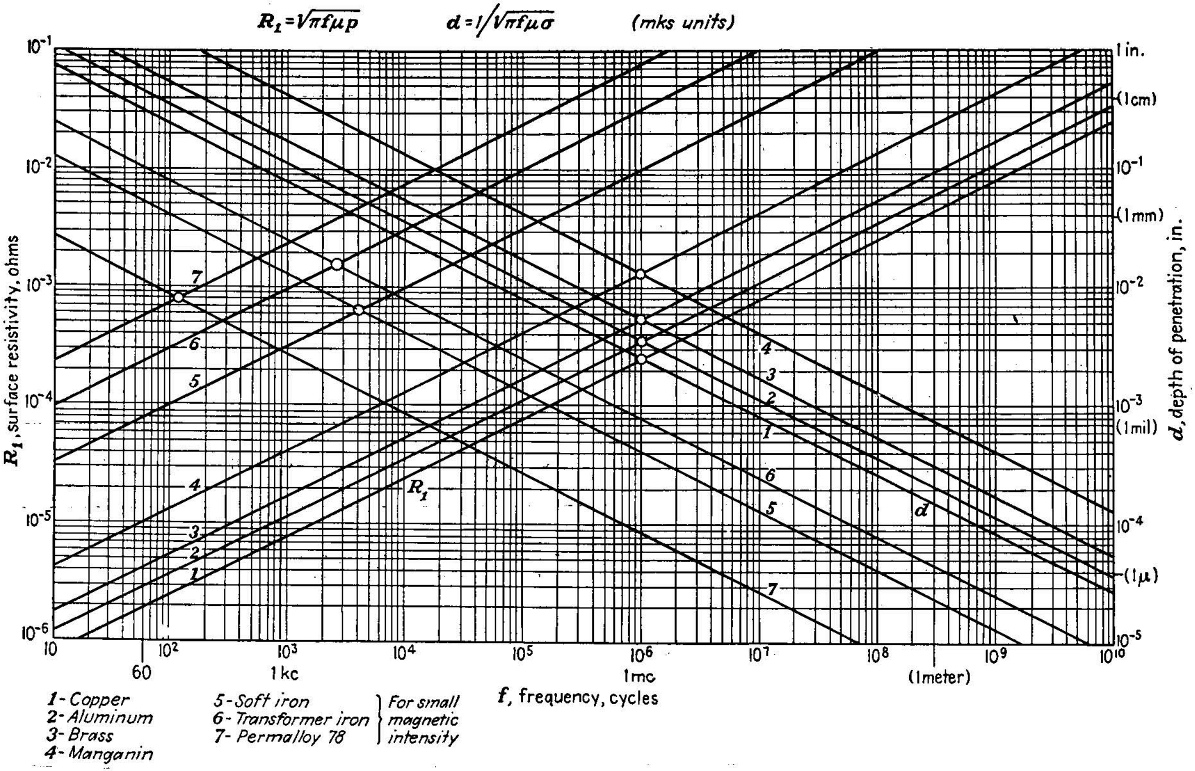
APPENDIX IV

CHARACTERISTICS OF FLUORESCENT SCREENS

RMA* designation	Substance	Activator	Formula	Fluorescent color	Phosphorescence, sec
P1	Zinc silicate	Manganese	$Zn_2\ SiO_4 \cdot Mn$	Green	Medium 0.03–0.05
P2	Zinc sulphide	Copper	$ZnS \cdot Cu$	Blue-green	Long
P3	Zinc beryllium silicate	Manganese	$Zn\ Be\ SiO_3 \cdot Mn$	Yellow-green	Medium 0.05
P4	P3 and zinc sulphide	Silver	$ZnS \cdot Ag + P3$	White	Short 0.005
P5	Calcium tungstate	....	$Ca\ WO_4$	Blue	Very short 5μ sec
P6	Zinc sulphide	Silver	$ZnS \cdot Ag$	....	Medium 0.005
	Zinc cadmium sulphide	Silver	$ZnCdS \cdot Ag$	White	
P7	Zinc sulphide	Silver	$ZnS \cdot Ag$	Blue	Medium 0.005
	Zinc cadmium sulphide	Copper	$ZnCdS \cdot Cu$	Yellow	Long
P11	Zinc sulphide	Silver with a nickel quencher	$ZnS \cdot Ag \cdot Ni$	Blue	Very short 10μ sec.

* Radio Manufacturers Association.

APPENDIX V



Surface resistance and depth of penetration of current resulting from skin effect.

APPENDIX VI

PRINCIPAL PROPERTIES OF THE BESSEL FUNCTIONS[1–4]

Differential equation of the Bessel function:

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right) y = 0$$

Form of solution:

$$y = AJ_n(x) + BN_n(x)$$

where J_n is the nth-order Bessel function of the first kind and N_n is the nth-order Bessel function of the second kind, also known as the "Neumann function."

Series expansion of the Bessel function:

$$J_n(x) = \frac{x^n}{2^n n!}\left[1 - \frac{x^2}{2^2(n+1)} + \frac{x^4}{2^4 2!(n+1)(n+2)} + \cdots + \frac{(-1)^k x^{n+2k}}{2^{n+2k}k!(n+k)!} + \cdots\right]$$

Small-value approximations (x less than 1/10 of first root):

$$J_n(x) = \frac{x^n}{n!2^n} \qquad N_0(x) = -\frac{2}{\pi}\ln_\epsilon \frac{2}{1.781x}$$

$$N_n(x) = \frac{-(n-1)!}{\pi}\frac{2^n}{x} \qquad (n = 1, 2, \cdots)$$

Large-value approximations (x larger than third root):

$$J_n(x) = \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{2n+1}{4}\pi\right)$$

$$N_n(x) = \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{2n+1}{4}\pi\right)$$

---

[1] JAHNKE, E., and F. EMDE, "Tables of Functions," Teubuer, Berlin, 1933.

[2] BURRINGTON, R. S., and C. C. TORRANCE, "Higher Mathematics," pp. 432–442, McGraw-Hill, New York, 1939.

[3] HANSEN, W. W., and V. R. WOODYARD, "A New Principle in Directional Antenna Design," *Proc. I.R.E.*, vol. 26, p. 338 March, 1938.

[4] SMITH, D. B., L. M. RODGERS, and E. H. TRAUB, Zeros of Bessel Functions, *Jour. Franklin Inst.*, vol. 237, pp. 301–303, April, 1944.

Roots of the Bessel function: u_{nm} = mth root of nth-order function.

$u_{01} = 2.405$ $u_{11} = 3.832$ $u_{21} = 5.135$ $u_{31} = 6.380$
$u_{02} = 5.520$ $u_{12} = 7.016$ $u_{22} = 8.417$ $u_{32} = 9.761$
$u_{03} = 8.654$ $u_{13} = 10.173$ $u_{23} = 11.620$ $u_{33} = 13.015$
$u_{04} = 11.792$ $u_{14} = 13.324$ $u_{24} = 14.796$ $u_{34} = 16.223$

Roots of first derivative of Bessel function: u'_{nm} = mth root of nth-order function.

$u'_{01} = 3.832$ $u'_{11} = 1.841$ $u'_{21} = 3.054$ $u'_{31} = 4.201$
$u'_{02} = 7.016$ $u'_{12} = 5.331$ $u'_{22} = 6.706$ $u'_{32} = 8.015$
$u'_{03} = 10.174$ $u'_{13} = 8.536$ $u'_{23} = 9.970$ $u'_{33} = 11.346$
$u'_{04} = 13.324$ $u'_{14} = 11.706$ $u'_{24} = 13.170$ $u'_{34} = 14.586$

Integral definition of the Bessel function:

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos (x \sin \phi - n\phi)\, d\phi$$

Other important relations:

$$\frac{dJ_n(x)}{dx} = -\frac{n}{x} J_n(x) + J_{n-1}(x)$$

$$\frac{dJ_n}{dx} = \frac{n}{x} J_n - J_{n+1}$$

$$\frac{dJ_n}{dx} = \frac{1}{2} J_{n-1} - \frac{1}{2} J_{n+1}$$

$$J_0' = -J_1 \qquad J_1' = J_0 - \frac{J_1}{x}$$

$$\cos (z \sin x) = J_0(z) + 2[J_2(z) \cos x + J_4(z) \cos 4x + \cdots]$$
$$\sin (z \sin x) = 2[J_1(z) \sin x + J_3(z) \sin 3x + \cdots]$$
$$\cos (z \cos x) = J_0(z) - 2[J_2(z) \cos 2x - J_4(z) \cos 4x + \cdots]$$
$$\sin (z \cos x) = 2[J_1(z) \cos x - J_3(z) \cos 3x + \cdots]$$

