THEORY OF THERMIONIC VACUUM TUBES

ELECTRICAL ENGINEERING TEXTS

A series of textbooks outlined by a committee of well-known electrical engineers of which Harry E. Clifford, Dean of the Engineering School, Harvard University, is Chairman and Consulting Editor.

Berg—
HEAVISIDE'S OPERATIONAL CALCULUS

Chaffee—
THEORY OF THERMIONIC VACUUM TUBES

Dawes—
COURSE IN ELECTRICAL ENGINEERING
Vol. I.—Direct Currents. Second Edition.
Vol. II.—Alternating Currents. Second Edition.

Dawes—
INDUSTRIAL ELECTRICITY—PART I
INDUSTRIAL ELECTRICITY—PART II

Langsdorf—
PRINCIPLES OF DIRECT-CURRENT MACHINES Fourth Edition

Lawrence—
PRINCIPLES OF ALTERNATING CURRENTS

Lawrence—
PRINCIPLES OF ALTERNATING-CURRENT MACHINERY

Laws—
ELECTRICAL MEASUREMENTS

ELECTRICAL ENGINEERING TEXTS

THEORY OF THERMIONIC VACUUM TUBES

FUNDAMENTALS—AMPLIFIERS—DETECTORS

BY

E. LEON CHAFFEE, PH.D.

Professor of Physics, Harvard University

FIRST EDITION
THIRD IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK AND LONDON
1933

THE MAPLE PRESS COMPANY, YORK, PA.

TO
ALICE

PREFACE

This book is based on the author's lecture notes for a course on vacuum tubes given at Harvard University since 1922. As the preparation of the manuscript progressed, it became apparent that not all of the material could be contained in a single book. Consequently, only the theory of the operation of vacuum tubes at low power is presented here; the remaining material, including the theory of power amplifiers and oscillators, gas-content tubes, rectifiers, etc., will, according to present plans, appear in a second book.

Although this book is written primarily as a textbook, it is hoped that it will serve also as a reference book. The author has endeavored to present only the fundamental principles of the subject, avoiding discussion of the multifarious circuits in which the vacuum tube may be used. The circuits and applications of vacuum tubes change from year to year but the fundamental theory is the same for all time. With an understanding of the principles, any circuit and any application can be analyzed.

Certain sections, which go into considerable detail, may to advantage be omitted on the first reading of the book. For the guidance of the reader these sections are indicated by an asterisk(*).

The author takes this opportunity to express his gratitude to his wife, always a companion, coworker, and inspiration in the preparation of the manuscript; to David P. Wheatland who so generously assisted in collecting experimental data and in reading the proofs, and to all others who have aided in various ways. The author wishes especially to acknowledge his obligation to Prof. H. E. Clifford, Dean of the Harvard Engineering School, for his many valuable suggestions and corrections in editing the manuscript.

E. L. C.

CRUFT LABORATORY, CAMBRIDGE.
March, 1933.

CONTENTS

Chapter IV

EMISSION OF ELECTRONS

Chapter V

PRACTICAL SOURCES OF EMISSION AND SOME GENERAL PHYSICAL ASPECTS OF VACUUM TUBES

CHAPTER IX

DYNAMIC MEASUREMENT OF TRIODE COEFFICIENTS

CHAPTER X

EFFECTS OF GAS IN A TRIODE

CHAPTER XI

INPUT AND OUTPUT ADMITTANCE OF A TRIODE

CHAPTER XII

LOW-POWER AMPLIFIER

CHAPTER XIII

REGENERATION WITH TRIODES

CHAPTER XIV

REGENERATION IN COUPLED CIRCUITS WITH SMALL SIGNALS

CHAPTER XV

COMBINATIONS OF TRIODES AS LOW-POWER RESISTANCE-COUPLED AMPLIFIERS

Chapter XVIII

METHODS OF REDUCING ENERGY INTERCHANGE BETWEEN GRID AND PLATE CIRCUITS OF A TRIODE. "NEUTRALIZATION"

Chapter XIX

SMALL-SIGNAL DETECTION BY A DIODE

Chapter XX

SMALL-SIGNAL DETECTION BY TRIODE

Chapter XXI

THEORY OF THE OPERATION OF NONLINEAR CIRCUITS WITH LARGE ELECTRICAL VARIATIONS WITH SPECIAL REFERENCE TO DETECTION WITH LARGE SIGNALS

Chapter XXII

EXPERIMENTAL TREATMENT OF DETECTION AT HIGH SIGNAL VOLTAGES

CHAPTER XXIII

TETRODES AND PENTODES

APPENDIX A

THEORY OF SUPERIMPOSED CURRENTS

APPENDIX B

METHOD OF TESTING AN AUDIO TRANSFORMER

LIST OF SYMBOLS

A	Denotes the average point
A	$r_p C_{pg} \omega$
B	Susceptance or imaginary components of admittance not a function of current or potential
b	Susceptance or imaginary component of admittance which is a function of current or potential
C	Capacitance
D	Durchgriff
E	R.m.s. voltage
$\bar{E}$	Steady voltage
$\overline{E}$	Average value of voltage
$\widehat{E}$	Maximum value of sinusoidal voltage
$\mathbf{E}$	Vector or complex r.m.s. value of voltage
$\widehat{E}$	Peak value of voltage
$\breve{E}$ $[E]$	Fictitious or equivalent value of voltage
$(\widehat{E}_p)_l$	Maximum value of plate-to-cathode voltage of low frequency
$(E_p)_l$	Plate-to-cathode voltage of low frequency. R.m.s. value
$(\Delta E_p)_l$	R.m.s. value of small alternating component of plate voltage of low frequency
$\Delta^2\overline{E}_p$	Steady value of increment of plate voltage of second order in magnitude
$(\widehat{\Delta^2 E_p})_l$	Maximum value of alternating component of plate voltage of low frequency and of second order in magnitude
$(\Delta^2 E_p)_l$	R.m.s. value of alternating plate voltage of low frequency and of second order in magnitude
$(\boldsymbol{\Delta}^2\mathbf{E}_p)_l$	Complex r.m.s. value of plate voltage of low frequency and of second order in magnitude
$[(\Delta^2 E_p)_l]$	R.m.s. value of fictitious alternating plate voltage of low frequency and of second order in magnitude

Similar quantities with subscript h denote alternating quantities of high frequency.

e	Electronic charge
e	Instantaneous value of voltage
$\tilde{e}$	Instantaneous value of sinusoidal voltage
$\bar{e}$	Instantaneous value of voltage measured from quiescent value
$\bar{\bar{e}}$	Instantaneous value of voltage measured from average value
ϵ	Base of Naperian logarithms
F	Denotes a function or a factor
G	Real component of admittance not a function of current or potential
g	Real component of admittance which is a function of current or potential
h	Planck's constant
h	Coefficient of regeneration
I	R.m.s. value of current
$\bar{I}$	Steady value of current
$\bar{\bar{I}}$	Average value of current
$\hat{I}$	Maximum value of current
$\mathbf{I}$	Vector or complex r.m.s. value of current
$\hat{I}$	Peak value of current
$(\hat{I}_p)_l$	Maximum value of plate-to-cathode current of low frequency
$(I_p)_l$	Plate-to-cathode current of low frequency. R.m.s. value
$(\Delta I_p)_l$	R.m.s. value of small alternating component of plate current of low frequency
$\Delta^2\bar{I}_p$	Steady value of increment of plate current of second order in magnitude
$(\widehat{\Delta^2 I_p})_l$	Maximum value of alternating component of plate current of low frequency and of second order in magnitude
$(\Delta^2 I_p)_l$	R.m.s. value of alternating plate current of low frequency and of second order in magnitude
$(\boldsymbol{\Delta}^2\mathbf{I}_p)_l$	Complex r.m.s. value of plate current of low frequency and of second order in magnitude

Similar quantities with subscript h denote alternating quantities of high frequency.

i	Instantaneous value of current
$\tilde{\imath}$	Instantaneous value of sinusoidal current
$\bar{\imath}$	Instantaneous value of current measured from quiescent value

$\bar{\bar{i}}$	Instantaneous value of current measured from average value
j	Operator which rotates vector 90 deg. (numerically = $\sqrt{-1}$)
K	Conductance, not a function of potential or current
$\bar{K}$	Conductance offered to steady current, not a function of potential or current
$\tilde{K}$	Conductance offered to alternating current, not a function of potential or current
k	Variational conductance which is a function of potential or current
$\bar{k}$	Conductance evaluated from steady current and potential, but which is a function of current or potential
$\bar{\bar{k}}$	Conductance evaluated from average values, but which is a function of current or potential
k_p	Plate variational conductance of vacuum tube
k_g	Grid variational conductance of vacuum tube
k_{pp}	Plate variational conductance of multielectrode tube
k_{gg}	Grid variational conductance of multielectrode tube
L	Self-inductance
M	Mutual inductance
$\mathfrak{M}$	Critical value of mutual inductance for regeneration
m	Degree of modulation
n	A whole number or a frequency
p	A whole number
Q	Denotes the quiescent point
q	Electric charge
R	Resistance, not a function of potential or current
$\bar{R}$	Resistance offered to steady current, not a function of potential or current
$\tilde{R}$	Resistance offered to alternating current, not a function of potential or current
r	Variational resistance which is a function of current or potential
$\bar{r}$	Resistance evaluated from steady potential and current, but which is a function of potential or current
$\bar{\bar{r}}$	Resistance Evaluated from average values, but which is a function of potential or current
r_p	Variational plate resistance of vacuum rube
r_g	Variational grid resistance of vacuum tube
r_{pp}	Variational plate resistance of multielectrode tube

r_{gg}	Variational grid resistance of multielectrode tube
s	Variational transconductance of a vacuum tube or amplifier
$\bar{s}$	Transconductance evaluated from average values, but which is a function of current or potential
s_p	Grid-to-plate variational transconductance of vacuum tube
s_g	Plate-to-grid variational transconductance of vacuum tube
s_{pg_n}	Plate variational transconductance from nth grid to plate of multielectrode tube
s_{g_np}	Variational reflex transconductance from the plate to the nth grid of multielectrode tube
T	Temperature or periodic time
t	Time
u_p	Variational voltage ratio or amplification factor of a vacuum tube
u_{pg_n}	Variational voltage ratio with respect to nth grid of multielectrode tube
u_{g_np}	Variational reflex factor with respect to nth grid of multielectrode tube
w	Thermionic work function
X	Reactance
$(X)_l$	Reactance to low frequency
$(X)_h$	Reactance to high frequency
x	Reactance which is a function of potential or current
x	Coordinate
Y	Admittance not a function of potential or current
y	Admittance which is a function of potential or current
y	Coordinate
Z	Impedance not a function of potential or current
z	Impedance which is a function of potential or current

Greek Letter Symbols

α	An angle
β	A reactance $\left(L\omega - \frac{1}{C\omega}\right)$ divided by $L\omega$
β	Special use in Chap. XXII, definition: Equation (803), page 576
γ	Shielding factor in a triode or multielectrode tube

Δ	Before a symbol indicates small but finite value
ϵ	A small quantity
η	Ratio of resistance to inductive reactance
θ	An angle
κ	Boltzmann's constant
λ	Wave length
ν	Frequency of radiation
π	3.1416
ρ	Volume density of charge
Σ	Sum of . . .
τ	Coefficient of coupling
ϕ	An angle
ψ	An angle or indicates function of
ω	Angular velocity

General Notes

Subscripts A, B, C denote sources of power in filament, plate, and grid circuits of a triode.

Subscripts a, b, c denote quantities associated with the external filament, plate, and grid circuits of a triode.

Subscripts f, p, g denote quantities directly associated with the filament, plate, and grid of a triode.

Subscripts h and l outside parentheses denote that the quantity is of or for high and low frequencies.

THEORY OF THERMIONIC VACUUM TUBES

CHAPTER I

INTRODUCTION

1. Classification of Vacuum Tubes.—The expression *vacuum tube* signifies a variety of devices all of which have in common a closed envelope or bulb made usually of glass or quartz, into which are sealed one or more electrodes and from which the air has been mostly exhausted or replaced by some other gas at reduced pressure. These tubes are capable of passing electric currents between their electrodes and, because of the passage of the electric currents, various effects are obtained which make the devices useful. As specific examples of such vacuum devices we may enumerate the following: Geissler tube, mercury-arc lamp, mercury rectifier, X-ray tube, tungar rectifier, photoelectric tube, and radio vacuum tube.

All vacuum tubes may be divided into two classes. Examples of the *first* class are the Geissler tube, the mercury-arc lamp, the X-ray tube, etc., which are useful because of radiation into which some of the electrical energy supplied to them is transformed. Since these tubes are sources of useful radiation, they are output devices in the same sense as are generators, motors, incandescent lamps, electric bells, etc. This book is not concerned with this class of vacuum-tube radiators but with the second class now to be described.

The *second* class includes those vacuum tubes which are useful because of the action they have in, or the effect they have upon, the circuits in which they are connected. The useful effects are usually localized in some part of the circuit remote from the vacuum tube itself. This class of devices may be called *circuit elements*, for, like resistances, inductances, capacitances, switches, relays, etc., they are devices used as component parts of electric

circuits which work together to cause some useful effect in the circuits. These tubes are not themselves output devices, as were the tubes of the first class, but are *controlling means.*

Among these controlling or circuit-element tubes may be listed the following:

1. Mercury rectifier.
2. Tungar rectifier.
3. Short-path gaseous rectifier.
4. Thermionic vacuum tube.
   - *a.* Two-electrode tube used as Fleming valve detector.
     Rectifier (Kenotron).
   - *b.* Three-electrode tube used as:
     Amplifier or repeater.
     Oscillator.
     Modulator.
     Detector.
   - *c.* Four-electrode tube used as:
     Amplifier.
     Detector.
     Oscillator.
   - *d.* Five-electrode tube used as:
     Amplifier.
     Detector.
     Oscillator.
   - *e.* Magnetron tube used as:
     Amplifier.
     Oscillator.

The above list is not complete but is given merely to illustrate the suggested classification. Although the first three kinds of controlling tubes are briefly treated, more emphasis is given to the most important kind of controlling tube, the *thermionic vacuum tube.*

2. General Considerations.—The passage of electricity through the space between the electrodes of all types of vacuum tubes is effected by a migration of discrete charged particles or carriers. If the tube is well evacuated, the current through the tube consists of a drift of minute negative charges, all of the same unitary charge of electricity, called *electrons.* If, on the other hand, an appreciable amount of gas is present in the tube, a part of the conduction may be due to charged particles of gas called *ions.* These ions may be either negatively or positively charged. Even

if the conduction of electricity is partly by gaseous ions, electrons are always present and indeed absolutely necessary in order to start the conduction. Electrons and gaseous ions play such important rôles in the operation of electronic vacuum devices that the second and third chapters are devoted to the fundamental principles concerning these carriers of electricity.

Since electrons are essential for the conduction of electricity through a vacuum tube, there must be some *source* of electrons inside the tube. Some possible sources are radioactive substances which emit electrons spontaneously, matter when acted upon by ultra-violet light or X-rays, and matter when strongly heated. The last source provides the most copious supply of electrons and is also the most convenient one. A heated electrode or filament, often called the *cathode*, is the most frequently used source in the electronic vacuum tubes to be discussed in this book. O. W. Richardson has given the name *thermions* to the electrons emitted from hot bodies and, although these electrons are not at all different from the electrons derived from other sources and are strictly not ions at all, the term has come into such general use that the inconsistency of the expression would best be disregarded. Vacuum tubes using thermions are called *thermionic vacuum tubes*. The subject of thermionic emission of electrons is discussed in Chap. IV and a discussion of the practical emitters is given in Chap. V.