APPENDIX VII

VALUES OF α^2 AS A FUNCTION OF $\frac{r_c}{r}$ FOR USE IN EQ. (15.63)*

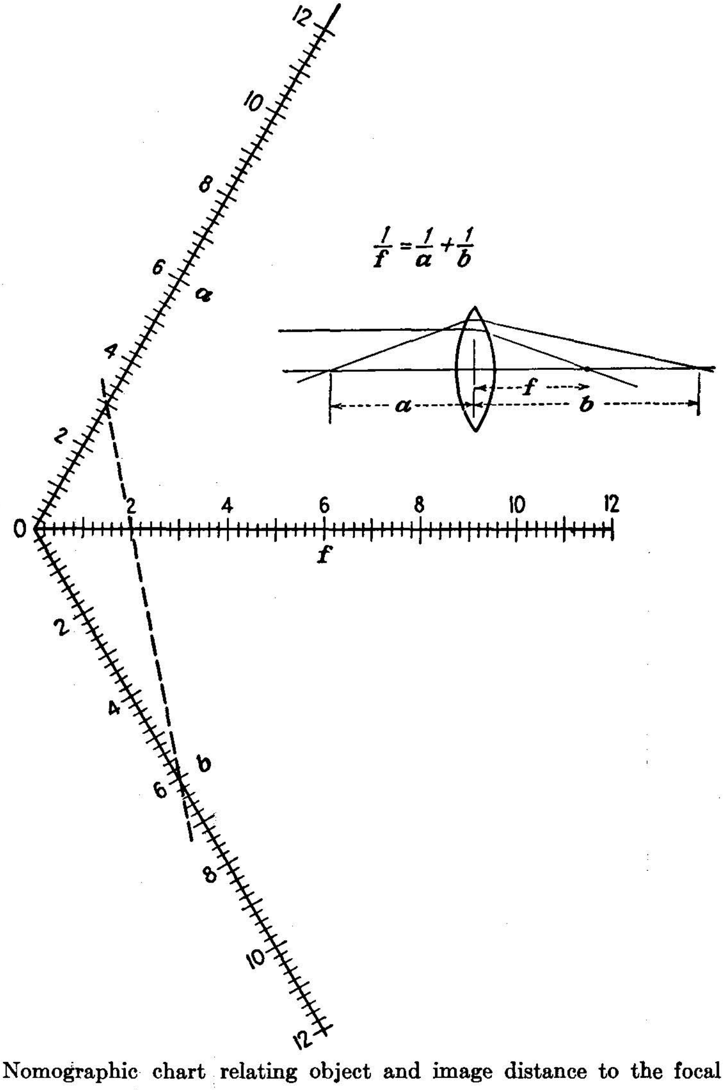
(r_c = radius of emitter; r = radius at any point P; α^2 applies to case where P is outside emitter, $r > r_c$; $(-\alpha)^2$ applies to case where P is inside emitter, $r_c > r$)

$\frac{r}{r_c}$ or $\frac{r_c}{r}$	α^2	$(-\alpha^2)$	$\frac{r}{r_c}$ or $\frac{r_c}{r}$	α^2	$(-\alpha)^2$
1.0	0.0000	0.0000	6.5	1.385	13.35
1.05	0.0023	0.0024	7.0	1.453	15.35
1.1	0.0086	0.0096	7.5	1.516	17.44
1.15	0.0180	0.0213	8.0	1.575	19.62
1.2	0.0299	0.0372	8.5	1.630	21.89
1.25	0.0437	0.0571	9.0	1.682	24.25
1.3	0.0591	0.0809	9.5	1.731	26.68
1.35	0.0756	0.1084	10	1.777	29.19
1.4	0.0931	0.1396	12	1.938	39.98
1.45	0.1114	0.1740	14	2.073	51.86
1.5	0.1302	0.2118	16	2.189	64.74
1.6	0.1688	0.2968	18	2.289	78.56
1.7	0.208	0.394	20	2.378	93.24
1.8	0.248	0.502	30	2.713	178.2
1.9	0.287	0.621	40	2.944	279.6
2.0	0.326	0.750	50	3.120	395.3
2.1	0.364	0.888	60	3.261	523.6
2.2	0.402	1.036	70	3.380	663.3
2.3	0.438	1.193	80	3.482	813.7
2.4	0.474	1.358	90	3.572	974.1
2.5	0.509	1.531	100	3.652	1144
2.6	0.543	1.712	120	3.788	1509
2.7	0.576	1.901	140	3.903	1907
2.8	0.608	2.098	160	4.002	2333
2.9	0.639	2.302	180	4.089	2790
3.0	0.669	2.512	200	4.166	3270
3.2	0.727	2.954	250	4.329	4582
3.4	0.783	3.421	300	4.462	6031
3.6	0.836	3.913	350	4.573	7610
3.8	0.886	4.429	400	4.669	9303
4.0	0.934	4.968	500	4.829	13015
4.2	0.979	5.528	600	4.960	
4.4	1.022	6.109	800	5.165	
4.6	1.063	6.712	1000	5.324	
4.8	1.103	7.334	1500	5.610	
5.0	1.141	7.976	2000	5.812	
5.2	1.178	8.636	5000	6.453	
5.4	1.213	9.315	10000	6.933	
5.6	1.247	10.01	30000	7.693	
5.8	1.280	10.73	100000	8.523	
6.0	1.311	11.46			

---

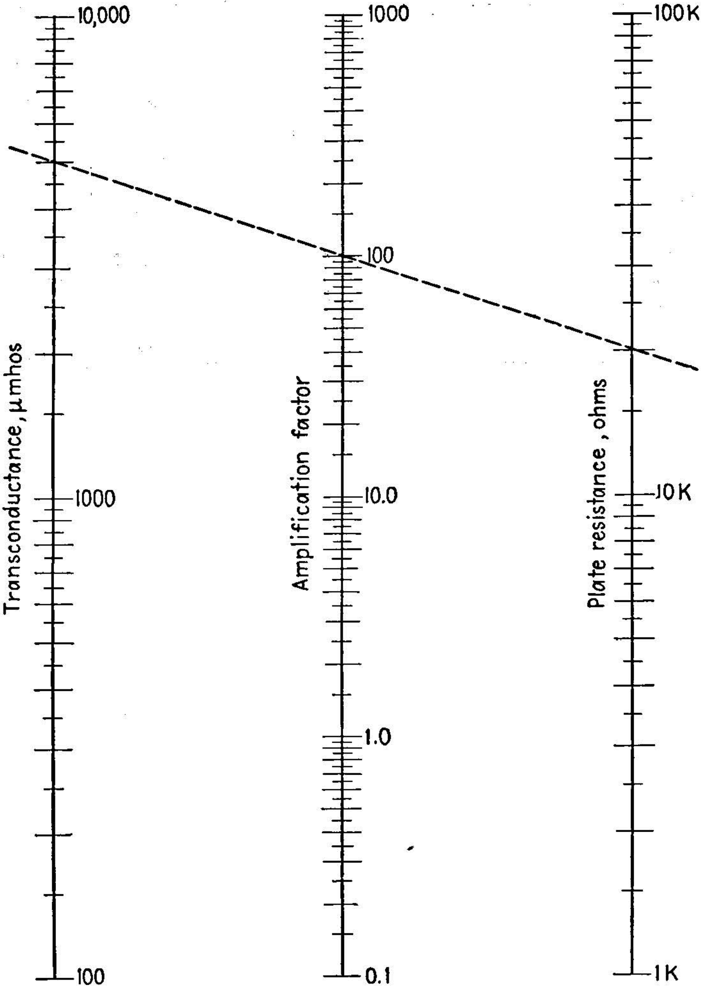
* LANGMUIR, I. L., and K. BLODGETT, Currents Limited by Space Charge between Concentric Spheres, *Phys. Rev.*, vol. 24, p. 53, July, 1924.

APPENDIX VIII



Nomographic chart relating object and image distance to the focal length of a thin lens.

APPENDIX IX



Nomographic chart relating amplification factor, mutual conductance, and plate resistance of a vacuum tube.

APPENDIX X

DESIGNATION OF FREQUENCY BANDS[1]

Title	Abbr.	Wave length	Frequency
Very low frequency	VLF	33.3– 10 km	10– 30 kc
Low frequency	LF	10– 1 km	30– 300 kc
Medium frequency	MF	1,000–100 meters	0.3– 3 mc
High frequency	HF	100– 10 meters	3– 30 mc
Very high frequency	VHF	10– 1 meter	30– 300 mc
Ultra-high frequency	UHF	1– 10 cm	300–3,000 mc
Super-high frequency	SHF	10– 1 cm	3– 30 kmc

---

[1] As announced by Federal Communications Commission, Mar. 2, 1943.

PROBLEMS

CHAPTER 4

4.1. What fraction of the electrons emitted from an oxide coating at a temperature of 1000°K can overcome a retarding voltage of 0.5 volt?