The thermionic vacuum tubes have, besides the hot electrode, one or more cold electrodes. In the early forms of tubes, one of these cold electrodes was a flat metal plate, or two plates connected together, one on either side of the filament. Hence this cold electrode is called the *plate*. Some writers prefer the term *anode* for this electrode, since the conventional idea of the flow of electricity makes this plate electrode one of the electrodes by which the current enters the tube. Although the term *anode* may be scientifically more elegant, the simpler term *plate* is generally used, even though in many tubes now constructed the original flat plate has evolved into other shapes, such as cylinders or flattened cylinders. There is in most tubes no difficulty in identifying the plate, since it is usually the most substantial electrode of the tube and is usually made of sheet metal. The other cold electrode or electrodes are usually of open mesh construction. This third electrode (there may be two or more electrodes of this type instead of one) is usually situated between

the hot electrode and the plate and in the early tubes was made up of a grid work of wires and hence called a *grid.* Although the grid electrode takes various forms in modern tubes, the term grid is sufficiently descriptive of this electrode to leave no question in one's mind as to which electrode is the grid. Since usually the function of the grid electrode is to effect control of the current to the plate, the grid electrode is sometimes called the *control electrode.*

Many names have been applied to the two-electrode tube, such as Fleming valve, diode, and kenotron, and to the three-electrode tube, such as audion, pliotron, oscillion, and radiotron. It is to be deplored that so many expressions have arisen for the same device. Most of the terms have been invented by manufacturers to distinguish their products. Although this may be admirable advertising tactics, it is believed that such trade names should not be used in scientific literature unless for the express purpose of specifying a certain manufacturer's product. It is, however, desirable to differentiate the two- and three-electrode tubes because these two types of tubes have different characteristics. The terms *two-electrode tube* and *three-electrode tube* are definite and satisfactory except for their length, so the simpler terms *diode* and *triode* will be used for the two types of thermionic tubes. If there are four or five electrodes, the device may be termed a *tetrode* or a *pentode.*

Returning to the consideration of vacuum tubes in general, electrons, and ions if gas be present, pass between the electrodes under the directing force of the potential differences impressed between the electrodes. Except in one or two special types of tubes where magnetic fields are used to influence the passage of the carriers, vacuum tubes are electric devices obedient to the laws of electrostatics.

The action of a tube as a circuit element is described in terms of the functional relation between the impressed voltage and the current, together with the influence of certain other variables such as temperature and time. With some forms of circuit element, this functional relation is expressible in a simple mathematical equation. For example, for a resistance, $e = Ri$; for an inductance, $e = +L\frac{di}{dt}$; for an arc, $e = a + \frac{b}{i} + \frac{c}{i^2}$; where e is the instantaneous potential difference impressed on the device, i is the instantaneous current through the device, t is time, and the other letters are constants. The functional relation for some

other devices, such as most of the thermionic tubes, cannot be expressed in such simple mathematical form, so that recourse must be had to graphical methods. The current through a diode is a function of the impressed voltage and the temperature of the electron emitter, unless the variations of potential are extremely rapid or gas is present, in which cases time may be an important factor. In any case, however, these functional relations are best expressed by families of curves. Such graphs are discussed in Chap. IV. A triode has two current paths which are mutually dependent and the functional relations between the variables are more complex. These relations also have to be expressed graphically and are discussed in Chap. VII.

Since the historical order is not followed in this book, a very brief historical sketch of the development of the thermionic tube is given here. This sketch will also serve to lay before the reader some of the important uses to which these thermionic tubes are put.

3. Brief Historical Sketch.—One of the earliest experiments which contributed to knowledge at the foundation of vacuum-tube development is that performed in 1873 by F. Guthrie.[1] He observed that a negatively charged electroscope was discharged when a metal ball heated to a dull red was brought near it, but that a positively charged electroscope was not discharged unless the ball was heated to a higher temperature. An electroscope charged either positively or negatively became discharged with the ball at a white heat. This experiment proves that the metal ball when very hot furnishes both positively and negatively charged carriers of electricity but furnishes only positively charged carriers when at the lower temperature. In the light of present-day knowledge, we know that a metal when heated to a dull red sometimes gives off positive ions of gas, originating probably from the gas occluded in the metal. At a higher temperature the metal gives off electrons. Guthrie's experiment can be conveniently repeated by placing at a short distance above the terminal of a charged electroscope, a wire which can be electrically heated to various temperatures.

Between the years 1882 and 1889, Elster and Geitel,[2] two German scientists, made a more extensive study of the conductivity of the gas near heated solids and flames. They investigated

[1] GUTHRIE, *Phil. Mag.*, 4th ser., **46**, 257 (1873).

[2] ELSTER and GEITEL, *Ann. d. Physik*, **16**, 193 (1882); **19**, 588 (1883); **26**, 1 (1885); **31**, 109 (1887); **37**, 319 (1889).

the conduction effects in different gases under reduced pressure. In some of their later work, they used a piece of apparatus which was a two-electrode vacuum tube, consisting of a straight electrically heated filament of carbon or metal located inside a bulb and directly below a cold plate sealed into the top of the bulb. Elster and Geitel were interested only in the scientific side of their experiments and, although they noted the unilateral conductivity of their two-electrode device, they did not mention its possibilities as a rectifier of alternating currents.

Thomas A. Edison,[3] in 1883, with apparently no knowledge of the experiments of Guthrie or the early experiments of Elster and Geitel, made an important discovery. Some of the carbon filaments of incandescent lamps, which Edison had recently invented, developed hot-spots due to a diminished cross section at certain points. These hot-spots not only shortened the life of the lamp but caused the bulbs to blacken, owing to carbon particles projected in straight lines from these hot-spots and deposited on the inside of the glass bulb. In the investigation of this phenomenon, he sealed into the bulb a plate situated between the legs of the horseshoe-shaped filament. He then discovered that a galvanometer, connected between the plate and the positive end of the filament, indicated a current, but that no current flowed when connection was made to the negative end of the filament instead of to the positive end. In the first case, the current evidently passed through the vacuous space between the plate and the filament. This phenomenon has since been known as the *Edison effect* and is probably the real starting point of the modern thermionic tube.

In 1884 and 1885 Preece[4] in England made some quantitative measurements of the Edison effect. Preece showed that the kind of metal used for the plate had no influence upon the effect, but that the current was greatly influenced by the distance between filament and plate, the temperature of the filament, and the potential difference applied between the plate and filament. He explained the phenomenon by assuming that the current was carried by negatively charged molecules or particles of carbon shot off in straight lines from the filament.

Attention is also directed to the early work of W. Hittorf[5] who showed that a current could be passed through a vacuum

[3] *Engineering*, p. 553, Dec. 12, 1884.

[4] PREECE, *Proc. Roy. Soc.* (*London*), **38**, 219 (1885).

[5] HITTORF, *Ann. d. Physik*, **21**, 119 (1884).

tube with a very small potential difference if the cathode was incandescent; and to the work of E. Goldstein,[6] who showed that, using a hot filamentary cathode of carbon or platinum, a current could be made to pass through a discharge tube even with low potentials, when the gas pressure was so low that, with the cathode cold, the highest voltages caused no perceptible discharge through the tube.

J. A. Fleming[7] of the University of London, who was in touch with the work of Edison and of Preece, conducted an extended series of experiments on the Edison effect between the years 1889 and 1896. Although the unilateral conductivity of the glow lamp provided with a cold plate was known, and hence its ability to rectify low-frequency alternating currents was a natural consequence, yet Fleming was the first to suggest and use this device for detecting high-frequency oscillations by converting them into direct current through the rectifying action of the vacuum tube. He took out a patent in Great Britain[8] in 1904, and in the United States[9] and Germany[10] in 1905, covering the use of the "oscillation valve," as he termed the device, as a detector or receiver in wireless telegraphy. The vacuum tube consisting of an electrically heated filament and a cold plate sealed into an exhausted bulb is a two-electrode tube and is often justly called a "Fleming valve."

The remarkable work of J. J. Thomson[11] in 1897 on the measurement of the mass of the electron, and the proof that all electrons derived from any source whatever are of the same mass and charge, furnished the true explanation of the Edison effect. The reader is referred to J. J. Thomson's books "Conduction of Electricity through Gases" and "The Corpuscular Theory of Matter" for an account of these experiments. In the light of Thomson's work, it is clear that the current of electricity is not carried through the vacuous space between the filament and plate by charged particles of matter but by minute unitary charges of negative electricity called *electrons*, all of the same charge, which

[6] GOLDSTEIN, *Ann. d. Physik*, **24**, 79 (1885).

[7] FLEMING, *Proc. Roy. Soc.* (*London*), **47**, 118 (1890); **14**, 187 (1896); *Phil. Mag.*, **42**, 52 (1896).

[8] British patent 24,850, Nov. 16, 1904.

[9] U. S. patent 803,684, Apr. 19, 1905.

[10] German patent 186,084, Klasse 21a, Gruppe 68, Apr. 12, 1905.

[11] THOMSON, J. J., *Phil. Mag.*, **44**, 293 (1897).

are emitted by the hot filament and attracted by the positively charged plate.

Many of the early experimenters in the field of thermionics believed that the emission of electricity from hot bodies, even if by electrons, was the result of some sort of chemical reaction between the small residue of gas remaining in the tube and the material of the filament. In fact, the observed diminution of the current from plate to filament, as more and more of the gas is removed, supported this view. On the other hand, O. W. Richardson, in 1901, working along both theoretical and experimental lines, advanced the theory that the electron emission at elevated temperatures is a property alike for all conductors and not at all dependent upon the presence of gas about them. Richardson derived mathematical laws for the electron emission as dependent upon temperature and these laws agree well with experiment. The reader is here referred to Richardson's book "The Emission of Electricity from Hot Bodies" for a more complete account of Richardson's valuable theoretical contributions to this field.

In 1903 and 1904 A. Wehnelt[12] discovered and investigated the copious emission of electricity from metal filaments coated with certain oxides, notably of the rare earths, such as the oxides of strontium, barium, calcium, etc. A heated filament or electrode of other form coated with one or more of these oxides and used as the negative electrode of a vacuum tube is known as a *Wehnelt cathode* and is one very common form of emitting electrode used in modern tubes.

The thermionic emission from pure tungsten was very thoroughly investigated by Irving Langmuir. His first paper was presented before the Institute of Radio Engineers, April 7, 1915, and in it he gives further support to Richardson's theory that pure electron emission is a property of metals independent of the presence of gas. Subsequent work by Langmuir and Dushman has added greatly to the practical knowledge of the properties of tungsten filaments.

We shall now consider the application made by Fleming of the two-electrode vacuum tube to the detection of high-frequency oscillations. In his earlier experiments and in his first patent applications of 1904 and 1905, he made use of the simplest property of the vacuum tube; namely, when a potential is

[12] German patent 157,845, Klasse 21g, Jan. 15, 1904; *Ann. d. Physik*, **14**, 425 (1904).

impressed between the plate and the negative end of the filament, a current flows when the plate is made positive, but very little or no current flows when the plate is made negative. Figure 1 gives one scheme of connections used by Fleming. *A* represents the antenna connected to ground through the tuning inductance *P*. The tuned circuit *SC* is coupled to the antenna coil *P*. The "oscillation valve" is diagrammatically represented by *V*. When oscillations occur in circuit *SC*, induced from the antenna circuit, oscillations of potential which take place across condenser *C* are impressed between the filament and plate of tube *V*. The rectified impulses then cause a deflection of the galvanometer *G* or produce a sound in a telephone receiver which may be substituted for *G*.

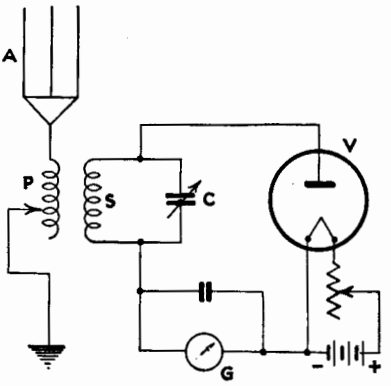


FIG. 1.—Fleming value used as a detector of radio signals.

In 1908 and 1909 Fleming filed patent applications in Great Britain and in the United States[13] covering a more efficient method of using a diode as a receiver of radio messages. This second method, which will be explained more fully in Chap. XIX, makes use of the rectifying action at a sharp bend of the characteristic curve of the diode. In this method the potential oscillations of the message are superimposed upon a steady potential and their sum is impressed between the filament and plate of a diode. The resulting net transfer of electricity in one direction in the plate circuit, due to the message, is considerably greater than in the use of the device as a simple valve, and louder signals are obtained.

The diode used to rectify the exceedingly small power of a radio signal is made in small sizes. The development of the diode for the rectification of large amounts of power at high potentials was a logical sequence but required the overcoming of certain technical difficulties. The presence of gas in the diode used for receiving messages is not always detrimental, for the sensitiveness of the detector is often materially increased by a small amount of gas. When high potentials and larger currents

[13] British patent 13,518, June 25, 1908; U. S. patent 945,619, Jan. 2, 1909.

are used, the presence of even a very small amount of gas is disastrous. It was necessary, therefore, in the construction of high-power vacuum tubes, to exhaust to a point that was possible only after the timely invention and development of better high-vacuum pumps. In 1913 Gaede of Germany introduced the molecular pump and in 1915 the diffusion pump. Langmuir shortly afterwards described an improved mercury pump which he called a "condensation pump." Using one of these improved methods of exhaustion, Langmuir, with other engineers, advanced the knowledge of pure electron emission and designed high-power rectifying diodes which were called *kenotrons*.

Another line of development of the diode rectifier resulted in the *tungar rectifier*, which is a two-electrode tube containing pure argon gas at several centimeters pressure. These tubes are suitable for rectifying relatively large currents at low potentials and are in common use for charging storage batteries from an a-c. source.

In 1907 Lee de Forest[14] made an original and radical improvement in the vacuum tube by interposing, between the filament and plate, a third electrode in the form of a wire grating or grid. This third electrode is situated in the most favorable position to control electrostatically the flow of electrons from the filament through the meshes of the grid to the plate. From this controlling action the three-electrode tube or triode derives all of its advantages over the diode. The triode was first used by de Forest, as shown in Fig. 2, and proved to be a great improvement on the detectors of the prior art. The potential fluctuations are impressed between the grid and filament, and the galvanometer or telephone is connected in series with a battery between the plate and filament.

With the introduction of the third electrode, the vacuum tube acquired new properties which were destined to open up tremendous fields of commercial utility. Under proper conditions, a variation of potential of the grid with respect to the filament causes corresponding variations of current in the plate circuit

[14] U. S. patent 879,532, filed Jan. 29, 1907; British patent 1,427, filed Jan. 21, 1908. For a complete description of the early patent history concerned with the diode and triode, see book by Fleming, "The Thermionic Valve and Its Developments in Radiotelegraphy and Radiotelephony," The Wireless Press, Ltd., London, 1919.

and this is accomplished with little or no flow of current to the grid. The device then acts as an *amplifier* or relay, because power can be controlled in the plate circuit by the expenditure of little or no power in the grid circuit. The triode was first used to amplify alternating currents of audible frequencies, but as development progressed, it was successfully employed in amplifying currents of radio frequencies.

One of the most important applications of the amplifying triode is the telephone repeater. The American Telephone and Telegraph Company and the Western Electric Company have spared no expense in research and development of triodes used in telephone practice. The major part of this development has

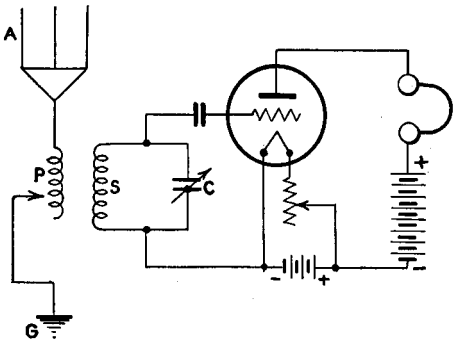


FIG. 2.—DeForest "Audion" or triode used as a detector.

taken place since 1912. As an indication of the rapidity of development, direct-wire telephony from New York to San Francisco was made possible as early as 1914 by the installation of triode repeaters along the line. The triode has been the direct cause of great advances in communication, some of which are long-distance telephony, multiplex telephony and telegraphy, and transoceanic telephony. Contemporary development has been carried on notably in England, France, Germany, and Italy.

Since a triode used as an amplifier of alternating currents gives out more power in the plate circuit than is fed into the tube at the grid, a small amount of the plate-circuit power can be spared to be fed back into the grid circuit, thereby very much increasing the plate power. This feeding back of power in such a way as to cause a building up of the effect in the plate circuit is known as *regeneration.* The discovery of regeneration

has generally been credited to E. H. Armstrong.[15] A decision of the Circuit Court of Appeals has credited de Forest with the origination of the idea of regeneration, at least at low frequencies, although Armstrong was probably the first to succeed in obtaining the regenerative effect at radio frequencies.

If the feed-back of power from the plate circuit to the grid circuit of a triode is increased, a point is reached when self-sustained oscillations are produced, the energy coming from the plate battery and the frequency being determined by the capacitance and inductance of the attached circuits. The vacuum tube thus acting as a generator of alternating currents of any frequency came into still another field of usefulness. Alexander Meissner[16] was probably the first to recognize this new use for a three-electrode tube and filed a patent application April 10, 1913. In a science commanding the attention of so many experimenters and developing so rapidly, it is sometimes difficult to determine the originator of any particular idea. The invention of the oscillating property of a tube is also attributed to Armstrong and to de Forest in America, and to C. S. Franklin and H. J. Round in England.

The discovery of the oscillating properties of the vacuum tube found application at the transmitting station for exciting oscillations in the antenna. This new application led naturally to the development of larger triodes which could convert more power. After overcoming many obstacles, tubes were made capable of generating several hundred watts, and, at present, tubes are made with an output of several hundred kilowatts. The limiting effect in increasing the output power is the amount of heat which it is possible to dissipate at the plate; and if the plate is sealed *inside* the vacuum tube, the only way of dissipating its heat is by radiation.