4.2. What is the emission-current density predicted by the emission equation [Eq. (4.3)] for tantulum at 2500°K? What is the corresponding emission-current density of tungsten at 2500°K? At what temperature will the emission-current density of tungsten be five times as great as at 2500°K?

4.3. Using the data given in Table 2, calculate the operating characteristics and life for a 10 per cent evaporation of mass of an ideal tungsten filament having a length of 2 cm and a diameter of 0.25 mm when heated to 2600°K.

a. Power radiated

$$W = W'ld = 263.0 \times 2 \times 0.025 = 13.17 \qquad \text{watts}$$

b. Resistance

$$R = R'\frac{l}{d^2} = 98.66 \times 10^{-6}\frac{2}{0.025^2} = 0.3155 \qquad \text{ohm}$$

c. Filament current

$$I_f = I_f' \times d^{3/2} = 1.632 \times 0.025^{3/2} = 6.45 \qquad \text{amperes}$$

d. Voltage drop

$$V_f = V_f' \times \frac{l}{d^{1/2}} = 161.1 \times 10^{-3} \times \frac{2}{0.025^{1/2}} = 2.04 \qquad \text{volts}$$

e. Emission current

$$I_e = I_e'ld = 2.25 \times 2 \times 0.025 = 0.1125 \qquad \text{ampere}$$

f. Ratio of hot to cold resistance

$$\frac{R_T}{R_{293^\circ}} = 14.12$$

g. Life for 10 per cent reduction in mass

$$\text{Life} = \frac{\text{volume} \times \text{density}}{10(\text{sec per hr})M} \qquad \text{hr}$$

r since $M = M'ld$ and density is 19,

$$\text{Life} = \frac{4.15 \times 10^{-4}d}{M'} = \frac{4.15 \times 10^{-4} \times 0.025}{2.76 \times 10^{-8}} = 376 \qquad \text{hr}$$

4.4. Calculate the operating characteristics and life for a 10 per cent evaporation of mass of an ideal tungsten filament having a length of 1 in. and a diameter of 10 thousandths of an inch when operated at 2850°K.

4.5. Design a tungsten filament ½ in. long that will give an emission of 0.500 ampere and have a life of 1,000 hr. Find the emission efficiency of this filament in milliamperes per watt.

4.6. Calculate and plot the emission efficiency of tungsten filaments in milliamperes per watt over a temperature range of 2000 to 3000°K. Show that the emission efficiency is independent of the length and diameter of the filament.

4.7. Calculate the emission of an ideal tungsten filament whose length is 4 cm and whose diameter is 0.5 mm over the temperature range of 2000 to 3000°K. Plot the results on power-emission paper to show that the curve is a straight line when presented in this form. Plot contours of constant emission efficiency in milliamperes per watt on this same sheet.

4.8. What is the emission-current density from a tungsten filament 1 cm in length and 0.2 mm in diameter operating at a temperature of 2700°K when the surface gradient of potential is 500 volts per cm? What is it when the surface gradient results from a cylindrical electrode surrounding the cathode that is 2 cm in diameter and raised to a potential of 500 volts?

4.9. Determine the emission constants A and b appearing in Richardson's equation for thoriated tungsten and barium–strontium oxide from the intercept and slope of the lines of Fig. 4.5.

4.10. What are the operating characteristics of a thoriated tungsten filament 1 in. in length and 10 thousandths of an inch in diameter operating at 2100°K? Use the data of Table II for heating power and the constants determined in Prob. 4.9 to determine the emission.

4.11. What are the relative emission-current densities of a pure tungsten filament and a thoriated tungsten filament at 2500°K? For the case of a filament 2 cm in length and 0.1 mm in diameter what are the relative emission efficiencies in milliamperes per watt?

4.12. Using coefficients determined as in Prob. 4.10, determine the emission-current density of a barium–strontium oxide coating at 1000°K.

4.13. Using the emission efficiency data of Fig. 4.7, estimate the emission of the oxide-coated cathode of a type 27 tube. The cathode dimensions are 0.065 in. in diameter by 14 mm in length. The cathode is heated by a voltage of 2.5 volts, which produces a current of 1.75 amperes. How does the emission current compare with the rated space-charge-limited current of 5 ma? Suggest how you could measure the emission current without damaging the tube.

CHAPTER 5

5.1. Two particles are suspended by strings of the same length, L, from the same point. Each has a mass m and a charge q. As a result of the forces arising from the like charges the particles will separate. Show that the angle θ which each string makes with the vertical in the equilibrium position is given by

$$4mgL^2 \sin^3 \theta = q^2 \cos \theta (4\pi\varepsilon_0)$$

5.2. Two *point* charges are located as follows:

$$+200 \text{ coulombs at } x = 0, \qquad y = 0 \text{ meters.}$$
$$-100 \text{ coulombs at } x = 1, \qquad y = 0 \text{ meters.}$$

a. *Sketch* a curve showing how potential varies along the line passing through the charges, outside the charges, and between.
b. At what points on the line is the potential zero?
c. At what point on the line is a gradient of potential zero?

5.3. Two parallel *line* charges are spaced 1 meter apart. If the first is located at the point (0,0) and has a positive charge of +2 units per meter and the second is located at (0,1) and has a charge of −1 unit per meter, sketch the relative potential along a line passing through the two line charges and perpendicular to both. If the potential midway between the wires is zero and if it is −100 volts at (0,0.9), where else is it zero? Where is the gradient of potential zero?

5.4. Show that the electric intensity inside of an infinitely long straight cylindrical rod of radius a which has a charge of λ per unit length uniformly distributed throughout its cross section is

$$E_r = \frac{\lambda r}{2\pi\varepsilon_0 a^2}$$

5.5. Obtain the potential plot about two parallel equally charged wires by drawing logarithmically spaced equipotential circles about individual wires, obtaining the potentials at the intersections of the circles by addition, and then drawing equipotential contours through points of the same value of potential. Let the wire diameter be one-twentieth of the spacing between wires, and assume that each wire is charged to +100 volts.

5.6. Work Prob. 5.5 for the case of one wire charged to +100 volts and the other charged to −100 volts.

5.7. Prove that the electric intensity inside a uniformly charged spherical shell is zero.

5.8. What is the gradient of potential between the conductors of a concentric cable whose outer and inner radii are r_2 and r_1, respectively, whose inner-conductor potential is zero, and whose outer-conductor potential is V_1? Find the potential at any radius between the conductors.

5.9. Evaluate the potential at a point that lies a distance c from a uniform spherical distribution of charge of radius a. Let the charge per unit volume of the spherical distribution be p. Show that the resulting potential outside the charge is the same as though the total charge were concentrated at the center of the sphere. Do this by integrating the effects of elements of charge in spherical coordinates.

5.10. Given a linear distribution of charge along a line segment of length l and density of λ coulombs per meter. Show that the potential gradient at a distance a from the end of the line segment of charge along the extended line segment is

$$\frac{dV}{dx} = \frac{-\lambda l}{4\pi\varepsilon_0(a + 1)}$$

Solve this problem by taking a summation of intensities.

5.11. Solve Prob. 5.10 by evaluating the potential at any point along the extended line segment of charge and then taking the derivative.

5.12. Sketch the potential-flux pattern around an exterior right angle of a conductor, *i.e.*, the field in the vicinity of a corner of a long square charged conductor. Make use of the properties listed in the text.

5.13. Calculate and plot equipotential and flux lines outside of a right-angle corner of a conductor by means of the function $W = Z^{1/2}$. Compare the results with the sketch of Prob. 5.12.

5.14. Obtain by integration a solution of Laplace's equation in one dimension for rectangular coordinates. Show that the potential varies linearly with distance while the gradient of potential is constant.

5.15. Obtain by integration a solution of Laplace's equation in polar coordinates when there is no variation of potential with angle. Show that potential varies logarithmically with radius while the gradient of potential varies inversely as radius.

5.16. A concentric conductor cable consists of a circular inner conductor 2 cm in diameter inside of an outer conductor of square cross section that measures 4 cm per side. Assume that the inner conductor is at a direct potential of 100 volts while the outer conductor is at a direct potential of 0 volts. Sketch flux and potential lines in the space between conductors. Estimate the gradient of potential at

a. The surface of the center conductor opposite a corner of the outer conductor.
b. The surface of the center conductor closest to the outer conductor.
c. The surface of the outer conductor closest to the center conductor.
d. At a corner of the outer conductor.

Estimate the capacity per unit length of line by taking the ratio of charge to potential. Remember that each flux line terminates on one unit of charge when the field plot is given by curvilinear squares and the adjacent equipotentials are separated by unit potential.