About the year 1911 wireless telephony began to be investigated by a great many experimenters. The continuous oscillations necessary for successful wireless telephony were at first generated by arcs or quenched sparks. With the advent of the vacuum-tube oscillator, the development of wireless telephony made rapid strides. In the early experiments on wireless telephony, weak currents were used and the ordinary microphone transmitter, placed in the oscillatory circuit, served as a means

[15] ARMSTRONG, *Proc. I.R.E.*, **3**, 215 (1915).

[16] MEISSNER, German patent, 291,604, Apr. 10, 1913.

for directly varying the amplitude of the oscillations in synchronism with the sound vibrations of the voice, a process known as *modulation*. The microphone was then used as the *modulating* device. As vacuum tubes were made for greater power, other means of modulating the powerful oscillating currents became necessary. Experiments showed that no other device could perform this modulating function as well as the same type of vacuum tube that was used for generating the oscillations.

In the year 1915 the engineers of the American Telephone and Telegraph Company and of the Western Electric Company succeeded in transmitting speech from the Naval Station at Arlington, Va., to Paris, and also from Arlington to Honolulu, the greatest distance being about 5,000 miles. The transmitting station utilized about 300 vacuum tubes, each of 25 watts capacity, some used as oscillators, others as modulators, and still others as power amplifiers.

The importance that radio communication assumed during the World War gave added stimulus to the development of the vacuum tube and its applications. During the period of hostility, there was little scientific *liaison* among the various countries, so that much duplication of work occurred. The vacuum tube was made in enormous quantities for use by the armies, resulting in more or less standardized types of tubes.

After the war came the period of broadcasting, which really began in America about 1921. The rapid growth of the radio industry since 1921 has been unparalleled. This rapid development from the beginning of the war up to the present time has been the result of the efforts of countless investigators. In this brief outline we may enumerate only a few outstanding developments, leaving the explanation of their operation to later chapters.

Many improvements have been made in the receiving apparatus, thereby extending the range of radio signaling and the extent of application of the triode. In December, 1919, Armstrong[17] described a new system for receiving radio signals, which he called the *superheterodyne*. The superheterodyne principle has proved to be of considerable value and is used in many modern radio-receiving sets.

In 1922 Armstrong[18] announced another system of reception, which he called *superregeneration*. This system has failed to

[17] ARMSTRONG, *Proc. I.R.E.*, **9**, 3 (1921).

[18] ARMSTRONG, *Proc. I.R.E.*, **10**, 244 (1922).

come into extensive use except at very high frequencies because of the difficulty of adjustment and the fact that weak signals are not amplified so much as strong signals.

C. W. Rice,[19] in 1920, and L. A. Hazeltine,[20] in 1924, described methods of preventing the oscillation of radio amplifiers by using neutralizing circuits which balanced the regeneration through the internal capacitance of the tube. This development enormously increased the stability and sensitivity of radio-frequency amplifiers.

W. Schottky,[21] in 1919, first suggested the use of a second grid to serve as an electrostatic screen between the control grid and the plate. The *screen-grid* tube was later developed by Hull and Williams[22] and made stable radio-frequency amplifiers possible without the necessity of neutralizing circuits. These four-electrode, or screen-grid, tubes came into common use in this country about 1928 or 1929 and are now extensively used in modern receiving sets.

With the advent of broadcasting, receiving sets were installed in a great many homes. To operate the filaments of the tubes then manufactured, a storage battery was required. The cost of the battery and the difficulty of recharging without additional expensive apparatus gave rise to a demand for tubes which could be operated from dry batteries. During the latter part of the war, the Western Electric Company developed for the Navy a small receiving tube which required only 0.25 amp. to heat the filament, instead of about one ampere required by most other American tubes of that period. Later, tubes of the same general characteristics appeared on the market, to be followed by new designs requiring only 0.06 amp. for the filament.

The tubes having filaments of low-power consumption which can be operated from dry batteries mark a decided advance by making it possible to operate a small receiving set entirely on dry batteries, thus eliminating the storage battery. Even dry batteries for operation of the filament and plate supply are objectionable, in that their life is short when several tubes are operated, and this makes the expense of renewal of batteries an important consideration. The next important development was in the design of

[19] RICE, U. S. patent 1,334,118, Mar. 16, 1920.
[20] HAZELTINE, U. S. patent 1,489,228, April, 1924; 1,533,858, April, 1925.
[21] SCHOTTKY, *Arch. Elektrot*, **8**, 299 (1919).
[22] HULL and WILLIAMS, *Phys. Rev.*, **27**, 433 (1926).

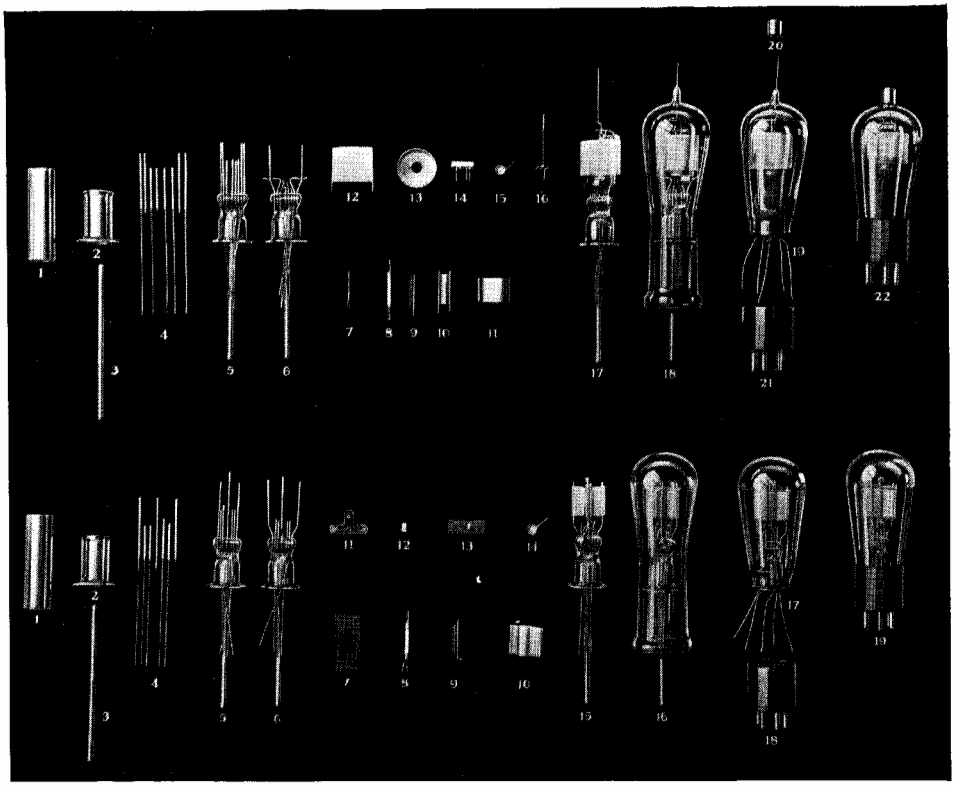


PLATE I.—Construction of modern vacuum tubes. (*Courtesy of R.C.A. Radiotron Company.*) Above: parts of screen-grid tetrode UY-224; below: parts of triode UY-227.

ABOVE

1. Tube for stem
2. Flare for stem
3. Pump tube
4. Leading-in wires
5. Stem
6. Finished stem
7. Heater filament
8. Complete cathode
9. Control grid
10. Inner screen
11. Plate
12. Outer screen
13. Top shield
14. Insulating support
15. Getter cup
16. Control-grid connection
17. Complete mount
18. Bulb
19. Sealed and exhausted tube
20. Grid cap
21. Base
22. Completed tube

BELOW

1. Tube for stem
2. Flare for stem
3. Pump tube
4. Leading-in wires
5. Stem
6. Finished stem
7. Heater filament
8. Complete cathode
9. Control grid
10. Plate
11. Insulating support
12. Insulator anchor
13. Filament hook
14. Getter cup
15. Complete mount
16. Bulb
17. Sealed and exhausted tube
18. Base
19. Completed tube

(*Facing page* 14)

tubes and receiving sets so that the sets could be entirely operated from a-c. lighting mains. At first, various types of B-battery eliminators appeared. These were devices making use of vacuum tubes or electrolytic rectifiers, operated from the house-lighting mains, to supply the high voltage for the plate circuits of the tubes. Then tubes appeared having separate filaments operated on low-voltage alternating current obtained from step-down transformers, these filaments serving as heaters for the emitting cathodes. From about 1927, receiving sets appeared which were operated entirely from the house-lighting mains and by the end of 1928, only about 11 per cent of the sets manufactured were battery-operated sets.

The experiment of 1915, in which speech was transmitted across the Atlantic Ocean, showed that the vacuum tube, when used in large numbers and connected in parallel, was capable of generating electric oscillations of large power. Because the use of a large number of small tubes is both inefficient and inconvenient, and believing that the trend of development would be toward more and more power, the engineers of the Western Electric Company began experiments with a view of increasing the power of vacuum-tube oscillators. With the ordinary type of tube, with internal plate, all of the heat generated at the plate is dissipated by radiation alone. The practical limit to the size of tubes of this type is from 1 to 2 kw. With the invention by W. G. Housekeeper of a method of sealing copper to glass, tubes were made in which the plate was a copper tube serving as a portion of the containing vessel as well. With this design the plate can be water- or oil-cooled, and practical tubes up to 500-kw. capacity each have been made.

Perhaps enough has been said to prove that the vacuum tube has become one of the most important of electrical devices. To emphasize further this statement, one might visit a modern broadcasting station where one finds vacuum tubes used as oscillators, modulators, amplifiers, telltale devices of various sorts, and as power rectifiers for supplying the necessary d-c. power for the oscillators and modulators. Again, an examination of a modern receiving set reveals vacuum tubes performing the functions of detector, radio amplifier, audio amplifier, oscillator, and in many cases one tube may perform two or more of these functions simultaneously. Finally, an inspection of a modern telephone exchange reveals hundreds of vacuum tubes function-

ing as repeaters; and in some exchanges where the carrier system is used, vacuum tubes are seen operating for many different purposes.

In this historical outline, emphasis has been placed upon the use of vacuum tubes in the field of communication. Although the communication field is the field in which the vacuum tube finds its widest application, there have been noteworthy advances in other applications of the vacuum tube.

Without going into detail, mention may be made of the use of the triode as a vacuum gauge, as a very convenient voltmeter of inestimable value in research, as a relay and a control device, etc. The two-electrode tube has found useful application as both high-power and low-power rectifier. The photoelectric cell, one form of diode, has been developed into a valuable tool in television and in investigation generally. One might go on to mention many other developments, but it seems best not to extend this outline but to leave further description of the development and uses of the vacuum tube to the more detailed discussion in the following chapters.

CHAPTER II

MOLECULES, ATOMS, AND ELECTRONS

4. Matter.—It has long been known and is now universally recognized that all matter is made up of submicroscopic pieces, these pieces being the smallest particles into which the mass can be divided and still retain the properties of the original mass. These unitary blocks of the mass are called *molecules*. Each kind of matter has its own distinctive molecule.

Many kinds of molecules are chemically complex and in the last analysis can be split up into elementary chemical substances which are called *chemical elements*, of which there are ninety-two. All of these elements have been isolated with the possible exception of two about which there is at present some doubt. When the molecule is thus split up, the original matter loses its identity. The smallest particles or units of these chemical elements are called *atoms*. Thus we see that there are known ninety-two kinds of atoms, such as atoms of hydrogen, sulphur, carbon, iron, copper, and lead.

From a physical standpoint we recognize three states of matter: solid, liquid, and gaseous. A *solid* is a substance in which the cohesive force between the molecules is so great that the substance is rigid and has a definite boundary. A *liquid* is a state of matter in which the attraction between molecules is not so great as in solids and the molecules are not fixed in relative position. Although a liquid has a definite boundary, it is not rigid but flows readily. Finally, a *gas* is a state of matter in which the molecules are so far apart that there is very little attraction between them. A gas ordinarily has no definite boundary.

5. Kinetic Theory.—According to the kinetic theory,[1] the heat that is contained in a portion of matter exists as the kinetic energy of motion of the molecules. That this continuous random motion of the molecules actually takes place is sufficiently proved indirectly by experiment. Of course, all the molecules in a given

[1] The reader is referred to Chap. I of Dushman's book "High Vacua," or to the *Gen. Elec. Rev.*, **15**, 952, 1042, 1159 (1915), for a brief treatment of the kinetic theory.

body do not possess the same kinetic energy, but corresponding to every temperature there is a mean kinetic energy; some molecules have energies greater and some less than this mean kinetic energy. The higher the temperature, the greater is this mean kinetic energy and the greater the mean velocity of motion of the molecules. This subject will be treated in more detail in Chap. IV on Emission of Electrons.

Many physical properties of a gas can be explained in terms of the kinetic theory. For example, the pressure of a gas in an enclosure is due to the bombardment of the walls of the enclosure by the rapidly moving molecules. The variation of pressure with volume and temperature is satisfactorily explained by this theory. As further examples of the applications of the kinetic theory, diffusivity, viscosity, vapor pressure, heat conductivity, and other properties are deduced and expressed as functions of temperature and pressure.

TABLE I.—MEAN FREE PATH OF GAS MOLECULES

Gas	Pressure = 1 bar		Pressure = 10^6 bars	
	0°C., centimeters	25°C., centimeters	0°C., centimeters	25°C., centimeters
H_2	....	19.2	.....	19.2×10^{-6}
He	....	29.6	.....	29.6×10^{-6}
N_2	....	10.0	.....	10.0×10^{-6}
O_2	....	10.7	.....	10.7×10^{-6}
A	....	10.6	.....	10.6×10^{-6}
CO	....	9.92	.....	9.92×10^{-6}
CO_2	....	6.68	.....	6.68×10^{-6}
Hg	3.24	.....	3.24×10^{-6}	
H_2O	6.03	.....	6.03×10^{-6}	

6. Mean Free Path.—One of the most important factors which enters into the explanation of heat and electrical conductivity, diffusivity, ionization, etc., is the mean free path of molecules and electrons in matter. The *mean free path* may be defined, with sufficient accuracy for our purpose, as the mean distance a particle travels between collisions. The mean free path of molecules of a gas at constant temperature varies inversely as the pressure but depends upon the particular gas. The mean free path is

connected directly with the coefficient of viscosity of the gas and increases somewhat with an increase of temperature.

Table I gives the values of the mean free path L for several gases at 0°C. or at 25°C., for a pressure of 1 bar and for 10^6 bars.[2]

Figure 3 is a logarithmic plot of the mean free path against the pressure in bars for a few gases at 20°C.

The mean free path of an electron, a particle very small in comparison with the size of the molecule of a gas, is considerably

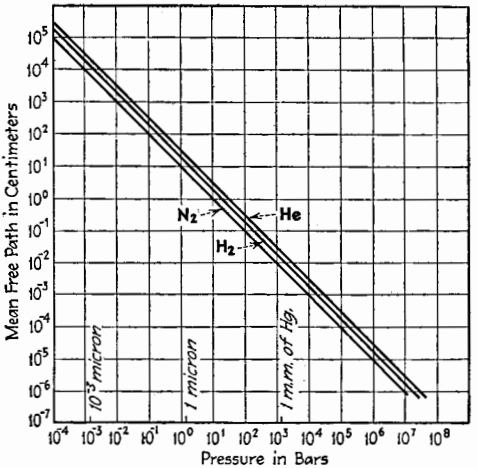


FIG. 3.—Mean free path of gas molecules of hydrogen, helium, and nitrogen at 20°C.

greater than the mean free path for the molecules themselves. Maxwell deduced that the mean free path of a particle which is small compared with the molecule is $4\sqrt{2}L$, where L is the mean free path of the molecule of the gas. This value has been used for the mean free path of electrons. Investigations have shown that the mean free path of electrons varies with their velocity. Ramsauer[3] has shown that the mean free path of electrons decreases as their velocity decreases and may pass through one or more minima and finally increase when the velocity of the

[2] 1 bar is a pressure of 1 dyne per square centimeter. One bar is, therefore, a pressure of 0.00075 mm. of mercury, and 0.001 mm. of mercury or 1 micron is a pressure of 1.333 bars. One standard atmosphere is a pressure of 760 mm. of mercury, so that 1 bar is 1.0133×10^{-6} atmosphere.

[3] RAMSAUER, *Ann. d. Physik*, **64,** 513 (1921); **66,** 546 (1921); **72,** 345 (1923).

electrons is very low. As a consequence of these studies, Maxwell's theoretical value of $4\sqrt{2}L$ for the mean free path is inaccurate when applied to electrons having low velocities.