5.17. One section of a plane-electrode triode is approximated by the following potential lattice

100 volts	100 volts	100 volts	100 volts	100 volts
a	*b*	*c*	*b*	*a*
d	*e*	*f*	*e*	*d*
g	*h*	−10	*h*	*g*
i	*j*	*k*	*j*	*i*
0 volts	0 volts	0 volts	0 volts	0 volts

The top row represents the plate at a potential of 100 volts. The bottom row represents the cathode at a potential of 0 volts. The grid is represented by the number in the fourth row of the third column and is at −10 volts. The points *a*, *d*, *g*, *i* are midway between grid wires on a line of symmetry. Find potentials at the lettered points by first assuming reasonable values and then correcting several times around by means of Eq. (5.44).

5.18. Prove that the function $W = \ln Z$ is analytic for finite values of Z other than zero and infinity.

5.19. Prove that the function $W = Z^4$ is analytic for finite values of Z other than infinity.

5.20. Separate the function $W = \ln Z$ into real and imaginary parts U and V, respectively. Show that $U =$ const and $V =$ const form orthogonal families of curves. Show that both $U =$ const and $V =$ const are solutions of Laplace's equation. Show also that the Cauchy-Riemann conditions are satisfied.

5.21. The transformation $W = Z^{1/2}$ transforms the upper half of the Z plane into the first quadrant of the W plane, giving rise to the field configuration associated with an inside right-angled corner. Show that the equipotential lines inside the right-angled corner are given by rectangular hyperbolas. Show that the flux lines are also hyperbolas. Show that the gradient of potential along the u and v axes in the W plane is normal to the axis and proportional to the distance from the origin.

5.22. Use the function $W = Z^{1/4}$ to obtain the flux and potential plot for an internal 45-deg corner between two plane conducting surfaces. Do this by letting $W = R\angle\phi$ and $Z = r\angle\theta$ and then transforming the lines $x = r \cos \theta =$ const and $y = r \sin \theta =$ const by means of the transforming function.

5.23. The function $W = Z^2$ transforms the upper half of the Z plane into the entire W plane and gives the potential configuration about the edge of a sheet conductor corresponding to the positive real axis of the W plane. Find the equations of the potential and flux lines in the W plane. Show that these are orthogonal sets of parabolas. Find the gradient of potential at any point in the W plane.

5.24. Show that the transformation $W = \ln \sin Z$ gives the field configuration of a row of parallel equidistant line charges having the same charge, *i.e.*, the field about a grid of parallel wires.

5.25. Show that the transformation $W = \ln \tan Z$ gives the potential about a row of parallel equidistant line charges with alternate positive and negative charges.

5.26. Show that the function $W = \ln \left(\dfrac{Z - a}{Z + a}\right)$ gives the potential and flux pattern about a two-wire transmission line having wires located at $(a,0)$ and $(-a,0)$ in the Z plane. Find the equations for the flux and potential lines to show that these are orthogonal families of circles.

5.27. Show that the function $W = \ln (Z^n - 1)$ gives the field about n line charges uniformly distributed around the unit circle, *i.e.*, the field of a squirrel-cage grid.

CHAPTER 6

6.1. An electron is liberated with zero velocity at the cathode of a plane-electrode diode whose electrode spacing is 5 mm and whose cathode-plate potential difference is 100 volts. With what velocity does the electron strike the plate? What energy has the electron acquired in moving from cathode to plate? How long does it take the electron to make the trip? If a singly charged hydrogen ion and a doubly charged oxygen ion are liberated at the plate, give the velocity, energy, and time associated with their arrival at the cathode.

6.2. An electron is liberated with zero velocity at the cathode of a cylindrical-electrode diode whose cathode radius is 0.2 cm and whose concentric plate radius is 1.0 cm. The cathode-plate potential difference is 100 volts. With what velocity and energy does the electron arrive at the plate? How long does the trip take? If a singly charged hydrogen ion and a doubly charged oxygen ion are liberated at the plate, with what velocity and energy and at what time will they arrive at the cathode? Refer to Fig. 8.14 for time factors.

6.3. An electron with a velocity acquired by falling through 10 volts is injected into a region with a retarding potential gradient of 2 volts per cm. How far will the electron travel before having its direction reversed? How long will it take the electron to return to its starting point? With what velocity will the electron return?

6.4. An electron is injected into the region between two parallel planes separated 1 cm and differing in potential by 50 volts, the resultant field being retarding. If the electron has a velocity acquired by falling through 100 volts of potential, find the point at which the electron will strike one of the electrodes, the velocity components with which it will strike, and the time of flight when the angle with which the electron enters is 0, 30, 45, and 60 deg with the normal to the electrodes. Tabulate results.

6.5. Solve Prob. 6.4 when the potential between the plates is 50 volts and the field is accelerating.

6.6. In Prob. 6.4 find the location of points closest to the second plate on trajectories of those electrons which are returned to the first plate.

6.7. An electron is injected at an angle of 60 deg with the normal to the plates into a region between two parallel plates separated 1 cm and having a retarding field of 20 volts per cm. There is a small hole in the second plate displaced 3 cm from the point at which the electron enters. Assuming that the transverse component of electron velocity is in line with the point of entrance and the hole in the second plate, with what velocity must the electron enter the retarding field region in order to pass through the hole in the second plate?

6.8. Derive Eq. (6.25).

6.9. Through what potential must an electron fall in order to be accelerated to 0 1, 0.5, 0.9, 0.95 of the velocity of light? What is the relative transverse mass of the electron at each of these velocities?

6.10. At what velocity is the transverse mass of an electron increased 1, 10, and 100 per cent? What are the corresponding accelerating potentials?

6.11. Derive Eq. (6.38).

6.12. Calculate and plot curves of the transverse and longitudinal mass of an electron as a function of $\frac{v}{c}\cdot$

6.13. Calculate and plot curves of the transverse and longitudinal mass of an electron relative to the rest mass as a function of potential.

6.14. Derive an expression for the deflection of a cathode-ray-tube beam by a set of deflecting plates, the expression to include the relativity correction for mass and velocity. Express the deflection as a fraction relative to the deflection in the absence of relativity effects.

6.15. Calculate and plot a curve of electron velocity in equivalent volts required to produce a circular path 1 cm in diameter when an electron is moving in a magnetic field ranging in intensity from 100 to 10,000 gausses.

6.16. Singly ionized lithium atoms with atomic weights of 6 and 7 are accelerated by a potential of 2,000 volts and then injected into a region of constant transverse magnetic field of density 800 gausses. The atoms are allowed to traverse a half circle before striking a photographic plate. What will be the separation of the marks on the photographic plate corresponding to the two isotopes of lithium?

6.17. An electron is accelerated through a given potential and then injected perpendicular to the elements of a cylinder, 10 cm in diameter, that has a constant magnetic field of strength 10 gausses parallel to its axis. There is a hole in the cylinder a quarter of a full circumference around the cylinder on a circle at which the electron enters. Through what potential must the electron be accelerated before entering the cylinder in order to pass out of the cylinder through this hole? There is a second hole a quarter of a circumference around the cylinder but displaced 3 cm axially along the tube. With what potential and at what angle with the axis must an electron directed toward the axis enter the cylinder in order to pass out through this second hole?

6.18. In a cyclotron a uniform magnetic field is used to cause ions to move in segments of a circular arc. Every half revolution the ions are subjected to an accelerating gradient of potential at the gaps of two D-shaped electrodes so that the radius after each semicircle of motion is greater than before. The accelerating field is supplied by a r-f voltage impressed upon the two D's and appears as an alternating field across the gap. The frequency of the field is regulated so that the ions cross the gap twice each cycle. If the magnetic-flux density is 10,000 gausses, what must the frequency of the applied voltage be when singly charged light hydrogen ions are used (atomic weight unity)? If each passage across the gap increases the energy of the ions by 40,000 volts, how many such passages are required to produce a 2,000,000-volt particle? What will be the diameter of the last semicircle of circular motion?

6.19. What will be the final diameter of the path of a 2,000,000-volt heavy hydrogen ion (atomic weight 2) and what will be the frequency of the applied voltage for the cyclotron of Prob. 6.18? Assume the same magnetic field and energy increase per gap passage.

6.20. What will be the final diameter of the path of a 2,000,000-volt argon ion and what must be the frequency of the voltage producing the accelerating field for the cyclotron of Prob. 6.18? Assume the same magnetic field and energy increase per gap passage.

6.21. An electron's velocity is x-directed in a region of uniformly directed electric field of strength 50 volts per cm and uniform z-directed magnetic field of strength 500 gausses. What must the electron velocity in equivalent volts be in order that its net deflection is zero?

6.22. An electron is emitted with zero velocity from a plane surface where it is subjected to an accelerating gradient of field of strength 50 volts per cm and a transverse magnetic field of 500 gausses. What is the maximum travel in the direction of the electric field in the resulting cycloidal path? What is the velocity at this point of maximum separation from the plane of emission? Where does the electron again return to the plane of emission? What is the elapsed time between departure and return to the emission plane?

6.23. Given the field conditions of Prob 6.22, but with an electron injected with a velocity equivalent to 20 volts normal to the plane. Find the position and velocity with which the electron again returns to the plane.

6.24. A diode consists of a straight filamentary cathode of radius r_c surrounded by a concentric circular plate of radius r_p. If the plate voltage is low enough, the magnetic field of the filament current may cause the electrons to curve strongly enough in their paths so that the tube will be cut off. Derive an expression for cutoff in such a tube in terms of the plate potential, the filament current, and the cathode and plate radius.

CHAPTER 7

7.1. Consider an idealized type 210 plane-electrode triode for which $d_{cg} = 0.050$ in., $d_{gp} = 0.075$ in., $a = 0.050$ in., and $r_g = 0.0025$ in. Using the low-mu formulas, calculate and plot potential profiles along lines perpendicular to the plane electrodes and passing (1) through a grid wire and (2) midway between grid wires for

a. Grid at twice cutoff voltage.
b. Grid at cutoff voltage.
c. Grid at half cutoff voltage.
d. Grid at zero potential.
e. Grid positive and at its "natural" potential.
f. Grid positive and at plate potential.

Assume a plate potential of 100 volts.

7.2. Find the diameters of the cathode, grid, and plate cylinders in the Z-plane equivalent of the W-plane triode representation of the tube whose dimensions are given in Prob. 7.1. Use the transformation of Eq. (7.3).

7.3. A cylindrical-electrode triode has a cathode diameter of 0.020 in. and a plate diameter of 0.750 in. There are 10 grid wires each of 0.012 in. diameter arranged to form a squirrel cage of grid wires evenly spaced around a grid-wire circle of diameter 0.262 in.

a. Calculate the amplification factor of the tube.
b. Calculate the equivalent-diode radius of the tube.
c. Calculate the interelectrode capacities of the active portion of the tube if this is 1 in. long.

7.4. Calculate and plot potential profiles of the cylindrical-electrode triode of Prob. 7.3 in planes through the axis and (1) through a grid wire and (2) midway between grid wires for a plate potential of 100 volts and

a. Grid at twice cutoff potential.
b. Grid at cutoff potential.
c. Grid negative but at half the cutoff potential.
d. Grid at zero potential.
e. Grid positive at its "natural" potential.
f. Grid at plate potential.

7.5. Find the diameters of the cathode, grid, and plate cylinders in the Z-plane equivalent of the triode of Prob. 7.3 if this latter be considered the W-plane configuration. Use the transformation of Eq. (7.15).

7.6. A plane-electrode triode has a cathode-grid spacing of 1 mm, a screening fraction of 0.14, grid-wire diameter of 0.1 mm, and a grid-plate spacing of 2.5 mm. Determine the amplification factor and the equivalent-diode spacing.

7.7. A cylindrical-electrode triode has the dimensions of the tube of Prob. 7.3 except that there are 14 grid wires evenly spaced around the grid-wire circle, instead of 10. Calculate the amplification factor and equivalent-diode radius.

7.8. A plane-electrode triode is to have an amplification factor of 10. If the screening fraction is $\frac{1}{8}$ and there are 50 grid wires per in., specify the grid-wire radius and the grid-plate spacing.

7.9. A plane-electrode triode has a grid-plate spacing of 0.050 in. and a square mesh grid of 0.005-in.-diameter wire spaced 0.015 in. Find the amplification factor of the tube.

7.10. A cylindrical-electrode triode has a plate radius of 0.500 in. and a grid consisting of parallel rings of 0.250 in. diameter and of 0.005-in. wire spaced 0.015 in. There are four grid-ring supports of 0.010-in. wire parallel to the axis of the tube and evenly spaced around the grid. Find the amplification factor of the tube. Cathode diameter is 0.10 in.

7.11. A cylindrical triode has a cathode diameter of 0.10 in. and a plate diameter of 0.500 in. The grid is a helix of 0.01-in.-diameter wire wound so that the largest circular cylinder that can be passed through it is 0.245 in. in diameter. The helical grid has a pitch of 0.08 in. between turns. There are two support wires for the grid of 0.025-in. wire parallel to the axis of the tube. Determine the amplification factor of the tube.

7.12. A plane-electrode triode has a grid-cathode spacing of 8 mils, a grid-wire spacing of 16 mils, grid-wire radius of 1 mil, and a grid-plate spacing of 20 mils. Determine the variation of amplification factor along the cathode. What are the maximum, minimum, and average values of amplification factor that appear? How do these compare with the values of amplification factor that assume large cathode-grid spacing?

7.13. Suggest means of measuring the amplification factor of a triode, given a current-flow model containing a suitable electrolyte.

7.14. Prove that the amplification factors of two geometrically similar tubes are equal.

7.15. From a comparison of Eqs. (7.33) and (7.43) obtain expressions for the cathode-grid and cathode-plate capacities of the fundamental triode of Fig. 7.1*a*. From a comparison of Eqs. (7.34) and (7.44) obtain an expression for the grid-plate capacity.

7.16. From the results of Prob. 7.15 and the transformation of Eqs. (7.4) and (7.5) obtain expressions for the interelectrode capacities of a plane-electrode triode per unit of area.

7.17. From the results of Prob. 7.15 and the transformation of Eqs. (7.16) and (7.17) find the interelectrode capacities of a cylindrical-electrode triode per unit of axial length of structure.

7.18. Find the interelectrode capacities per unit area of the triode of Prob. 7.1. Consider that the grid and plate exist only on one side of the cathode.

7.19. Calculate the interelectrode capacities per unit of axial length of the cylindrical-electrode triode of Prob. 7.7.

7.20. Calculate the capacity per unit length of a five-wire transmission line made of wires of 3-mm-diameter wire. Four of the wires are located at the corners of a square whose dimension is 10 cm on an edge in the cross-sectional view and are connected together. The other wire is located at the center of the square and acts as a return wire. From the capacity per meter determine the characteristic impedance of the line neglecting losses, using the relation that the characteristic impedance in ohms is the reciprocal of the product of the capacity per unit length in farads per meter and the velocity of propagation in meters per second.

7.21. Given a plane-electrode triode with the dimensions of the tube of Prob. 7.1 except that the diameter of the grid wires is twice as large. Calculate the amplification factor by the formula of Vodges and Elder and by the Ollendorf second and third approximations, and compare results.

7.22. Given a cylindrical-electrode triode with the dimensions of the tube of Prob. 7.3 except that the grid wires are twice as large in diameter. Calculate the amplification factor by the formulas of Vodges and Elder and the Ollendorf second and third approximations, and compare results.

7.23. Derive the amplification-factor formula given in Fig. 7.19*a* for the electrode geometry shown.

7.24. Derive the amplification-factor formula given in Fig. 7.19*b* for the electrode geometry shown.

7.25. Derive the amplification-factor formula given in Fig. 7.19*c* for the electrode geometry shown.

7.26. Derive the amplification-factor formula given in Fig. 7.19*d* for the electrode geometry shown.

CHAPTER 8

8.1. In an ideal plane-electrode diode whose emission is space-charge-limited, the cathode-plate separation is 2 mm, and the potential difference is 100 volts. Find the transmitted-current density, the space-charge density at the plate, and the velocity of the electrons arriving at the plate. Find also the power dissipated per unit area of plate surface and the gradient of potential at the plate.

8.2. What must be the cathode-plate spacing of an ideal plane-electrode diode in order that the transmitted current per square inch be 250 ma when the potential difference between cathode and plate is 200 volts?

8.3. Calculate and plot curves similar to those of Fig. 8.6 for a cylindrical diode whose ratio of plate to cathode radius is 10.

8.4. Given an ideal cylindrical diode whose cathode diameter is 2 mm and whose plate radius is 1 cm. Find the transmitted current per centimeter of axial length for a potential difference of 100 volts. Find also the velocity with which the electrons arrive at the plate, the space-charge density at the plate, the gradient of potential at the plate, and the power dissipated per centimeter of axial length at the plate.

8.5. Find the potential, gradient of potential, electron velocity, and space-charge density midway between cathode and plate in the diode of Prob. 8.4.

8.6. Solve Prob. 8.4 on the assumption that the outer electrode is the cathode and the inner electrode is the plate, dimensions and potentials being otherwise unchanged.

8.7. Calculate and plot curves showing the location and magnitude of the maximum gradient of potential in a cylindrical diode as a function of the ratio of plate radius to cathode radius when the inner electrode is the cathode.

8.8. Calculate and plot curves similar to those of Fig. 8.6*a* but for a cylindrical diode whose outer electrode is the cathode and whose inner is the plate.

8.9. Given a diode whose electrodes are concentric spheres, the inner being the cathode and the ratio of diameters being 2 to 1. If the plate diameter is 2 cm, what will be the plate current for a potential difference of 100 volts?