7. Vacuum.—In order to get some sort of physical picture of a vacuum, first consider the structure of a gas. There are about 2.52×10^{19} molecules of gas per cubic centimeter at a temperature of 20°C. and a pressure of 760 mm. of mercury. The atom of hydrogen is roughly 10^{-8} cm. in diameter. Therefore, only about one-millionth of the space in hydrogen gas under the above conditions of temperature and pressure is occupied by the atoms. Although this calculation is admittedly very crude, because it assumes that the hydrogen molecule consists of two spherical atoms, it is safe to say that there are great voids between the molecules even at atmospheric pressure. When a gas is compressed, it is this empty intermolecular space that is reduced.

When the pressure of a gas is reduced, the empty space increases directly in proportion to the reduction in the pressure. The term *vacuum* is a very indefinite term and means that the pressure of gas in an enclosure is reduced below atmospheric pressure and usually connotes a considerable reduction in pressure. For example, if the pressure is reduced to one-millionth of atmospheric pressure, the vacuum is said to be fairly good or high. The pressure would then be 0.000760 mm. of mercury or 0.987 bar. The hydrogen molecules at this pressure occupy roughly one-millionth of a millionth of the space, which makes such a vacuum seem quite free of matter.

Examine this vacuum from another point of view. We find that at the pressure of 0.987 bar there are 2.52×10^{13} molecules of any gas per cubic centimeter, or 2.52×10^{10} molecules per cubic millimeter. This view of the vacuum gives one the idea that the vacuous space is very much occupied.

The modern developments of vacuum pumps permit a pressure of 0.0001 bar to be attained. Vacua corresponding to pressures less than 1 bar are usually classed as high vacua.

8. Atoms.—The physical picture of the structure of the atom has undergone many changes during the last few decades, and even at present physicists are none too sure of its exact structure. Up to a few years ago, physicists entertained the Bohr or miniature-solar-system picture of the atom, in which *discrete* unitary charges of negative electricity called *electrons* rotate in definite orbits around a positively charged central structure called the

nucleus. This picture has been immensely valuable in the explanation of a great many experimental results, such as the X-ray and ordinary line spectra of the elements, conduction of electricity through gases, chemical phenomena, and many others. The idea of discrete electrons has been attested by many other convincing experiments too numerous to list here. Certain recent experiments, among which may be mentioned those of Davisson and Germer[4] on the reflection of electrons from crystals of metal, and of G. P. Thomson[5] on the diffraction of electrons, cannot be explained by the theory of discrete electrons alone, but can be explained if the electrons possess the properties of waves whose length is determined by their velocity. It then becomes necessary to consider, as L. de Broglie[6] first suggested, that the electron has associated with it waves which manifest themselves in certain phenomena. This idea, together with the theoretical work of Schrödinger, Heisenberg, and others, has given rise to the theory of wave mechanics, which has been eminently successful in more completely correlating theory and experiment.

The result of these newer ideas is to spoil to a certain extent the definite picture of the atom by surrounding it with a certain indefiniteness. The electron may not be quite so discrete as it at first was pictured, and the positions of the electrons in the atom are expressed in terms of probabilities rather than by definite coordinates. For the purposes of this book, it seems best to adhere to the idea of discrete electrons, which are, however, accompanied by waves, and to use the picture of the atom which has been so helpful in explaining many phenomena of physics, *i.e.*, that of the miniature solar system, to be described more in detail below.

The normal atom has a number of electrons revolving about its nucleus, the number being different for the different elements, starting with hydrogen possessing one electron, helium with two, and so on in orderly sequence up to the heaviest element uranium, with ninety-two electrons. The number of electrons in the normal atom is the atomic number z of the element.

[4] DAVISSON, *Bell System Tech. J.*, **7**, 90 (1928); DAVISSON and GERMER, *Proc. Nat. Acad. Sci.*, **14**, 317 (1928); **14**, 619 (1928).

[5] THOMSON, *Nature*, **122**, 279 (1928); *Proc. Roy. Soc.*, **117**, 600 (1928); and **119**, 651 (1928); *Phil. Mag.*, **6**, 939 (1928).

[6] DE BROGLIE, *Ann. d. Physik*, **10**, 322 (1925).

Since the normal atom is electrically neutral, the nucleus possesses a positive charge equal to z times the charge of a single electron. The nucleus is probably rather complex in structure, as might be expected from the following facts: the mass of the atom is almost entirely associated with the nucleus; radioactive phenomena arise from the nucleus; and some of the properties of the atom are attributed to the structure of the nucleus. It is known that the nuclei of all of the elements except hydrogen contain electrons as well as positive electricity.

The size of the atom is more or less indefinite but may be considered to be roughly the size of the outer electronic orbits. The electrons and nucleus are exceedingly minute compared to the size of the atom. Here again we have difficulty in assigning any definite size to either the electron or the nucleus, because they have no definite boundary. To get a rough idea of the relative size of the atom, the electron, and the nucleus, suppose the normal hydrogen atom to be magnified until the diameter of the orbit of the single electron is 100 ft. Then the nucleus would have a diameter less than 0.01 in., and the electron, if it has any size, would probably be as small as the hydrogen nucleus. Many of the atoms of higher atomic number are larger than the hydrogen atom because of the larger electronic shells, and also the nuclei of these heavier atoms are larger than the hydrogen nucleus. Even if we cannot assign definite dimensions to the components of an atom, it is apparent that there is much empty space in atoms as well as in matter. This picture is valuable when considering the penetration through matter of fast-moving electrons and alpha particles from radioactive substances.

The electrons revolve around the nucleus of a normal atom in more or less definite orbits. If the atom is violently disturbed by certain external agencies, one or more of the electrons may be displaced so as to revolve in larger orbits, the sizes of which have certain fundamental relations with one another, as dictated by the quantum theory. The atom is then said to be *excited.* When any electron jumps from one of its large orbits, occupied only when the atom is excited, back to any intermediate orbit or back to its normal orbit, the atom gives out radiation. This radiation may be in the visible region of light, it may be of shorter wave length known as ultra-violet light, or, if still shorter, known as X-rays.

If the atom is unhampered in its behavior by other atoms, as, for instance, if the atom is one of a gas or vapor, the radiation emitted has definite wave lengths determined by the particular orbital jumps. In general, when the outermost electrons are raised to larger orbits and fall back, the radiation is visible or in the near ultra-violet, but when the innermost electrons are disturbed, the wave length of the radiation is shorter and the radiation may be in the X-ray range.

Many of the physical properties of the atoms, such as the optical and X-ray spectra and ionizing potentials, together with the periodic nature of many of the physical and chemical properties, are explained if the electrons in an atom are confined within definite shells. A study of the various properties of atoms leads to the allocation of two electrons to the first or innermost shell, eight to the second and third shells, eighteen to the fourth and fifth shells, thirty-two to the sixth shell, and the remaining electrons to the seventh shell in elements having atomic numbers greater than eighty-six. For example, the atom of helium has two electrons so that the first shell is complete. The next element, lithium, has three electrons, the third electron being alone in the second shell. The chemical properties of an atom are largely dependent upon the stray electric field from the outer-shell electrons. Lithium is very active chemically because of the single electron in the outer shell, while helium with its completed outer shell is chemically inert. The second shell is completed with the inert gas neon, the third shell is completed with the inert gas argon, the fourth with krypton, and the next with xenon. The atom, whose outer shell lacks one electron to complete it, such as that of fluorine and of chlorine, is also chemically active because of its strong affinity for the electron which would complete the shell.

9. Molecules.—The smallest unit of an element or compound that can exist in a normal state and possess the properties of the element or compound is called a *molecule.* For certain elements, such as helium, argon, neon, and mercury, the atom is the molecule. Such elements are called *monatomic.* The molecules of the other elements and of all compounds are made up of atoms held together by the electric forces due to the charges of which the atoms are composed. The molecules of hydrogen, oxygen, nitrogen, iodine, and some others are composed of two atoms and are called *diatomic* molecules. It happens that most

of the metals and permanent gases are monatomic and most of the chemically active gaseous elements are diatomic. Molecules of some elements have more than two atoms, as, for instance, the molecule of phosphorus has four atoms, the molecules of selenium and of sulphur are composed of eight atoms.

The molecules of compounds are often very complex, being composed of atoms of different elements. Sometimes, as with hydrochloric acid (HCl), carbon monoxide (CO), potassium chloride (KCl), the molecules are diatomic, but more often there are more than two atoms in the molecule, and there may be as many as ten, twenty, or even more atoms in complex molecules.

10. The Electron.—Although the atom is the natural habitat of electrons, an electron can be detached from an atom by various physical agencies. The electric forces in a beam of radiation, especially ultra-violet light and X-rays, falling upon the surface of matter may pull electrons from the surface atoms. This phenomenon is known as the *photoelectric effect.* Again, certain radioactive atoms spontaneously explode or, more properly, disintegrate, throwing off electrons known as *beta rays.* Most of the electrons from radioactive substances come from the nuclei uninfluenced by any known physical agency, but one or more of the electrons in the outer shell of a normal atom may be knocked off when the atom is struck by a rapidly moving electron or by another atom. The latter process is known as *ionization,* the end products being one or more freed electrons and a *positive ion,* which is the remaining positively charged atom.

The electrons of metallic atoms and of certain other atoms seem to be very loosely bound and so can easily wander from atom to atom when the atoms are close together as in a piece of metal. These loose electrons, although probably only temporarily free, are known as *free electrons* and move more or less freely through the matter when an electric force is applied. This motion constitutes a current of electricity and the matter is said to be a *conductor.* If the temperature of conducting matter is sufficiently elevated, the velocities acquired by some of these free electrons, due to their heat agitation, are sufficient to project them outside the sphere of attraction of the stationary nuclei, and we have what is known as *thermionic emission.*

Our information concerning the properties of the electron has come largely from the researches of J. J. Thomson and his students in the Cavendish Laboratory at Cambridge, England,

and from the investigations of many other eminent physicists, such as Townsend, Rutherford, Millikan, C. T. R. Wilson, Langmuir, and K. T. Compton. Thomson was the first to show that the electron possesses mass and that the ratio e/m_0, where e is the charge and m_0 the mass at rest, has the same value for *all* electrons. The value of e/m_0, corrected by more recent investigation, is

$$\frac{e}{m_0} = 1.758 \times 10^7 \text{ absolute electromagnetic units}$$

This value of e/m_0 is about 1,840 times the value for the hydrogen ion, the smallest known positive ion, consisting only of a single charged hydrogen nucleus. Since it is an experimental fact that the charge of the electron is the same as the charge of the positive ion, the mass of the electron must be about 1/1,840 the mass of the smallest positive ion. The electrons, therefore, contribute very little to the mass of the atom, confirming the statement previously made that the mass of the atom is largely associated with its nucleus.

The value of the charge on an electron has been measured by several methods. The best value is probably that obtained by Millikan[7] in a remarkably ingenious series of experiments begun in 1909 and is

$$e = 4.770 \times 10^{-10} \text{ absolute electrostatic unit}$$
$$e = 1.591 \times 10^{-20} \text{ absolute electromagnetic unit}$$

Knowing the value of the charge, the mass can be calculated from the known ratio e/m_0. The mass of a stationary electron is

$$m_0 = 0.9035 \times 10^{-27} \text{ gram}$$

According to the theory of relativity, the mass of an electron moving rapidly with respect to the observer increases as its velocity approaches that of light and becomes infinite at the velocity of light. The value of m for any velocity v is

$$m = \frac{m_0}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \tag{1}$$

where c is the velocity of light. This increase in mass has been checked experimentally by Kaufmann,[8] Bucherer,[9] and

[7] MILLIKAN, *Phil. Mag.*, **19**, 209 (1910).
[8] KAUFMANN, *Ann. d. Physik*, **19**, 487 (1906).
[9] BUCHERER, *Ann. d. Physik*, **28**, 513 (1909).

others,[10] and lately by the author.[10] This change of mass of a moving electron does not much concern us, however, for there is very little change unless the voltage drop acting on the electron is greater than about 10,000 volts.

11. Electric Field about an Electron.—Since a stationary electron has a charge, there exists about the electron a radial electrostatic field which extends indefinitely in all directions. In fact it is only by this electrostatic field that the presence of a stationary electron is made evident. About all we know of the structure of the stationary electron is that it is the point of convergence of its electrostatic field or lines of force.

12. Moving Electron and Its Magnetic Field.—A moving electron is a moving charge and hence constitutes a transfer of electricity, which is an electric current. Since an electric current has an associated magnetic field, the motion of the electron with its associated electrostatic field sets up a magnetic field. The mass of the electron is directly assignable to the energy and inertia of this magnetic field.

As already stated, another attribute of a moving electron is a system of waves associated closely with the electron and having a wave length dependent upon its velocity and given by the relation

$$\lambda = \frac{h}{mv} \tag{2}$$

where h is Planck's radiation constant, m the mass of the electron, and v its velocity.

13. Forces Acting upon a Moving Electron.—An electron which is acted upon by an electrostatic field moves in the opposite direction to that of the field itself, because the conventional direction of the field is taken as the direction in which a free positive charge would move. The electron is urged by a force whose magnitude is equal to the product of the strength of the field and the electronic charge. Since the electron has a mass m, it is accelerated and acquires kinetic energy according to the familiar mechanical laws for accelerated motion. If, however, in its flight through free space it collides with an atom, its energy may be in part lost to the bombarded atom. If an electron moves between two points having a difference of potential E, as between two electrodes of a vacuum tube, the work done on the

[10] PERRY and CHAFFEE, *Phys. Rev.*, **36**, 904 (1930). See bibliography for other recent determinations of e/m.

electron by the field caused by E, whether the field is uniform or not, is, by definition of potential, Ee units of work. The work is in ergs if E is expressed in absolute electrostatic or electromagnetic units and e in terms of the corresponding absolute unit of charge. If the electron encounters no atoms between the points or electrodes, the electron acquires kinetic energy and hence a velocity v, given by the expression

$$Ee = \int_0^v v\, d(mv). \tag{3}$$

Substituting in Eq. (3) the value of m from Eq. (1),

$$Ee = m_0c^2\left[\frac{1}{\sqrt{1 - (v/c)^2}} - 1\right] \tag{4}$$

If v/c is small, Eq. (4) becomes

$$Ee = \tfrac{1}{2}m_0v^2 \tag{5}$$

If the electron is suddenly stopped after acquiring a velocity v, the energy, expressed by Eq. (4) or Eq. (5), is given up by the electron and usually appears as heat. The high temperature of the plate of a power vacuum tube is due to the heat produced by the impinging electrons.

A moving electron constitutes a current flow of magnitude proportional to the charge of the electron and to its velocity. If the electron moves a distance Δl in time Δt, the rate of transfer of electricity is $e/\Delta t$, which in proper units is the strength of the current. The product $e/\Delta t \cdot \Delta l$ is therefore equivalent to a current $i = e/\Delta t$ flowing in a path of length Δl. Hence

$$i\Delta l = e\frac{\Delta l}{\Delta t}$$

or

$$i\Delta l = ev \tag{6}$$

From elementary principles of electricity, the force acting on a length dl of wire carrying a current i, when the wire makes an angle θ with a magnetic field H, is

$$\Delta f = Hi \sin\theta\, \Delta l \tag{7}$$

The direction of the force is perpendicular to the plane containing Δl and H, as shown in Fig. 4*a*. Using the equivalence given in Eq. (6), the force on an electron is,

$$f = Hev \sin\theta \tag{8}$$

The direction of motion of the electron is opposite to the conventional flow of a positive current (see Fig. 4*b*). If the magnetic field is uniform and θ is equal to $\pi/2$, the acceleration perpendicular to the motion of the electron is constant and the electron moves in a circular path. The path is helical if θ is not equal to $\pi/2$.

Positive and negative ions, being charged masses, follow the above laws in electric and magnetic fields. Their masses being at least 1,840 times greater than that of an electron, the velocities which the ions acquire are much less than that acquired by an electron under the same conditions.

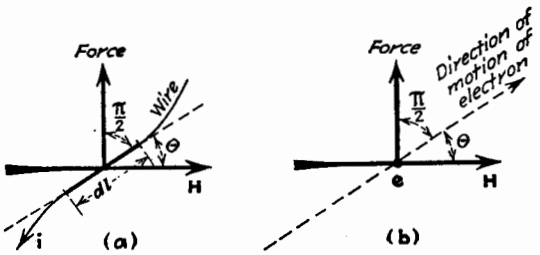


FIG. 4.—Force on an element of wire carrying a current and on a moving electron, both in a magnetic field.