8.10. Solve Prob. 8.9 with the outer electrode considered the cathode and other conditions unchanged.

8.11. A plane-electrode triode has the dimensions of the tube of Prob. 7.1. Calculate the mutual conductance for a plate potential of 100 volts and a grid potential of half the cutoff value. Determine also the plate current per square inch under these conditions.

8.12. A plane-electrode triode has a grid-plate spacing of 30 mils. Grid wires are spaced 15 mils, and the screening fraction is $\frac{1}{10}$. What must be the cathode-grid spacing to give a mutual conductance of 5,000 micromhos per in.^{2} if the plate voltage is 200 volts and the grid voltage is 1 volt negative?

8.13. Calculate the equivalent-diode spacing of the tube of Prob. 7.1 for a plate voltage of 100 volts and a grid voltage of 2 volts negative by the formulas of Eq. (7.53) and Eq. (8.45), and compare results.

8.14. Derive Eqs. (8.49) and (8.50).

8.15. From Eq. (8.49) obtain an expression for the equivalent-diode radius of a cylindrical triode. Calculate the equivalent-diode radius of the triode of Prob. 7.3 by this formula, and compare with the result obtained by using Eq. (7.58).

8.16. Calculate the mutual conductance and plate current for the tube of Prob. 7.3, assuming that the structure is 1 in. long and that the plate potential is 500 volts while the grid potential is -20 volts.

8.17. It is desired to design a triode for high-power audio service. Assume an ideal cylindrical structure. Assume that the cathode is to be 2 mm in diameter and the plate to be 1 cm in diameter. The electrode structure is to be 1 cm in length. What must be the grid dimensions in order that the tube will have ar amplification factor of 100 and an average mutual conductance of 5,000 micromhos when the plate potential is 500 volts?

8.18. Given a plane-electrode diode whose cathode is emitting electrons having an average velocity such that they can overcome a retarding potential of 2 volts. Let the electrode spacing be 5 mm, and let the current that reaches the plate be one-tenth of the emitted current. Find the location of the potential minimum and the magnitude of the plate-current density for a plate 25 volts more positive than the cathode. Use the relations of Eqs. (8.58) to (8.62).

8.19. The cathode of a plane-electrode diode is oxide-coated and operates at a temperature of 1000°K. What is the transmitted-current density to a plate at a potential that is 20 volts positive relative to cathode? Find the location and magnitude of the potential minimum. Find also the fraction of the emitted current that is transmitted to the plate.

8.20. A cylindrical diode is 1 in. long and has a plate diameter of ½ in. and a tungsten filament whose diameter is 5 mils. Neglecting end effects and initial velocities of electrons, calculate the plate current when the plate is 20 volts positive with respect to the negative end of the filament and the direct voltage drop along the filament is 10 volts. Calculate the plate current when the filament is excited by an alternating voltage whose rms value is 10 volts and one end of the filament is grounded. How does this differ from the current resulting when the filament is heated by alternating current with the same voltage drop but with the center tap of the exciting transformer grounded?

8.21. Derive Eq. (8.85).

8.22. Carry out the steps leading to Eqs (8.22) to (8.24).

CHAPTER 9

9.1. Three triodes with constants as follows are operated in parallel:

$$\begin{aligned} \mu_1 &= 10 & G_{m1} &= 2{,}000 \text{ micromhos} \\ \mu_2 &= 12 & G_{m2} &= 5{,}000 \text{ micromhos} \\ \mu_3 &= 30 & G_{m3} &= 3{,}000 \text{ micromhos} \end{aligned}$$

Calculate the equivalent amplification factor, mutual conductance, and plate resistance.

9.2. A plane-electrode triode has the following dimensions:

$$\begin{aligned} d_{cg} &= 40 \text{ mils} & r_g &= 2 \text{ mils} \\ a &= 30 \text{ mils} & d_{gp} &= 60 \text{ mils} \end{aligned}$$

Calculate the current-division factor. Calculate the ratio of plate to grid current when the grid and plate are both positive and the plate potential is five times as great as the grid potential, assuming that there is negligible secondary emission.

9.3. Estimate the current-division factor of the cylindrical triode having the dimensions of the tube of Prob. 7.3.

9.4. Calculate and plot contours of constant plate and grid current per square inch of electrode structure of the triode of Prob. 7.1. Let grid voltage range from −100 to +100 volts. Let plate voltage range from −500 to +500 volts. Show constant current contours in all four quadrants. Assume that secondary emission from both grid and plate is negligible.

CHAPTER 10

10.1. A plane-electrode beam-power tube has the following electrode dimensions:

$$\begin{aligned} d_{cg} &= 20 \text{ mils} & a &= 20 \text{ mils} \\ d_{gs} &= 30 \text{ mils} & r_g &= 1 \text{ mil} \\ d_{gp} &= 70 \text{ mils} & r_s &= 1 \text{ mil} \end{aligned}$$

Calculate the amplification factor by the formula of Eq. (10.11). Compare this value with that obtained from the product of the triode mu's as explained in Sec. 10.2.

10.2. In an idealized beam-power tube the density of the current injected into the screen-grid–plate region is 10 ma per cm². If the screen grid is at a potential of 100 volts positive and the plate is at a potential of 10 volts negative, at what point will the electrons come to rest and reverse direction. The screen-grid–plate spacing is 1 cm.

10.3. In the beam-power tube of Prob. 10.1 the plate voltage is raised to 10 volts positive. For the same injected-current density determine whether a type *B* distribution of potential is possible.

10.4. For the beam-power tube of Prob. 10.1 let the screen-grid potential be 100 volts positive, the plate voltage be 10 volts positive, and the injected-current density be variable. The potential distribution is of type *B*. Find the location of the virtual cathode when the current transmitted to the plate is 0.25, 0.5, and 0.75 of the injected current.

10.5. A beam-power tube has a screen-grid potential of 300 volts and a plate potential of 60 volts. The potential distribution is of type *C* with a potential minimum of 30 volts. What must be the screen-plate distance for an injected current of 48.5 ma per cm²?

10.6. A plane-electrode beam-power tube has a screen-grid–plate spacing of 0.5 cm. Screen grid and plate are kept at a potential of 50 volts. Indicate the position of the potential minimum or virtual cathode as the injected-current density is increased from zero to 50 ma per cm² and then reduced to zero again.

10.7. A beam-power tube has its plane screen grid and plate separated a distance of 0.8 cm. Plot a curve of current density transmitted to the plate against injected-current density as the injected-current density is increased from zero to 50 ma per cm² and reduced to zero again, when the screen grid is at a potential of 100 volts and the plate is at a potential of 50 volts.

10.8. A plane-electrode beam-power tube has its screen grid and plate separated a distance of 0.8 cm. Let the screen-grid potential be 100 volts and the injected-current density be held constant at 10 ma per cm². Plot a curve of plate current against plate voltage as plate voltage is raised from zero to 100 volts and then reduced to zero again.

CHAPTER 11

11.1. Derive an expression similar to that of Eq. (11.6) giving the maximum potential between suppressor-grid wires when the plate potential has the general value V_p not equal to V_2. Let $V_3 = 0$.

11.2. Justify Eq. (11.9) qualitatively.

11.3. An idealized plane-electrode pentode has the following electrode dimensions:

$$\begin{array}{lll} d_{c1} = 10 \text{ mils} & a_1 = 12 \text{ mils} & r_1 = 1.5 \text{ mils} \\ d_{12} = 60 \text{ mils} & a_2 = 18 \text{ mils} & r_2 = 1.75 \text{ mils} \\ d_{23} = 120 \text{ mils} & a_3 = 60 \text{ mils} & r_3 = 2.5 \text{ mils} \\ d_{3p} = 80 \text{ mils} & & \end{array}$$

Calculate the ratio of plate to screen-grid current.

11.4. For the pentode of Prob. 11.3 determine the cathode charge per unit area and the charge per unit length of each of the grids for the following potential values:

$$V_c = V_3 = 0 \qquad V_1 = -1 \text{ volt} \qquad V_2 = 200 \text{ volts} \qquad V_p = 250 \text{ volts}$$

11 5. For the pentode of Prob. 11.3 calculate the electrostatic amplification factors μ_{1p} and μ_{12}.

11.6. Assuming that the value of m in Eq. (11.1) is 0.2, calculate the true amplification factor of the pentode of Prob. 11.3.

11.7. Calculate the mutual conductance of the pentode of Prob. 11.3 for the electrode potentials of Prob. 11.4 and for an m of 0.2.

11.8. Calculate the plate resistance of the pentode of Prob. 11.3 for the electrode potentials of Prob. 11.4 and an m of 0.2.

11.9. Plot curves similar to those of Fig. 11.12 for the pentode of Prob. 11.3, assuming values of a_3 of 30 and 90 as well as 60 mils. Plot curves with $\frac{V_p}{V_2}$ as abscissa and $\frac{V_p - V_{\min}}{V_2}$ as ordinate. Note that the $V_{\min}$ involved here is the $V_{3\,\max}$ of Eq. (11.6).

11.10. Plot curves similar to those of Fig. 11.13 for the pentode of Prob. 11.3. Let d_{2p} have a constant value of 200 mils, but plot curves for d_{3p} equal to 40 and 120 as well as 80 mils. Plot curves with $\frac{V_p}{V_2}$ as abscissa and $\frac{V_p - V_{\min}}{V_2}$ as ordinate.

11.11. Plot curves similar to those of Fig. 11.18 showing the distribution of the sidewise component of velocity of electrons scattered by the three grids of the pentode of Prob. 11.3 operating with the electrode potentials of Prob. 11.4.

11.12. From the results of Prob. 11.11 calculate and plot the plate-voltage–plate-current characteristic.

CHAPTER 12

12.1. Find the rms voltage and current associated with the thermal-agitation noise in a 10,000-ohm resistor over a band width of 50,000 cycles.

12.2. What is the available noise power from the resistor of Prob. 12.1?

12.3. What is the noise power available from a 200-ohm resistor over a band width of 10 mc?

12.4. What is the noise power available from a parallel combination of a resistance of 50,000 ohms at 20°C and a resistance of 100,000 ohms at 200°C over a band width of 2 mc?

12.5. What is the noise voltage associated with a parallel combination of a 20,000-ohm resistor and a 0.1-microfarad condenser over a band width of 50,000 cycles?

12.6. The relative power gain of a resistance-capacity coupled amplifier is given by

$$G_p(f) = \frac{G_p(f_m)}{\left(1 + \frac{f_1^2}{f^2}\right)\left(1 + \frac{f^2}{f_2^2}\right)}$$

where $G_p(f_m)$ is the mid-frequency power gain, f_1 is the low-frequency 70.7 per cent point, and f_2 is the high-frequency 70.7 per cent point. What is the equivalent band width from Eq. (12.10)? Assume $\frac{f_2}{f_1} > 100$.

12.7. What direct current is required in a diode whose emission is temperature-limited to give an rms noise current of 20 microamperes over a band width of 5 mc?

12.8. What is the rms noise current in a diode whose emission is space-charge-limited when the cathode temperature is 1000°K, the band width is 1 mc, and the plate resistance is 10,000 ohms?

12.9. What is the rms noise current in a diode whose emission is space-charge-limited when the plate current is 10 ma, the plate voltage is 50 volts, and the band width is 0.5 mc?

12.10. What resistance in series with the grid circuit of a triode will produce as much noise in the plate circuit as does the tube itself if the mutual conductance is 5,000 micromhos, the amplification factor is 50, the cathode-grid spacing is 12 mils, and the grid-plate spacing is 25 mils?

12.11. What resistance in series with the grid of a triode will produce as much noise in the plate circuit as does the gas in a triode when the grid-circuit resistance is 1 megohm, the plate current is 10 ma, the positive-ion grid current is 0.1 microampere, and the mutual conductance of the tube is 2,000 micromhos?

12.12. What is the resistance whose noise when placed in series with the grid of an ideal pentode produces the same effect as does the actual tube when the mutual conductance is 6,000 micromhos, the plate current is 10 ma, and the screen current is 2 ma?

12.13. What is the noise figure of a secondary-emission multiplier tube using six stages of multiplication when the secondary-emission ratio per stage is 5?

12.14. What is the noise figure of an intermediate-frequency amplifier having a pentode input stage operating from a resistance of 600 ohms over a band width of 3 mc? The mutual conductance of the tube is 7,000 micromhos, the plate current is 10 ma, and the screen current is 2 ma. If the output impedance of the first stage is 2,000 ohms and the second tube in the amplifier has the same characteristics as the first, will it contribute appreciably to the output noise? What is the over-all noise figure including the effect of the second stage?

12.15. Derive an expression for the noise figure of three stages of amplification in cascade.

12.16. A receiver uses a crystal mixer without r-f preamplification. If the crystal conversion gain is 0.6 and its noise temperature is 2.3, what will be the noise figure of the receiver if the noise figure of the amplifier is 3.12?

CHAPTER 13

13.1. An electron moving with a velocity equivalent to 500 volts crosses a plane boundary into a region where the potential is uniformly 100 volts less. If the electron made an angle of 30 deg with the normal to the plane boundary before crossing it, what angle will it make after? By how much must the potential on the far side of the plane boundary be less than 500 volts for the same angle of incidence in order that the electron will just be reflected back?

13.2. Derive a series expansion for potential similar to that of Eq. (13.21) about a radial line of symmetry for two-dimensional potential fields expressed in terms of polar coordinates r and θ. What angle do the equipotential lines at a saddle point make with the radial line of symmetry in this case?

13.3. What is the radius of curvature of the equipotential line on the axis of an equal-diameter two-cylinder lens at a distance of one radius from the cylinder junction when the cylinder spacing is very small?

13.4. An electric lens consists of a circular aperture in a plate between two parallel plates. Plot the potential along the axis of the lens for the following potentials and dimensions:

$$\begin{aligned} V_1 &= 10 \text{ volts} & d_{12} &= 3 \text{ mm} \\ V_2 &= 2 \text{ volts} & d_{23} &= 9 \text{ mm} \\ V_3 &= 50 \text{ volts} & R &= 1.5 \text{ mm} \end{aligned}$$

where the notation is that of Eq. (13.37).

13.5. What is the focal length of the aperture lens of Prob. 13.4?

13.6. Calculate the two focal lengths and the location of the two principal planes of an equal-diameter two-cylinder lens for a voltage ratio of 4 to 1 by the method of linear axial-potential segments.

13.7. Solve Prob. 13.6 by the method of joined circular segments.

13.8. Solve Prob. 13.6 by the method of equivalent thin lenses.

13.9. For the lens of Prob. 13.6 what is the location of the image for an object located four lens diameters from the cylinder junction on the low-voltage side of the lens, and what is the corresponding magnification?

13.10. Calculate and plot the P-Q curves for the Hutter lens of Fig. 13.26.

13.11. It is desired to use a lens that will operate with an object distance of 3 cm and an image distance of 25 cm. The voltage ratio to be used is to be 5 to 1. If the resulting image is to have as small a size as is practically feasible, what lens should you use?

13.12. Given a lens with the following constants:

$$\begin{aligned} f_1 &= -1.8 \text{ mm} & f_2 &= 4 \text{ mm} \\ F_1 &= -2.6 \text{ mm} & F_2 &= 2.8 \text{ mm} \end{aligned}$$

Calculate and plot image distance and lateral magnification as a function of object distance.

CHAPTER 14

14.1. Derive Eq. (14.26).

14.2. Derive Eq. (14.30).

14.3. Derive Eq. (14.31).

14.4. Derive Eq. (14.57).

14.5. How many ampere turns are required for a magnetic lens that is to have a focal length of 2 cm when the coil diameter is 3 cm and the beam voltage is 800 volts? What is the rotation associated with such a lens?

CHAPTER 15

15.1. A cathode gun of the aperture type shown in Fig. 15.5 has the following dimensions: $R = 1$ mm, $d_{12} = 3$ mm, $d_{23} = 6$ mm. What is the amplification factor determining current cutoff?

15.2. What is the radius of the crossover section of the beam of a gun cathode for which $r_c = 2$ mm, $\theta = 0.1$ radian, $V_2 = 10$ volts, and the cathode is coated with an oxide emitter operating at 1000°K?

15.3. What magnetic-flux density is required to produce a deflection of 2 cm at a fluorescent screen 20 cm from the deflecting field if the region of uniform field is 2.5 cm long and the beam voltage is 1,000 volts?

15.4. An electron beam leaves an electron gun converging with a maximum angle of 5 deg with the axis. The current is uniformly distributed over the beam. If the total beam current is 1 ma and the beam voltage is 800 volts, where will the minimum-diameter cross section occur and what will be the value of this diameter? Assume that the original diameter of the beam is 2 mm. What will be the diameter of the beam on a screen 25 cm from the electron gun?

15.5. At what angle should a 1-kv beam with a current of 1 ma leave an electron gun in order that the cross section of minimum diameter will occur on a screen 25 cm away? Assume that the original diameter of the beam is 2 mm.

15.6. What is the maximum current that can be transmitted in the form of a beam through a cylinder 2.5 cm long and 0.5 cm in diameter without wasting any current in the absence of positive-ion neutralization? What is the impedance corresponding to this current? What will be the maximum current if it is permitted to waste current?

15.7. It is desired to construct a Pierce cathode with a convergent conical beam. The cathode diameter is to be 1 cm; the initial angle of convergence of the beam is to be 15 deg after the anode and 50 deg before the anode. What will be the size of the anode aperture, and what voltage will be required to produce a current of 500 ma? Indicate the shape of the cathode and anode electrodes outside the beam.