14. Collisions between Electrons and Molecules.—An electron moving in a space in which molecules are present is said to encounter or collide with a molecule when it comes into the force field of the molecule to an extent sufficient to alter materially the path of the electron, or to evoke some change in the molecule.

If the molecule is monatomic, such as those of the inert gases, and the velocity of the electron is low, the encounter is analogous to that between elastic spheres. Applying the conservation of momentum and energy equations, the fraction of the initial energy lost by the electron if the molecule is initially at rest, as calculated by ordinary laws of mechanics, is

$$\frac{W_{\text{lost}}}{W_{\text{initial}}} = \frac{4Mm}{(M+m)^2} \cdot \cos^2 \theta \tag{9}$$

where M and m are the masses of the molecule and electron, respectively, and θ is the angle between the trajectory of the electron and the radius of M to the point of impact. Since a large number of electrons distributed uniformly in space will

strike molecules with all values of θ between $-\pi/2$ and $+\pi/2$, the average fractional loss of energy for all the electrons is

$$\text{Average } \frac{W_{\text{lost}}}{W_{\text{initial}}} = 2.66\frac{Mm}{(M+m)^2} \tag{10}$$

If the molecule is in motion owing to the temperature of the gas, the average loss of energy is less than that given by Eq. (10). Since M is always more than 1,840 times greater than m, even the maximum loss of energy, when θ is equal to zero, as given by Eq. (9), is very small. Therefore, a low-velocity electron encountering only a few monatomic molecules loses a negligible amount of energy. The electron, however, suffers a change in direction of motion at each encounter.

If the gas molecule encountered by the slowly moving electron is diatomic instead of monatomic, the electron may impart to the molecule rotation as well as translation, with the result that the colliding electron may lose much more energy than in the case of the monatomic molecule.

So far we have considered encounters between molecules and slowly moving electrons. As the velocity of the electron is increased, at a certain velocity, different for each kind of molecule, the collision suddenly becomes inelastic, owing to a change in the structure of the molecule induced by the collision. At this critical velocity, the electron loses all of its energy, the energy being spent in causing an electron in an atom of the molecule to rotate in a larger orbit or to be completely detached from the molecule. The first of these events is known as *resonance* and the second as *ionization.*

15. Resonance.—As has been stated, the electrons in a normal atom rotate in more or less definite orbits. Any electron in the outer shell of the atom may be induced temporarily to rotate in certain other orbits, larger than its normal orbit. So, too, there is a group of other orbits, larger than the normal orbit, in which any electron of the next innermost shell can rotate, and so on, for all of the shells of the atom. When an electron is in any other than its normal orbit, it has associated with it energy greater than that corresponding to its normal orbit. The atom can absorb into its structure only parcels of energy just sufficient to change an electron from one permitted orbit to another. In other words, the energy absorbed can be only that equal to the difference in energy corresponding to any temporary orbit and

the normal orbit for that electron, or between only certain of the temporary orbits.

When one or more than one electron is in a temporary orbit, the atom is said to be in an excited state. The displaced electron tends to return to its normal orbit. It may do so in one jump, in which case the difference in energy is radiated as light or X-rays, the frequency being given by the relation

$$h\nu = W_2 - W_1 = \Delta W \tag{11}$$

where ν is the frequency of the radiation, h is Planck's constant and equal to 6.547×10^{-27} erg sec., and $W_2 - W_1$ is the difference in energy corresponding to the two orbits. On the other hand, the electron may return by way of one or more of the smaller temporary orbits, thus making two or more energy jumps. Each energy jump gives rise to a monochromatic radiation in accordance with Eq. (11), in which ΔW is the change in energy. An atom ordinarily remains in an excited state for a very short time, estimated to be of the order of 10^{-8} sec. With some atoms, however, certain temporary orbits seem to harbor an electron for a much longer time, even as long as a tenth of a second in some cases. These particular excited states are known as *metastable states.*

The energy required to excite an atom may be derived from radiation, in which case Eq. (11) gives the frequency ν_i required to cause an energy change ΔW in the atom, ΔW being the energy difference corresponding to two permitted orbits. The term *resonance* is applied to the excitation of an atom by radiation of that frequency which would be emitted by the atom if the same energy jump were made as occurs when the atom returns from the excited state to the normal state. It is easy to see how resonance radiation may "diffuse" through a gas by the reversible process of excitation and emission.

An atom may also be excited if an electron or positive ion collides with it with sufficient energy. The exciting projectile must possess at least an amount of kinetic energy equal to the smallest energy jump between orbits. If the projectile is an electron, the potential drop which would give this necessary kinetic energy is known as the *first resonance potential* of the atom. An atom may possess several resonance potentials given by the relation

$$Ee = \Delta W \tag{12}$$

where E is the resonance potential, e is the electronic charge, and

ΔW is a possible energy jump for the atom. As might be expected, an atom may be not only excited from its normal state but raised from one excited state to another, although the probability that a colliding electron may find an atom in an excited state is very small because of the short time of duration of an excited state. Of course, the probability of such an event is much greater if an atom is in a metastable state.

As examples of resonance potentials we shall cite only a few for some of the common gases. Others can be found in the "International Critical Tables" and in many books on this subject. The resonance potential for sodium vapor corresponding to the first excited state is 2.1 volts, and the resonance radiation is the familiar D line of the sodium spectrum. Mercury vapor has several resonance potentials corresponding to various excited states. These potentials are 4.66, 4.86, 5.43, 6.67, 7.69, 8.58 and 8.79 volts. The state corresponding to 4.66 volts is one of the metastable states.

16. Ionization.—If the exciting energy acting upon an atom is sufficient, an electron may be completely detached from the atom. This event is known as *ionization* and may be thought of as a limiting excited state for which the largest permitted orbit is at infinity. The positively charged part of the atom that remains is a positive ion. The potential which would impart to the ionizing electron the necessary energy is known as the *ionizing* potential and fits into Eq. (12), where ΔW is the difference between the energy of an electron at infinity and an electron in the normal orbit.

The simplest ionization act consists in the removal of a single outer electron, although more than one electron may be removed; and if the energy of the impacting electron is great, one of the inner electrons may be removed.

The ionization potentials of the alkali metals, which have only one or two electrons in the outer shell, are particularly low. For sodium the ionizing potential is 5.12 volts, for potassium 4.32 volts, and for caesium 3.9 volts. On the other hand, the inert gases with complete outer shells have high ionization potentials. For example, that for helium is 24.5 volts, for neon 21.5 volts, and for argon 15.7 volts.

An atom, by absorbing just the right amount of energy, may be ionized from an excited state, particularly from a metastable state, instead of from the normal state.

As an electron enters an ionized atom under the action of the electrostatic force, it may reach its normal orbit by one jump or by successive steps, stopping momentarily in one or more of the temporary orbits. Light is emitted in accordance with Eq. (11) for every energy jump executed. Hence, ionized gas is continually recombining, and accompanying this recombination there is luminosity, the light being the characteristic line spectrum of the gas.

The probability that an electron will ionize a molecule increases from zero, for velocities less than the ionizing velocity, up to a maximum when the energy of the electron is two or three times this ionizing energy. Thereafter the probability decreases for electrons of greater velocities. Hence, high-velocity electrons ionize less efficiently than those having velocities corresponding to from 50 to 200 volts. The probability also depends upon the pressure of the gas and the distance the electron travels. One way of expressing this varying probability of ionization is by plotting the number of positive ions produced by one electron traveling 1 cm. in a gas at 1-mm. pressure. The result for other pressures and lengths of path is proportional to the pressure and length of path. Various measurements have been made of this important quantity for several gases. Hughes[11] was one of the first to make these measurements. The plot given in Fig. 5 is a compilation of the results obtained by Compton and Van Voorhis,[12] Jones,[13] Bleakney,[14] and Buckmann.[15]

Positive ions colliding with a molecule are very inefficient in producing ionization. To produce the same number of ions per centimeter of path, the positive ions must be moving with a velocity comparable to that of an electron, which means that the energy of a positive ion must be thousands of times greater than that of an electron in order to ionize as efficiently.

17. Reflection of Electrons and Secondary Emission.—When electrons strike the surface of a solid, some electrons are reflected. The percentage reflected depends upon the angle, the kind of matter bombarded, the condition of the surface, and the velocity of the electron. Ordinarily the percentage reflected ranges

[11] HUGHES, *Washington University Studies*, 1924.
[12] COMPTON and VAN VOORHIS, *Phys. Rev.*, **26**, 436 (1925); **27**, 724 (1926).
[13] JONES, *Phys. Rev.*, **29**, 822 (1927).
[14] BLEAKNEY, *Phys. Rev.*, **35**, 139 (1930).
[15] BUCKMANN, *Ann. d. Physik*, **87**, 509 (1918).

from a few per cent up to perhaps 20 or 30 per cent, but in special cases more may be reflected. One would hardly expect that the reflection of electrons would be specular in nature considering the smallness of the electron in comparison with the surface atoms. The fact is, as shown by the angular distribution curves of Davisson and Kunsmann,[16] that ordinarily the reflection is diffuse, the actual distribution depending upon conditions of the experiment. Davisson and Germer[17] found that if the reflection takes place at the surface of a crystal of nickel, a small fraction

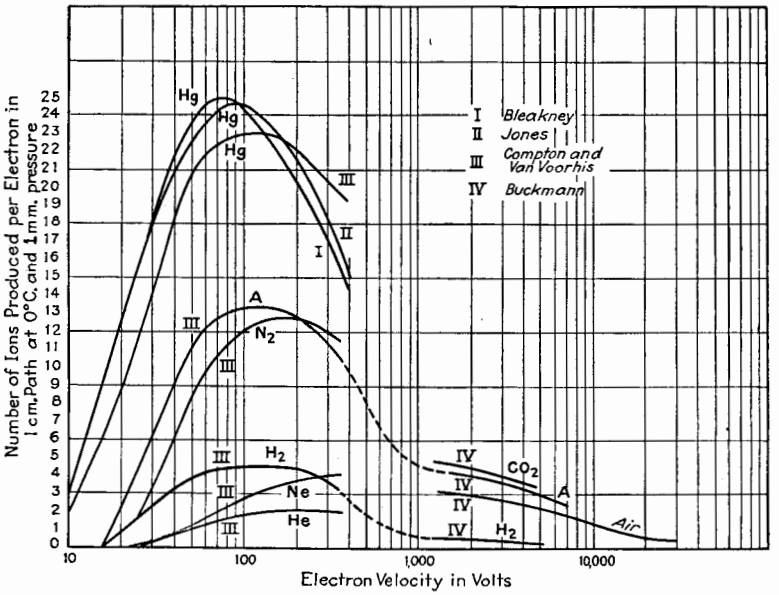


FIG. 5.—Probability of ionization.

of high-velocity electrons are specularly reflected at a certain angle which depends upon the velocity of the impinging electrons. This type of reflection is explained if the reflection of these electrons is determined by an associated system of waves of length given by Eq. (2).

In addition to the reflection of electrons striking a solid, something analogous to ionization also takes place; *i.e.*, elec-

[16] DAVISSON and KUNSMANN, *Science*, **64**, 522 (1921); *Phys. Rev.*, **22**, 242 (1923).

[17] DAVISSON and GERMER, *Phys. Rev.*, **30**, 705 (1927); *Proc. Nat. Acad. Sci.*, **14**, 317 (1928); DAVISSON, *J. Franklin Inst.*, **205**, 597 (1928).

trons are knocked out of the surface. This phenomenon is known as *secondary emission* and the electrons derived from the surface are known as *secondary electrons* or *delta rays*. These secondary electrons leave the surface with very low velocities corresponding to only a few volts, and their number depends greatly upon the material of the surface, its physical condition, and the velocity of the primary electrons. Secondary emission will be treated in greater detail in Chap. IV.

General References

TOWNSEND: "Electricity in Gases," Oxford University Press, New York, 1915.

MILLIKAN: "The Electron," Chicago University Press, Chicago, 1917.

CROWTHER: "Ions, Electrons, and Ionizing Radiation," Edward Arnold & Co., London, 1919.

SOMMERFELD: "Atomic Structure and Spectral Lines," Methuen & Co., London, 1923.

DARROW: "Introduction to Contemporary Physics," D. Van Nostrand Company, New York, 1926.

THOMSON, J. J.: "Conduction of Electricity through Gases," 3d ed., Cambridge University Press, Cambridge, 1928.

HOAG: "Electron Physics," D. Van Nostrand Company, New York, 1929.

RUARK and UREY: "Atoms, Molecules, and Quanta," McGraw-Hill Book Company, Inc., New York, 1930.

COMPTON and LANGMUIR: "Electrical Discharges in Gases," *Rev. Modern Phys.*, Vol. 2, No. 3 (1930).

DARROW: "Electrical Phenomena in Gases," Williams & Wilkins Co., Baltimore, 1932.

CHAPTER III

CONDUCTION OF ELECTRICITY

The operation of all types of vacuum tubes is directly dependent upon the characteristics of the conduction of the electric current through the tube. The object of this chapter is to give the reader a brief treatment of the principles of conduction of electricity under various circumstances, but with special emphasis on the two types of conduction which most concern us in dealing with vacuum tubes, namely, conduction through a high vacuum and conduction through a space containing an appreciable amount of gas.

18. The Conventional Direction of an Electric Current.— The motion of an unneutralized charge of electricity, whether the charge is positive or negative, constitutes an electric current. Since, however, the electrical effects of positive and negative electricity are opposite, the motion of a negative charge in one direction is equivalent to the motion of an equal positive charge in the opposite direction.

When the terms plus and minus were originally allocated to the two kinds of electric charge, the minus sign was given to that charge which appears on sealing wax, hard rubber, and other resinous substances, when rubbed with fur, silk, etc. The positive sign was given to the charge which appears on vitreous substances like glass. This choice of signs was apparently entirely arbitrary. With the signs of the two electric charges thus defined, the charge of the electron is negative and the charge of the atomic nucleus is positive.

For mathematical discussions a convention is necessary for the direction of an electric current. The conventional direction of a current is the direction of flow of positive electricity. If a current of electricity is due to a motion of positive ions, the conventional direction of the current and the actual motion of the carriers of electricity are the same, but if, as is usually the case, the electric current is due to the motion of electrons, the conventional direction is *opposite* to the actual

direction of motion of the electricity. In the study of vacuum tubes, this difference between conventional direction and the actual direction of flow is always before us, because the current through a thermionic tube is practically all due to the flight of electrons inside the tube from the hot cathode to the positive electrodes. In order, however, to conform with the familiar electrical principles of an electric circuit, the arrows representing the currents are usually placed to indicate the conventional direction of flow.

19. Metallic Conduction.—The most familiar kind of electrical conduction is metallic conduction. Because metals are the best conductors of electricity, metallic wires are used for conveying electricity and electrical energy.

A characteristic of the atoms of metallic solids or liquids is that the outer electrons are loosely attached to the atom so that these outer electrons temporarily drift away from the atoms into the spaces between, giving rise to what are known as *free electrons* in the metal. Electricity flows through metals only because of the motion of these free electrons which, urged by the applied electric field, either pass between the atoms or jump from atom to atom. The direction of motion of the negative electrons constituting the current is opposite to the conventional direction of the current.

As the electrons move, urged by the electric field impressed upon the conductor, they frequently collide with the closely packed atoms, giving up at each collision or encounter a part or all of the kinetic energy acquired between encounters. This energy goes into heat known as *joulean heat*, and this heat is set free throughout the conductor carrying the current. The opposition to the flow of the current, due to these encounters with the atoms of the conductor, is known as electrical *resistance*. The quantitative measure of resistance is the ratio of the potential difference to the current. One of the characteristic properties of metallic conductors is that the electrical resistance is independent of the current, provided the temperature remains constant.

The rate of transfer of electricity, which is the magnitude of the current, is determined by the number of electrons moving and their mean velocity in the direction of the electric field. The actual mean velocity of any electron is relatively very small, being of the order of a few meters per second for a potential gradient of a volt per centimeter. This velocity of the electron

must not be confused with the velocity of an electrical impulse through the conductor. The velocity of an electrical impulse or signal is the velocity of an electric wave through or in the space surrounding the conductor. It depends upon the configuration of the conductor and may be the velocity of light. For example, the velocity of an impulse in a circuit consisting of two parallel wires in free space is the velocity of light.

20. Electrolytic Conduction.—The conduction of electricity through a class of liquids known as *electrolytes* is entirely different from metallic conduction. Electrolytes are solutions of chemical compounds some of whose molecules are split up in solution into electrically charged parts known as *electrolytic ions*. The positively charged ions travel through the liquid in the conventional direction of the current, while at the same time the negatively charged electrolytic ions travel in the opposite direction. The velocities of the positive and negative ions for unit potential gradient vary somewhat according to the electrolyte, but are roughly the same. This type of conduction does not concern us here.

21. Gaseous Conduction.—We now come to a very important but very complex type of conduction of electricity, namely, conduction through a space which contains gas in a sufficient amount to play an important rôle in the conduction process. Much is still unknown concerning this type of conduction. Still, so much is known about gaseous conduction that in the space which is reasonably allotted to this subject we can give only the briefest outline of those principles which are necessary for the understanding of the operation of certain types of vacuum tubes. For more information on this subject the reader may consult the references at the end of this chapter.

Electricity can flow between two electrodes in a gaseous space only if there are carriers of electricity, *i.e.*, electrons or gaseous ions, present or produced in the space. The current then consists of the drift of negative carriers in the direction toward the positively charged electrode, accompanied by the drift of positive ions, if they exist, in the opposite direction. The electric current or discharge in the gas takes on different aspects and has different properties according to the number of carriers and the way in which they are produced and supplied to the discharge space. The characteristics of the discharge also depend greatly upon the pressure and kind of gas, the size, shape,

and distance between electrodes, and the size and shape of the enclosure.

Before describing the different types of discharge, we shall consider briefly the source of the electrons and ions which are so necessary for the initiation and maintenance of every gaseous discharge. First, we should point out that normally there exist in a gas very few ions, in number of the order of 1,000 per cubic centimeter at atmospheric pressure. These ions may be produced by small amounts of radioactive material in the neighborhood or by the penetrating cosmic radiation. When an electric field is impressed between electrodes, these stray ions are swept to the electrode of opposite sign to their charge and give rise to a very minute current, measurable only by the most sensitive instruments. Owing to these stray ions, a charged electroscope gradually loses its charge. The conductivity imparted to a space by these ions is so small as to be of little importance, except that this slight conductivity may serve to start a more powerful discharge such as a spark. Besides these stray ions, other ions may be liberated in a discharge space by X-rays from an external source, by the action of ultra-violet light upon the gas, or by collision of electrons and ions with the gas molecules. Electrons may also be supplied to the discharge space by photoelectric effect at one of the electrodes, or by thermionic emission.

The discharges through a gaseous space are conveniently divided into two general types:

1. *Non-self-sustaining discharge.*
2. *Self-sustaining discharge.*

The non-self-sustaining discharge is one which requires some external agency to supply carriers to the discharge space. Examples of this type of discharge are conduction through hot-cathode rectifiers, through photoelectric cells, etc. The discharge through high-vacuum thermionic tubes may be considered as a special type of non-self-sustaining discharge because the amount of gas is so small as to have no appreciable effect.

The self-sustaining discharge is one which maintains itself without aid from external agencies other than the source of applied potential and may or may not be self-starting. Usually this type of discharge is started by the stray ions which normally exist in a gas. In this type of discharge the carriers of electricity are replenished by the action of the discharge itself. Exam-

ples of this type of discharge are the glow, arc, and spark discharges.

1. *Non-self-sustaining Discharge.*—This type is the simpler of the two types of discharge. We may suppose, for example, that we have two electrodes in a glass enclosure containing gas at greatly reduced pressure. Suppose, first, that X-rays pass through the gas and produce, by ionization of the gas, a certain number of ions and free electrons per cubic centimeter per second. If the pressure is a few centimeters of mercury, the electrons soon find a neutral molecule to which they adhere, forming negative ions. If the pressure is much lower, most of the electrons remain free until they combine with positive ions. When an electric field is impressed between the electrodes, the positive and negative ions or electrons move in opposite directions, both contributing to the current. If the potential difference is low, the velocity of the carriers is small and considerable recombination of the electrons and ions takes place so that only a fraction of the ions produced reach the electrodes. As the potential difference is increased, the current increases up to a certain saturation value when all of the ions are swept out so rapidly that an inappreciable number are lost by recombination. The saturation value of the current depends upon the strength of the ionizing agent and is directly proportional to the volume of gas between electrodes and to the pressure of the gas.

In a discharge such as just described, the velocities of the positive and negative ions are roughly the same. When, however, the pressure of the gas is so low that free electrons exist and migrate across the space, these free electrons acquire a very large velocity compared to the velocity of the ions, because of the relatively small mass of the electron. Therefore, the current carried by the ions is generally only a few per cent of that carried by free electrons.

A discharge of this type is often used to measure the intensity of X-rays, or other ionizing radiation, by measuring the saturation current. A chamber in which the discharge takes place is known as an *ionization chamber.*

A discharge of the non-self-sustaining type, and similar in some respects to the above-described discharge, takes place if, instead of carriers being liberated throughout the gaseous space, electrons are liberated at the cathode terminal by photoelectric effect or by thermionic emission. In this case the applied poten-

tial gradient between the electrodes urges the electrons across the space and the current rises, as the potential difference increases, up to a saturation value in which case all of the electrons are swept across as fast as they are emitted.

In either of the above kinds of non-self-sustaining discharge, the electrons and ions in their passage through the gas may make frequent collisions with gas molecules if the pressure of the gas is such that the mean free path is small in comparison with the distance between electrodes. In such a case, if the maximum velocity of the electrons and ions is below ionizing velocity, only inelastic collisions, or collisions which excite resonance radiation, can take place. Such collisions cause a decrease in current only when the collisions involve a loss of energy. If, however, the velocity of the carriers is above ionizing velocity, additional ions and electrons are produced, which results in an increase in current. If the amount of gas between electrodes is small, only a limited number of additional ions can be produced, and the current, although increased, is still finite and stable and the discharge is still non-self-sustaining. If, however, there is considerable gas between electrodes, the mean free path of the electron is so small that each electron can make several ionizing collisions between electrodes, provided the potential difference is high enough to impart sufficient energy to the electrons in a distance equal to or less than the mean free path. Under these conditions cumulative ionization takes place and the current may build up to very high values and pass over to an arc or spark.

The potential difference across a non-self-sustaining discharge usually rises with increasing current, as will be explained in a later section of this chapter.

2. *Self-sustaining Discharge.*—The self-sustaining type of discharge may start as a non-self-sustaining discharge. When the potential difference between electrodes is well above ionization voltage and positive ions are produced in the gas and move toward the cathode, these ions, on striking the negative electrode, may knock out a sufficient number of electrons to maintain the discharge. No outside agency is necessary, therefore, as is necessary with the non-self-sustaining type of discharge. These electrons knocked out of the cathode move toward the anode, producing more positive ions and electrons. The positive ions thus produced move to the cathode to liberate more electrons. Thus the discharge, under favorable conditions of gas pressure,

geometry of the electrodes, and potential difference, becomes self-sustaining through mutually dependent actions of electrons and ions.

Positive ions are inefficient in liberating the necessary electrons from the cathode. Metastable atoms, *i.e.*, atoms in which one or more electrons are displaced from their normal orbits to temporary orbits, seem to be more active in liberating electrons from the cathode. Such metastable atoms are produced in the discharge and may bombard the cathode and liberate electrons. In addition, the light of short wave length emitted by the excited atoms in the discharge may cause some of the electrons to be emitted at the cathode by the photoelectric effect. The bombardment of the cathode by the positive ions may be so intense as to heat the cathode to incandescence, in which case electrons are emitted by the thermionic effect. Such an action is the principal source of electrons in an arc discharge and is the main reason that the arc is self-sustaining. If the cathode terminal of an arc is cooled, the arc ceases. Although in the arc discharge the current is large, the potential difference is relatively small, ordinarily of the order of 10 to 100 volts. The current is carried mainly by the electrons, and the larger the current the more heat is produced at the cathode and hence the more electrons are liberated. Such an action causes the potential difference necessary to maintain the current to fall as the current increases. Arcs may be maintained in a gas at atmospheric pressure; or in gas at much reduced pressure as in the case of the familiar mercury arc. It should be noted, however, that the reason for the emission of electrons at the cathode of a mercury arc is still in dispute.

Self-sustaining discharges can be made to take place at reduced pressure when there is no thermionic emission at the cathode induced by the heat of the discharge. In such cases the cathode emission is induced by the other causes already given, *i.e.*, bombardment and photoelectric effect of the light from the discharge. Such discharges usually require much higher potential differences than do arcs, *i.e.*, voltages of the order of several hundred to many thousand volts according to the pressure of the gas and the distance between electrodes. The discharges in the familiar Geissler and Crookes tubes and in the old type of gas-filled X-ray tube are examples of these high-voltage self-sustaining discharges.

In discussing this form of high-voltage self-sustaining discharge, suppose that we have a cylindrical glass tube, three or four centimeters in diameter and 50 cm. long, in the ends of which are sealed disk-shaped aluminum electrodes, the planes of the disks being perpendicular to the axis of the tube. Suppose the tube to be connected to an exhaust pump and the electrodes to a source of high potential. If the pressure of the gas is atmospheric and the potential high enough, *i.e.*, of the order of 150,000 volts, a spark may take place between the electrodes. As the pressure is reduced, the spark becomes more fuzzy and requires less voltage to cause it to pass. Soon, as the pressure becomes of the order of a few millimeters of mercury, the discharge nearly fills the whole tube with a reddish glow and the

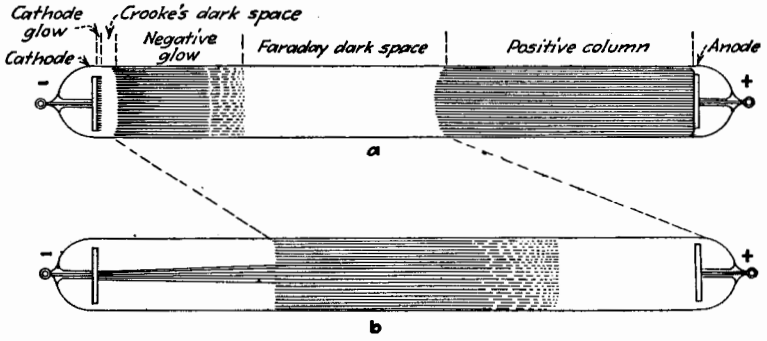


FIG. 6.—Electric discharge through a gas at reduced pressure.

potential difference required is only a few thousand volts. We may also notice that the discharge at the negative terminal has one or more nonluminous regions and exhibits a more bluish color than the rest of the discharge. To be more exact, there is a dark region close to the negative terminal, known as the *Crookes dark space*, as shown in Fig. 6*a*. Then, proceeding toward the anode, there is a bluish region, if the gas is air, sharply defined at its limit toward the cathode, known as the *negative glow*. The negative glow gradually fades into a second dark space, known as the *Faraday dark space*, diagrammatically shown in Fig. 6*a*. Finally, the rest of the discharge space is filled with a reddish glow, known as the *positive column*, which may be uniform or striated. Besides the characteristic portions of the discharge as described, there is a violet-colored glow which emanates from the cathode into the Crookes dark space

and is called the *cathode glow.* This glow at high pressures covers the surface of the cathode with a thin region of intense luminosity. As the pressure decreases, the cathode glow expands longitudinally but contracts laterally, so that at low pressures it may emanate from a small region at the center of the cathode as shown in Fig. 6*b*.

The potential gradient is very different at various parts of the discharge. Near the cathode and in the Crookes dark space it is high and greater than at any other region in the discharge. This large potential gradient is often called the *cathode fall of potential.* On the other hand, the potential gradient in the negative glow is very small but increases gradually through the Faraday dark space, attaining a nearly constant value in the positive column when it is uniform.

The negative glow is a region of ionization. The positive ions produced therein fall through the Crookes dark space where they acquire velocity sufficient to liberate electrons at the surface of the cathode. The distance from the cathode to the negative glow is largely dependent upon the mean free path of the electrons in the gas and hence upon the pressure of the gas. This distance is roughly inversely proportional to the pressure.

As the pressure of the gas is reduced, the negative glow recedes from the cathode and becomes fainter because of the lower density of the gas. The positive column moves toward the anode and appears to be swallowed by it until eventually it completely disappears. At a pressure of the order of 0.05 mm. of mercury, or 50 microns, the only glow evident in the gas is a faint negative glow (much fainter than illustrated in Fig. 6*b*) and a purplish beam which starts from the center of the cathode and extends along the axis of the tube. The beam diverges somewhat as it leaves the cathode and can be deflected by a magnetic field, showing that the luminosity marks the path of negative charges projected from the cathode in nearly straight lines. This beam is known as *cathode rays.* It consists of the electrons which are knocked out of the cathode and are then accelerated by the electric field between the electrodes. The glass at the positive or anode end of the tube usually shows a green fluorescence, if the tube is of soft or lime glass, and a blue fluorescence, if the tube is of lead glass or pyrex glass. This fluorescence is due to the bombardment of the glass by the electrons of the

cathode rays and takes place only when the gas pressure is so low that the electrons reach the glass without many collisions with gas molecules.

As the pressure of gas is further reduced, the voltage necessary to maintain the discharge increases rapidly and the negative glow becomes fainter and more remote from the cathode until it fades out entirely. The beam of luminosity emanating from the cathode becomes less visible. The fluorescence of the glass due to the bombardment by the cathode rays becomes at first more brilliant, because of reduced obstruction by gas molecules to the flight of the electrons. At lower pressures, however, it also fades because the current through the tube decreases and finally ceases altogether when the amount of gas is insufficient to allow a self-sustaining discharge to exist.

When the pressure of the gas is very low, *i.e.*, of the order of from 1 to 10 microns, the mean free path of the electron is in general much greater than the distance between electrodes. For this reason, very few positive ions are produced in a self-sustaining discharge at these pressures. In view of the inefficiency of positive ions and metastable atoms in knocking electrons out of a metal surface, it is difficult to see how these few positive ions can knock out of the cathode a sufficient number of electrons to maintain the discharge obtained in high-vacuum discharge tubes, such as the old-type gas-filled X-ray tubes. Janitsky[1] performed some interesting experiments which demonstrated that at low pressures the discharge passes much more readily to an anode which has not been degassed than to one which had been strongly heated in a vacuum to rid it of occluded gas. Janitsky interpreted his experiments to indicate that positive ions may be drawn out of the anode by the intense electric field and thus increase the stream of positive ions which bombard the cathode. He supposed that these ions drawn from the anode were atoms of gas adsorbed on the anode surface as positive ions.

The charges on the walls of the discharge tube often have a great influence upon the discharge. If the pressure of the gas is low, and consequently there are very few positive ions present in the tube, the electrons lodge on the inside surface of the tube, thus charging the walls negatively. This negative charge is sufficient to repel even high-velocity electrons, thus preventing

[1] JANITSKY, *Zeits. f. Physik*, **11**, 22 (1922); **31**, 277 (1925); **35**, 27 (1925).

any bombardment of the walls. When there is no bombardment of the walls by electrons, there is no glass fluorescence. This is the case in the modern Coolidge X-ray tube. The negative charge on the walls may greatly inhibit the discharge through a tube, especially if the tube is long and is small in diameter.

Occasionally, however, this negative charge may reverse to a positive charge at certain spots on the glass. The reason for this reversal is as follows: If, before any negative charge has become established on the glass wall, a few high-velocity electrons strike the glass in a spot, secondary emission from this spot may then take place to such an extent that more electrons leave the spot than reach it. This spot then becomes positively charged and attracts more high-velocity electrons to it. Thus a stable condition is set up, in which the high-velocity electrons continue to bombard the glass, and secondary emission maintains the positive charge. The bombardment of the glass by these high-velocity electrons produces a great deal of heat which can easily result in melting a hole in the glass tube.

When the gas pressure is of the order of a micron or more, there are usually a sufficient number of positive ions present to neutralize the negative charge on the walls. Although the electrons may bombard the walls and lodge on the glass, the positive ions diffuse to the walls and combine with the electrons. In fact, a great deal of the recombination takes place at the walls of the tube. The continual bombardment of the walls, made possible by the continual neutralization of the electrons, produces the glass fluorescence already described.

22. Conduction in a High Vacuum.—The type of conduction now to be considered briefly is really a special case of the non-self-sustaining discharge described in the previous section as taking place when the amount of gas is so small as to have a negligible effect upon the characteristics of the discharge. Since most of the vacuum tubes to be considered in this book are high-vacuum devices, it seems pertinent to include a section at this point on this type of conduction, although the characteristics of the flow of electricity through a high vacuum are considered in detail in Chaps. IV, V, and VII.

For this type of discharge, electrons must be supplied at the cathode by the photoelectric effect, by thermionic emission, or by an equivalent means. The electric field impressed between the electrodes accelerates the electrons so that, as they travel

toward the anode, their velocity continually increases. Since there is no appreciable amount of gas, collisions of the electrons with gas molecules play no part in the conduction. There is no visible glow in the tube, and the walls, if bombarded by electrons, become highly charged negatively unless secondary emission is started, in which case the sign of the charge on the glass may reverse, as described previously.