15.8. Design a Pierce cathode that will pass the maximum current through a cylinder ½ cm in diameter and 6 cm in length at a voltage of 1,000 volts.

15.9. Design a Pierce cathode that will produce a parallel circular beam ½ in. in diameter and carrying a total current of 1 ampere at a voltage of 5,000 volts.

15.10. Design a Pierce cathode that will produce a strip beam 2 mm thick and carrying a current of 100 ma per cm^2 at a voltage of 600 volts.

15.11. A set of electrostatic-deflection plates for an 800-volt beam is 2 cm long with a spacing of ½ cm. At what frequency will the maximum deflection be reduced to half the d-c value?

15.12. Derive an expression for the static deflection of a two-wire transmission line of wires of diameter *d* and spacing *s*. The electrons are shot between the wires in a direction normal to the plane of the wires. Extend this expression to include transit-time effects.

15.13. Derive an expression for the spread of a sheet beam resulting from the mutual electrostatic repulsion between electrons.

15.14. Derive an expression for the spread of a circular beam of electrons including the effect of the magnetic forces involved.

15.15. At what angle must a beam whose initial diameter is 3 mm leave an electron gun in order to have the minimum possible diameter at a screen 30 cm away if the beam current is 0.1 ma and the beam voltage is 1,000 volts? What is the resultant minimum diameter at the screen?

15.16. Solve problem 15.15 for the case of a beam current of 1 ma, all other conditions being the same.

CHAPTER 16

16.1. Calculate the inductance of a straight piece of wire 2 in. long and 20 mils in diameter. What is the reactance of this inductance at 250 mc?

16.2. An ultra-high-frequency amplifier has a common grid and plate-circuit inductance consisting of a lead 1 in. in length and 50 mils in diameter. What is the component of input conductance due to this at 200 mc if the input capacity is 10 micromicrofarads and the tube conductance is 2,000 micromhos?

16.3. What is the input conductance of a tube due to electron transit-time effects at a frequency of 100 mc if the tube has a mutual conductance of 2,000 micromhos, the cathode-grid spacing is 10 mils, the grid-plate spacing is 50 mils, the effective voltage of the grid plane is 2 volts, and the plate potential is 100 volts?

16.4. A plane-electrode diode has a plate current of 100 ma at a plate potential of 50 volts. (Cathode is one-sided, and there is a plate only on the emitting side.) If the area of the cathode and plate are each 4 cm^2, what is the r-f impedance of the diode at a frequency of 400 mc?

16.5. Calculate the components of the equivalent circuit of Fig. 16.13 for the idealized 210 tube of Prob. 7.1 if the effective cathode and plate area are each 1 in^2. Take the plate potential as 100 volts, the grid potential as half the cutoff value, and the frequency as 200 mc.

16.6. What is the effect of voltage-scaling a triode by a factor of 2 in dimensions upon the various operating constants and properties of the tube?

16.7. What is the effect of completely scaling a tube in the direction of higher frequencies if the tube size is reduced by 2 and the tube is to operate at twice the frequency?

16.8. Two tubes are geometrically similar except that the smaller is half the size of the larger. If the smaller is to be used at three times the frequency to be applied to the larger, how will the tube constants and operating conditions compare?

16.9. At what frequency will the efficiency of a triode whose cathode-grid spacing is 20 mils, whose grid-plate spacing is 100 mils, and whose plate voltage is 1,600 volts have dropped to 90 per cent of its low-frequency value?

16.10. At what frequency will the efficiency of a triode whose cathode-grid spacing is 6 mils and whose grid-plate spacing is 30 mils have dropped to 90 per cent of its low-frequency value if the plate voltage is 250 volts?

16.11. What is the frequency at which the tube of Prob. 16.10 would stop oscillating if the amplification factor of the tube is 20?

16.12. What is the ratio of the grid-plate transit time to the cathode-grid transit time of the tube of Prob. 16.10 if the effective potential of the grid plane is 5 volts and the plate potential is 50 volts?

16.13. Obtain equations of motion for electrons in a plane-electrode diode similar to those of Eq. (16.79) for the case of temperature-limited emission and bias such that current flows for only 60 deg of the entire cycle.

CHAPTER 17

17.1. What is the skin depth of current penetration in copper at 4,000 mc? What is the corresponding surface resistivity?

17.2. What is the skin depth of current penetration in iron at 3,000 mc if the volume resistivity of the iron is six times that of copper and the permeability is 50? What is the corresponding surface resistivity?

17.3. The energy stored in the field of a cavity that is tuned to 4,500 mc drops by a factor of 10 db in 5 microseconds. What is the Q of the resonator?

17.4. What is the error in the approximate expression for the impedance of a parallel resonant circuit given by Eq. (17.13) when the value of δ is $\frac{1}{Q}$?

17.5. What are the equivalent series resistance, inductance, and capacity of a resonator whose shunt resistance is 100,000 ohms, whose Q is 15,000, and whose resonant frequency is 2,500 mc?

17.6. Derive an expression for the beam coupling coefficient of a parallel set of fine grids including second-order transit-time effects.

17.7. Obtain an expression for the beam coupling coefficient of a bunching gap consisting of two equal-diameter cylinders placed end to end without grids.

17.8. Construct a distance-time diagram for electrons bunched by a set of plane grids. The depth of modulation is 0.15, the beam voltage is 1,000 volts, and the frequency is 3,000 mc. Use at least 18 lines per cycle.

17.9. Repeat Prob. 17.8 for a depth of modulation of 0.30.

17.10. The bunching grids of a klystron amplifier are 2 mm apart. If the beam voltage is 2,000 volts and the drift space is 2.5 cm long, what must be the value of the r-f voltage at the bunching grids to produce a maximum fundamental component of current at the catcher? The operating frequency is 3,000 mc. For the same bunching grids, what must be the magnitude of an exciting voltage of 300-mc frequency to produce a maximum value of tenth-harmonic current at the catcher?

17.11. A klystron amplifier has buncher and catcher grids that are fine and plane and spaced 2 mm apart. The length of the drift space is 2.5 cm. The operating frequency is 3,000 mc. The beam current is 20 ma at 1,500 volts. Calculate and plot small-signal transconductance as a function of beam voltage for the following cases:

a. Emission is space-charge-limited so that beam current is proportional to the three-halves power of the beam voltage.
b. Beam current is constant as beam voltage is varied.

17.12. What is the theoretical power required to bunch a beam of 30 ma at 300 volts when the bunching gap spacing is 1.5 mm?

17.13. Construct a distance-time diagram for a reflex-klystron oscillator whose potential field in the reflector space is linear. Let the time spent by an unmodulated electron in the reflector space be 2.75 cycles. Let the depth of modulation be such that maximum power is obtained from the beam. (Construction is simplified if a template of the parabolic curve involved is cut and all curves are drawn from this.) Use at least 18 lines per cycle.

17.14. Plot the negative of the small-signal admittance spiral of a reflex-klystron oscillator whose beam conductance is 100 micromhos and whose beam coupling coefficient is unity. The tube has a resonator whose shunt resistance at the resonant frequency of 3,500 mc is 1,000 ohms and whose Q is 200. Plot the line showing the locus of resonator admittance as frequency is varied. On which transit-time mode will oscillations first occur? On which mode will the power output be maximum? What will be the limiting frequencies of oscillation on the two lowest modes? What will be the frequency stability in megacycles per volt at mid-mode for the lowest mode, assuming that cathode and reflector voltages vary proportionately?

17.15. Derive Eq. (17.89).

17.16. Derive Eq. (17.93).

17.17. Discuss qualitatively the factors determining optimum gap spacing in a reflex-klystron oscillator of the evacuated-cavity type from the standpoint of maximum output power.

17.18. What is the maximum power that can be obtained from a reflex-klystron oscillator having a beam current of 15 ma and operating with a beam voltage of 300 volts, if the gap spacing is 1.5 mm, the resonant frequency is 3,500 mc, the unloaded shunt resistance is 1,000 ohms, and the unloaded Q is 300? Consider that the Q and shunt resistance can be reduced by coupling an external resistive load through a lossless coupling loop. What is the resonator efficiency for a condition of maximum output power?

17.19. Plot curves similar to those of Fig. 17.48 for $k = 0.01$, $Q_1 = 100$, $\omega_1 = \omega_2$, but with Q_2 assuming values of 50 and 200 as well as 100.

17.20. What is the frequency stability in kilocycles per volt of a two-resonator klystron oscillator whose operating frequency is 3,000 mc and whose beam voltage is 1,500 volts if $Q_a = 250$, $Q_b = 60$, and $k = 0.02$, the transit angle between resonators being 8π radians?

17.21. Derive Eq. (17.134).

17.22. What is the starting current of the tube in Prob. 17.20 if the beam coupling coefficient is 0.7 and the mutual reactance between resonators is 50 ohms?

NAME INDEX

SUBJECT INDEX