When the electrons reach the anode, they are traveling with a velocity dependent upon the potential difference between electrodes. The numerical value of this velocity can be obtained from Eq. (4), or approximately from Eq. (5), page 27. The kinetic energy imparted to the electron by the field is given up to the anode, when the electron strikes it, and appears as heat. Very high temperatures can be produced in this way and may be sufficiently high to melt any metal.

Attention is again directed to the fact that the electrons move in a direction opposite to the conventional direction of the electric current. The same number of electrons pass every cross section in the tube per second so that, where the velocity of the electrons is large, their volume density in space is correspondingly small.

23. Cathode Rays.—Those electrons which emanate from the cathode and which normally travel in straight lines through the tube when the gas pressure is low constitute the cathode rays. The term is perhaps more descriptive of the phenomenon when the electrons pass through a narrow orifice, thus limiting the stream of electrons to a narrow beam.

The cathode rays cause many substances struck by them to fluoresce. The fluorescence of the various kinds of glass has already been described. Many gem crystals, such as ruby and diamond, are caused to fluoresce brilliantly by impact of cathode rays. Many salts, notably calcium tungstate, zinc silicate or willemite, platinobarium cyanide, etc., also exhibit very brilliant fluorescence under the action of these fast-moving electrons. Screens made by covering a flat surface with a thin layer of one of these salts are often used to show the place where cathode rays strike.

The *Braun tube*, a most useful device, consists essentially of a long glass tube having at one end a suitable cathode and limiting orifice which gives a narrow beam of cathode rays along the axis of the tube. A screen of fluorescent salt is placed at the other

end of the tube to show the point of impact of the electrons. Since the beam of electrons is equivalent to a current, it can be deflected by either a magnetic or electric field, and the amount of deflection is shown by the motion of the fluorescent spot on the screen. The Braun tube will depict very rapid changes in the deflecting fields because the cathode beam, being without inertia, is practically instantaneous in its response. For this reason the Braun tube is a most valuable oscillograph for high-frequency work.

The point of impact of high-speed cathode rays is the source of X-rays. In the X-ray tube, the cathode rays are focused by the electrostatic field about the properly designed cathode so that they strike the piece of metal known as the *target* in a small area from which the X-rays emanate. The X-ray tube is not within the scope of this book, so no further discussion of it will be given.

When a small amount of gas, insufficient to support a self-sustaining discharge, is present in a discharge tube in which a narrow beam of cathode rays is produced, enough ionization takes place along the path of the rays to render the path visible as a purplish-blue streak. Very instructive and beautiful experiments can be performed with such a tube by deflecting the beam by magnetic fields, or by electrostatic fields produced by charged electrodes within the tube.[2]

High-speed cathode rays can be projected through a thin film of metal, such as aluminum or molybdenum, used as a window of the tube, so that the electrons are obtained outside the tube. Their velocity is rapidly reduced by collisions in the dense gas of the outer space. Lenard[3] first projected cathode rays outside a tube, but his rays had only a small velocity. These rays outside the tube were called *Lenard rays*. More recently Coolidge[4] has projected 750,000-volt electrons outside the tube. A blue haze outside and near the window of Coolidge's tube marks the region of intense ionization. Crystals placed in this region fluoresce brilliantly owing to bombardment by the electrons.

24. Canal or Positive Rays.—Not only is it possible to produce a beam of rapidly moving electrons, but it is also possible

[2] BRÜCKE, *Phys. Zeits.*, **31,** 1011 (1930).

[3] LENARD, *Ann. d. Physik,* **51,** 229 (1894).

[4] COOLIDGE, *J. Franklin Inst.*, **202,** 693, 722 (1926); *Gen. Elec. Rev.*, **31,** 184 (1928).

to produce a beam of rapidly moving positive ions. A beam of positive ions was first recognized by Goldstein[5] in 1898 and called by him *canal rays*. Canal rays can conveniently be observed if a perforated cathode is used in a tube carrying a self-sustaining discharge at low pressure and high voltage. Some of the positive ions which move toward the cathode pass through the holes into the region back of the cathode and their path in the tube is visible, owing to the ionization of the gas. When these positive ions fall upon glass or some other materials, they produce fluorescence, though the color of the fluorescence is different in general from that produced by cathode rays.

The positive rays can be deflected by electric and magnetic fields, but because of the much greater mass of the positive ions, the amount of deflection is very much less than in the case of cathode rays.

25. Positive-ion Current.—It should be noted that, although conduction through a gas comprises a transfer of electricity by both positive and negative carriers, the two kinds of carrier do not necessarily contribute equally to the current.

The mass of the electron is only about 1/1,840 that of an atom of hydrogen, so that in hydrogen the positive-ion carrier has a mass about 1,840 times greater than the mass of the negative carrier having the same magnitude of charge. In other gases the ratio of mass of the positive and negative carriers is much greater. Making use of the simple laws of accelerated motion, if the mean free path is large compared with the distance between the electrodes, the velocities acquired by the positive and negative carriers in traversing the distance between electrodes are to each other as the reciprocals of the square roots of their masses. In the case of hydrogen, the ratio of velocities of the positive and negative carriers is about $1/\sqrt{1,840} = 1/42.9$. Since the amount of electricity transferred in a given time by carriers having the same charge is proportional to their velocities, the current carried by the positive ions in hydrogen is about $\frac{1}{42}$ of the current carried by the negative electrons. In other gases having positive ions of greater mass, the ratio of currents is less. In other words, the positive-ion current is seldom greater than $\frac{1}{43}$ of the negative-electron current and is usually much less.

26. The Current-voltage Characteristics of Conductors.—In the preceding sections, the various physical aspects of the several

[5] GOLDSTEIN, *Ann. d. Physik*, **64**, 38 (1898).

types of conduction have been considered. We shall now discuss conduction quantitatively in terms of the potential difference across the conductor and the current which flows through the conductor.

For many conductors the current is a function only of the impressed potential difference. This may be expressed in mathematical form by

$$i = f(e) \tag{13}$$

where i and e represent the instantaneous current and the difference of potential. In many cases, however, relation (13), although very general in its form, is inadequate to express the characteristic of the conductor. The current may depend upon the length of time the potential e is applied. In such a case the additional factor, time, must be included, and

$$i = f(e, t) \tag{14}$$

First, consider those conductors whose characteristics are expressed by relation (13). A metal maintained at constant temperature has so many free electrons that, for ordinary values of current, the current is directly proportional to the potential difference. This is expressed mathematically by the following simple form of Eq. (13):

$$i = Ke \tag{15}$$

where K is a constant known as the *conductance* of the particular piece of metal. The reciprocal of K is the *resistance*, R, so

$$e = Ri \tag{16}$$

Equations (15) and (16) are known as Ohm's law and the resistance of the conductor is known as an *ohmic resistance.*

Many types of conduction, however, for example, gaseous conduction as just described, do not follow Ohm's law but have more complicated forms of the relation (13). In general, the relation (13) cannot be expressed in mathematical form, and it is necessary to express the relation graphically by what is known as an *e-i characteristic curve.* Several such curves are given in Fig. 7, the type of conduction to which each curve applies being noted on the curve. The scales of voltage and current are not necessarily the same for the several curves.

Figure 7*a* is the graph of an ohmic resistance, as given by Eq. (15). The conductance K is represented by the slope of the resistance line.

In none of the other types of conduction illustrated in Fig. 7 is there a constant such as K or R, but instead, the ratio of the instantaneous current to the corresponding instantaneous voltage, *i.e.*, the instantaneous conductance, is a variable and a function of the current or potential difference. The arc, illustrated by the curve of Fig. 7*b*, exhibits a diminishing voltage as the current increases, which gives an increasing conductance with increasing current. Such a characteristic is often termed a *falling characteristic* because its slope is always negative. The curve shown in Fig. 7*c* illustrates, in general form, the course of the *e-i* curve for gaseous conduction. The exact shape of this curve depends

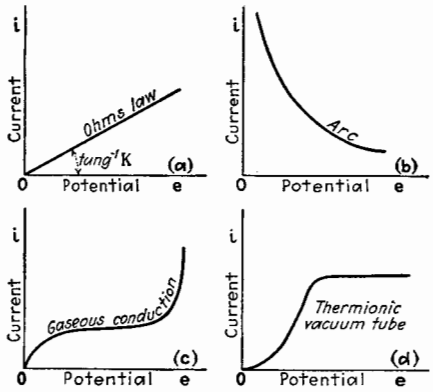


FIG. 7.—Forms of *e-i* characteristic curves.

upon various factors, such as gas pressure and kind of gas, and the final rapid rise of current may be entirely absent. The flat region is the saturation value of current. It occurs when all of the carriers are swept out as rapidly as they are produced. The rapid rise at the end is due to cumulative ionization. The curve of Fig. 7*d* is typical of the electronic conduction in a high-vacuum thermionic tube. The flat portion shows the saturation value of the current when all of the electrons are drawn to the positive electrode as rapidly as they are emitted by the heated cathode. The curves of Fig. 7, *c* and *d*, are rising characteristics because the slope of the curves is always positive. The instantaneous conductance in certain regions may increase with increasing current, while in other regions it may decrease with increasing current.

A common example of a type of conduction in which time enters is conduction through the filament of a tungsten lamp. Owing to the rise of temperature with increasing current, the conductance of the filament decreases by a factor of about ten from zero to normal current. The *e-i* characteristic for slowly changing current is shown by the curve so marked in Fig. 8. If, however, the increase and decrease of the current take place in so short a time that the temperature of the filament does not change appreciably, the *e-i* characteristic is given by the straight line marked "very rapid current change." If the current rises fairly rapidly and then decreases more slowly, some such

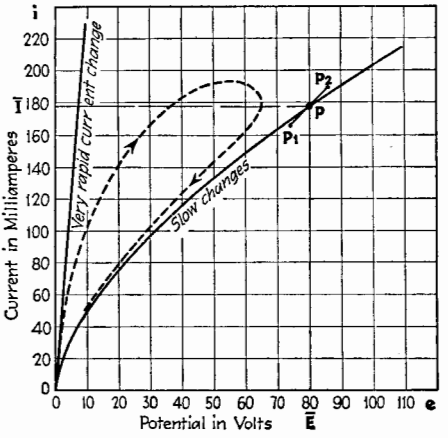


FIG. 8.—Current-voltage characteristic curve of a tungsten-lamp filament.

characteristic as is shown by the dotted line may be obtained. It is evident that for such a conductor there is no unique *e-i* characteristic, except for the limiting cases of very slow change and very rapid change of current. In the first case the temperature follows the change in current. In the second case the changes of current are so rapid that the temperature cannot follow them and assumes some mean value.

Actually, the conduction through a gas, illustrated in Fig. 7, *b* and *c*, is of the type in which the current-voltage curve depends upon the rate of change of the current. Time enters not only because the temperature changes as the current changes but also because the positive ions, which influence the conductance, move much more slowly than the electrons do and hence cause a lag in the effects when the current changes at a high rate. If the

rate of current change is increased to such values that the time of flight of the electron between electrodes becomes appreciable compared to the period of change of the current, even conduction through a high-vacuum thermionic tube is of the type involving time as expressed in Eq. (14). The conduction through a metal at constant temperature is also of the type containing time, for very rapid current changes, so that all conduction in the final analysis is of the type shown in Eq. (14). For some types of conduction the time factor can be neglected for very slow changes of current only, while for other types of conduction the time factor is negligible for fairly rapid current changes.

27. Variational or Incremental Conductance and Resistance.—Whenever an ohmic resistance of the type illustrated by Fig. 7*a* is an element of a circuit, which may also contain inductance and capacitance, it is a simple matter to calculate the current which flows in the circuit under a given impressed electromotive force. The resistance is the same whether the applied e.m.f. is constant or alternating, or if an alternating e.m.f. is superimposed on a constant e.m.f. When the type of conduction does not follow Ohm's law, complexities arise which will now be considered.

A type of conduction which does not follow Ohm's law is shown in Fig. 7*b*, also given in Fig. 9. If the instantaneous e.m.f. is e_1 (Fig. 9), the instantaneous current is i_1 and, by analogy with Ohm's law, the *instantaneous conductance* is the ratio of $i_1/e_1 = \bar{k}_1$. If e_1 changes to another value e_2, the instantaneous conductance changes to a new value $\bar{k}_2$. If $e_2 - e_1$, a small change in the e.m.f., is Δe, and the corresponding change $i_2 - i_1$, in i, is Δi, then

$$\text{limit}_{\Delta e=0} \frac{\Delta i}{\Delta e} = \frac{di}{de}$$

is called the *variational or incremental conductance* k. This new quantity is the slope of the characteristic curve and, in the case shown in Fig. 9, is a negative quantity. The quantity k is the conductance offered to a small variation of e.m.f. as, for example, for a small alternating potential superimposed upon a steady e.m.f.

Since the reciprocal of the conductance is the resistance, de/di is the *variational resistance*. In the work that follows, conductance is used instead of resistance, when simpler expressions result therefrom.

Referring again to the example given in Fig. 9, let P be the center point about which a small alternating variation of potential and current takes place. The voltage and current corresponding to P are $\bar{E}$ and $\bar{I}$. $\bar{E}$ is a steady potential upon which is

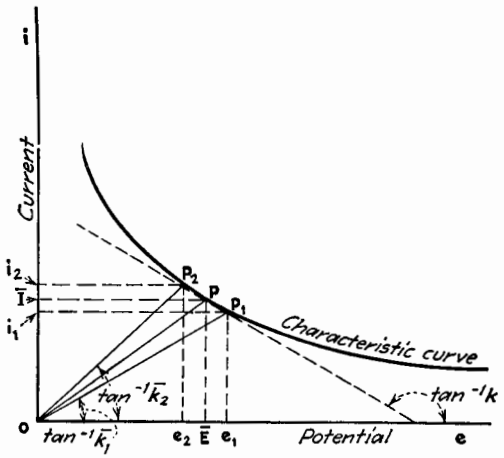


FIG. 9.—Characteristic curve of an arc.

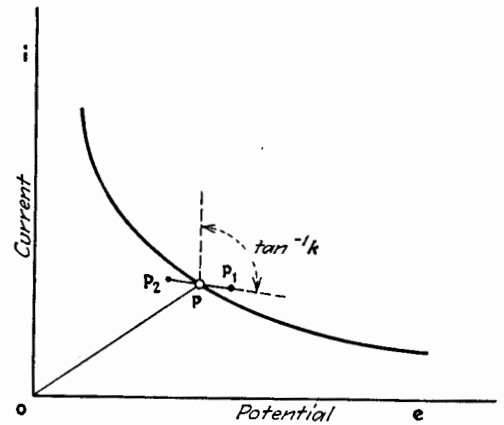


FIG. 10.—Path of operation for a small high-frequency alternating variation.

superimposed a small alternating potential of amplitude $e_1 - \bar{E}$. To the steady current $\bar{I}$, the arc offers a positive resistance given by $\bar{E}/\bar{I}$ or a positive conductance given by the slope of the line joining O with P. To the alternating variation, the arc offers a negative resistance or conductance.

In the example of Fig. 9 just described, the conduction follows the simple law given in Eq. (13), in which case the variational conductance is simply the slope of the *e-i* characteristic curve. If, however, the alternating variation is so rapid that the time factor enters, the variational conductance is not the slope of the characteristic curve but of some other line P_1P_2, as shown in Fig. 10.

A condition similar to that described for the arc obtains if a small alternating e.m.f. is superimposed upon a steady e.m.f. $\bar{E}$, applied to a metal-filament lamp. This case is illustrated by the a-c. line P_1P_2 in Fig. 8. The conductance for a small alternating current of high frequency is constant and of a value determined by the steady current $\bar{I}$.

Sometimes the condition is even more complex than is illustrated in Fig. 10. The a-c. path P_1P_2 for increasing current may not be the same as the path for decreasing current and a small ellipse or other figure may be traced between P_1 and P_2. In such a case the conduction not only offers resistance but acts as though it possesses inductance also.

If the a-c. variations are small, all of these cases can be handled mathematically by solving the steady-current and a-c. parts of the problem separately. As already stated, the value of resistance or conductance for steady current is different, in general, from the value for alternating current. The method of solution of this type of problems will be more fully described in a later chapter.

General References

TOWNSEND: "Electricity in Gases," Oxford University Press, New York, 1915.

LANGMUIR and MOTT-SMITH: *Gen. Elec. Rev.*, **27** (1924).

DARROW: "Introduction to Contemporary Physics," D. Van Nostrand Company, New York, 1926.

"Handbuch der Physik," vol. XIV, Geiger-Scheele, Julius Springer, Berlin, 1927.

"Handbuch der Experimental Physik," vol. XIII, Teil 3, Wien and Harms, Akademische Verlagsgesellschaft M. B. H., Leipzig, 1928.

THOMSON, J. J.: "Conduction of Electricity through Gases," 3d ed., Cambridge University Press, Cambridge, 1928.

EMELÉUS: "The Conduction of Electricity through Gases," E. P. Dutton & Co., Inc., New York, 1929.

COMPTON and LANGMUIR: Electrical Discharges in Gases, Part I, *Rev. Modern Phys.*, **2**, No. 3 (1930).

LANGMUIR and COMPTON: Electrical Discharges in Gases, Part II, *Rev. Modern Phys.*, **3**, No. 2 (1931).

CHAPTER IV

EMISSION OF ELECTRONS

28. Three Sources of Electron Emission.—By the emission of electrons is meant the evolution or giving off of electrons by solid or liquid bodies, usually not spontaneously but as an accompaniment and result of the action of certain definite physical agencies. There are three important means by which electrons are caused to be emitted:

a. Elevation of temperature of body.

b. Bombardment of body by rapidly moving ions, electrons, or metastable atoms.

c. Electromagnetic radiation of sufficiently short wave length falling upon body.

Electrons are spontaneously emitted from radioactive substances, but this form of emission is so weak as to be of little importance as a source of electrons for electron tubes and will not be considered further.

The three types of emission classified above according to the causes of emission are called *thermionic emission*, *secondary emission*, and *photoelectric emission*, respectively. The first of these types of emission is the most important and will be considered in detail.

I. THERMIONIC EMISSION

29. Maxwell's Distribution of Velocities.—The heat energy that a body contains, which determines its temperature, resides in the kinetic energy of motion of its atoms or molecules, which are in constant random to-and-fro motion, as briefly explained in Chap. II. The greater the heat energy, the higher the temperature and the higher the velocities of motion of the particles. The velocities of motion are not the same for all the particles. Some particles have a relatively low velocity while others have a very high velocity, but the majority have velocities which are not very different from a certain velocity known as the *most probable velocity*. Maxwell calculated, by the theory of probability, the distribution of velocities of the

particles of a gas and obtained the curve shown in Fig. 11. In this figure, any ordinate y represents the probability that a particle has a velocity x times the most probable velocity, where x is the abscissa of the ordinate y. The ordinate y is a function of x given by the relation

$$y = \frac{4}{\sqrt{\pi}} x^2 \epsilon^{-x^2} \tag{17}$$

The ratio of the area under the curve between any two ordinates, at x_1 and x_2, to the total area under the curve, gives the fraction of the total number of particles which have velocities ranging

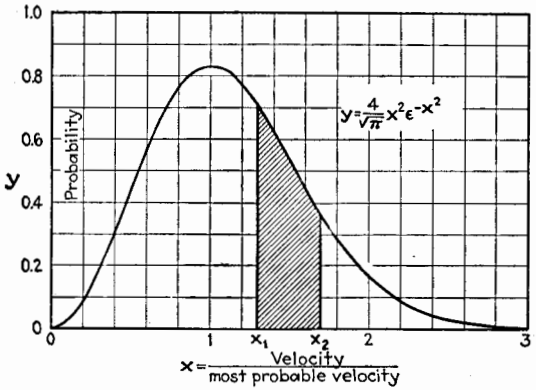


FIG. 11.—Maxwell's law of distribution of velocities of gas molecules.

between x_1 and x_2 times the most probable velocity. For example, only 3.1 per cent of the particles have velocities greater than 2.5 times the most probable velocity. If the ordinates are Δx apart, then $y\Delta x/A = \Delta n_x/N$, where Δn_x is the number of particles having velocities ranging between x and $x + \Delta x$ times the most probable velocity, N is the total number of particles, and A is the total area under the curve. $A = \int_0^\infty y dx = 1$.

The most probable velocity changes with the temperature and is proportional to the square root of the absolute temperature.

30. Work Function.—The free electrons within a conducting body are assumed also to have random motion somewhat similar in kind to the motion of the atoms. The classical theory, developed by Richardson, postulated that the electrons share with the atoms the heat energy and hence that the velocities of the

electrons follow Maxwell's distribution law. Modern theories, based on the statistical mechanics of Fermi and Dirac, and developed by A. Sommerfeld and others, give to the electrons a velocity distribution which is very different from that of Maxwell.

It is sufficient for our purpose at this point, without going into a discussion of the several theories, to picture the electrons within the conductor as having some sort of velocity distribution. If there were no restraint at the surface, all electrons meeting the boundary of the conductor and having a normal component of velocity would escape, thus giving enormous emission of electrons even at ordinary temperatures. It is, therefore, necessary, in order to explain experimental results in terms of this picture,* to postulate a surface restraint or barrier which only those electrons that have a normal component of velocity greater than a certain value can surmount. This surface restraint may be due mostly to the electrical attraction of the surface for the emerging electron but is most conveniently expressed in terms of a definite amount of work w. An electron must possess kinetic energy at least as great as w in order to pass out through the surface. This work w is known as the *thermionic work function.*

The following simple analogy may aid in the understanding of the physical meaning of the work function. Imagine a number of marbles contained in a box, the sides of which have a height h. Suppose the marbles are jostled about by some means such as a hopper at the bottom of the box. The random motion of the marbles is roughly analogous to the motion of electrons. Some of the marbles may possess sufficient kinetic energy to carry them over the side of the box and out. If m is the mass of a marble, the minimum kinetic energy necessary is $\frac{1}{2}mv_{\text{min.}}^2 = mgh$, where mgh is analogous to the thermionic work function w. A marble which is projected free from the box has lost no energy; the kinetic energy which it possessed has been transformed wholly or partly into potential energy when it is elevated to a height h above its normal position.

Returning now to the electrical case, only those electrons can escape whose velocities are equal to or greater than $v_{\text{min.}}$ given by the equation

* The picture described above and in the following sections should be considered in some respects as a very crude picture. It is difficult to describe the quantum mechanics of the electrons inside a conductor in terms of an accurate and at the same time simple picture.

$$\tfrac{1}{2}mv_{\text{min.}}^2 = w \tag{18}$$

The energy of an escaped electron is not lost but merely transformed in part into potential energy.

The work w is ordinarily expressed in terms of a potential Φ such that Φ times the electronic charge e is equal to the work w, or

$$\Phi e = w \text{ (in absolute units)}$$

If Φ is in volts and e in absolute electrostatic units,

$$\Phi e = 300w \tag{19}$$

The equivalent voltage Φ is often called the *electron affinity* and is different for different substances. The quantity Φ is of the greatest importance because it indicates the comparative difficulty that electrons experience in escaping from different substances. Usually, the substances having small electron affinities are better emitters of electrons at any given temperature than the substances having larger values of Φ. This important constant is treated more in detail in Sec. 33 of this chapter.

31. Mechanism of Emission.—As the temperature of a conductor is increased, the distribution of velocities of the free electrons inside the conductor changes in such a way that more electrons possess a velocity sufficient to carry them through the surface restraint. The electrons which are emitted charge the space outside negatively and leave the body positively charged. Hence, there exists an electrostatic field outside the body urging the emitted electrons back into the body. At any given temperature, equilibrium is established when just as many electrons return to the body as escape in any given interval of time. Thus, a cloud of electrons exists outside the body, having a density dependent upon the temperature and upon the distance from the surface. There is, however, a definite rate of emission of electrons at each temperature, probably independent of the density of the electron cloud outside the body. If an external electric field is applied which is sufficient to draw off the electrons as fast as they are emitted, a certain saturation current per unit area of surface of the emitting body is obtained at each temperature.

32. Emission vs. Temperature.—In 1901, O. W. Richardson,[1] applying the classical kinetic theory, deduced an equation for

[1] RICHARDSON, *Camb. Phil. Proc.*, **11**, 286 (1901).

the saturation emission current per unit area of emitting surface as a function of temperature. This equation is

$$I_s = A_1 T^{1/2} \epsilon^{-\frac{w}{\kappa T}} \tag{20}$$

where A_1 is a constant; w is the work per electron necessary for it to pass through the surface restraint and is assumed to be a constant for any substance; T is the absolute temperature; ϵ is the Napierian base; and κ is Boltzmann's gas constant for a single electron, or two-thirds the average kinetic energy possessed by an electron at 1° abs. ($k = 1.3709 \times 10^{-16}$ erg per degree). In deducing this equation, Richardson assumed that the electrons inside the conductor obey the laws of a perfect gas and share with the atoms the heat energy in the body. The velocities of the electrons inside the body were therefore assumed to follow Maxwell's distribution law.

Richardson's theory just referred to is now held to be incorrect because of discrepancies between the theory and certain experimental facts. For example, if the large number of free electrons which exist in a conductor share in the heat energy, the specific heat of the body should be much larger than is actually observed.

Later, Richardson,[2] using the reasoning of thermodynamics applied to the electrons outside the conductor and making no assumptions as to the velocity distribution of the electrons inside the conductor, arrived at the expression given in Eq. (21)

$$I_s = A T^{1/2} \epsilon^{\int \frac{w}{\kappa T^2} dT} \tag{21}$$

where A is a quantity independent of T. Richardson pointed out that if the work function w is a function of temperature, given by the approximate relation

$$w = w_0 + \tfrac{3}{2}\kappa T, \tag{22}$$

which is deduced by further thermodynamical reasoning concerning the electrons inside the metal, Eq. (21) reduces to

$$I_s = A T^2 \epsilon^{-\frac{w_0}{\kappa T}} = A T^2 \epsilon^{-\frac{b_0}{T}} \tag{23}$$

where w_0 is the value of the work function at absolute zero.

Equation (23) can be deduced from the quantum theory, using the statistical mechanics of Fermi and Dirac, as shown by

[2] RICHARDSON, *Phil. Mag.*, **28**, 633 (1914); "Emission of Electricity from Hot Bodies," Longmans, Green & Co., New York, 1916, rev. 1921.

Sommerfeld.[3] Equation (23), therefore, has a much stronger theoretical backing and is now universally adopted as the correct form of the emission equation.

Equation (23) has another point in its favor. As pointed out first by Richardson,[4] A comes out experimentally to have about the same value for many common metals. Later Richardson[4] and then Dushman[5] showed theoretically that A is a universal constant for pure metals.

Bridgman,[6] in a rigorous theoretical treatment of the emission equation, shows that, if the difference between the specific heat of the neutral metal and the specific heat of the electric charge on its surface is zero, A has the universal constant value given by Dushman's theory. There is, however, some slight evidence that A may be a function of Φ_0, where Φ_0 is the value of Φ at absolute zero, and possibly also of the reflection coefficient of electrons at the anode.[7]

It might seem that it would be easy to decide by experiment between the two possible expressions for the emission current given in Eqs. (20) and (23), but the exponential factor varies so much more rapidly with T than does the T^2 or $T^{1/2}$ term that it is practically impossible to decide from the experimental results which equation is the correct one. However, the experimental results of Dushman,[5] Schlichter,[8] Davisson and Germer,[9] Waterman,[10] and Dushman *et al.*,[11] are in support of Eq. (23).

Dushman, Rowe, Ewald, and Kidner[11] have found by experiment that the value of A for tungsten, tantalum, molybdenum, and thorium agrees well with the theoretical value of $60.2 \frac{\text{amp.}}{\text{cm.}^2 \text{ deg.}^2}$. The value of A is vastly different from this if there is a film of another element on the surface of the metal. A monatomic layer of an element more electropositive than the metal, such as a layer of caesium or thorium on tungsten,

[3] SOMMERFELD, *Zeits. f. Physik*, **47**, 1 (1928).

[4] RICHARDSON, *Proc. Roy. Soc.* (London), **A 91**, 530 (1915); "Emission of Electricity from Hot Bodies," p. 42, Longmans, Green & Co., New York, 1916.

[5] DUSHMAN, *Phys. Rev.* **21**, 623 (1923).

[6] BRIDGMAN, *Phys. Rev.*, **14**, 306 (1919); **27**, 173 (1926).

[7] COMPTON and LANGMUIR, *Rev. Modern Phys.*, **2**, 137 (1930).

[8] SCHLICHTER, *Ann. d. Physik*, **47**, 625 (1915).

[9] DAVISSON and GERMER, *Phys. Rev.*, **20**, 300 (1922).

[10] WATERMAN, *Phys. Rev.*, **24**, 366 (1924).

[11] DUSHMAN, ROWE, EWALD, KIDNER, *Phys. Rev.*, **25**, 338 (1925).

decreases the value of A, while a monatomic layer of a more electronegative element, such as oxygen or phosphorus, increases the value of A.

Equation (23) is shown plotted in Fig. 12 for tungsten and logarithmically in Fig. 13 for tungsten, tantalum, molybdenum, platinum, and thorium. With the exception of platinum, the differences in emission for the several metals are due to the differ-

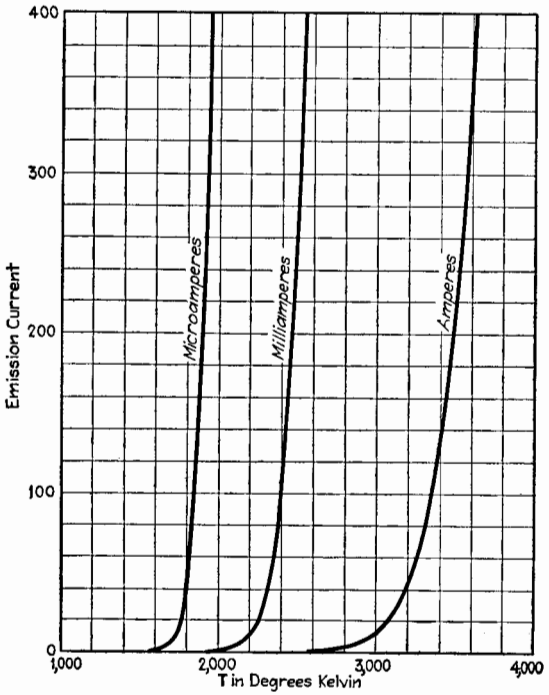


FIG. 12.—Saturation current for tungsten. $E = \infty$.

ent values of b_0 in Eq. (23). (The values of b_0 for these metals are given below in Table II.) The melting points of the several pure metals are indicated by the termini of the lines in Fig. 13. It is evident from the figure that at any temperature below about 2500°K. the emission from tantalum is at least ten times the emission from tungsten. The measurements of the emission constants of platinum vary widely, probably because of the impossibility of obtaining a clean surface uncontaminated with adsorbed oxygen.

It should be remembered that the saturation current given by Eq. (23) is obtained only when an electric field is impressed of sufficient intensity to draw all electrons away from the emitter as fast as they emerge. This condition prevailed in obtaining the results of Figs. 12 and 13. The manner in which the emission current varies with temperature, when the impressed voltage is small, is considered in Sec. 42 of this chapter.

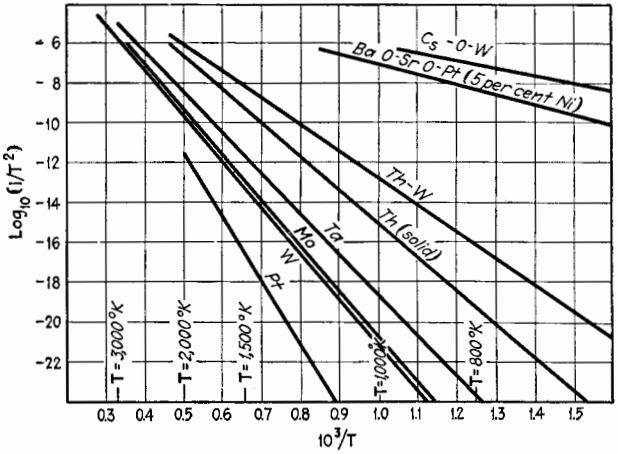


FIG. 13.—Comparison of emission from various kinds of surfaces.

33. Further Considerations Concerning the Work Function.—We return to further consideration of the work function w_0 or the electron affinity Φ_0. The value of w_0 for tungsten is approximately 7.20×10^{-12} erg, or the equivalent voltage Φ_0 is 4.52 volts. The value of w or Φ at room temperature is only slightly greater than the value of w_0 or Φ_0 as shown by Eq. (22).

If it is assumed in accordance with the classical theory that the thermal energy of the electrons inside the metal is the same as that of the electrons just outside, w is the work done against electrical forces when an electron leaves the surface. This statement and the conclusions that follow are not materially altered when the modern theory of emission is used.

Schottky[12] has shown that if certain rather rough assumptions are made, the calculation of this electrical work comes out to be

[12] SCHOTTKY, *Phys. Zeits.*, **12**, 872 (1914); **20**, 220 (1919); *Ann. d. Physik*, **44**, 1011 (1914); *Zeits. f. Physik*, **14**, 63 (1923).

about equal to the observed value for the work function. Analytically, the expression for the electrical work is

$$\text{Work} = \int_0^\infty F(x)dx \tag{24}$$

where $F(x)$ is the force on an electron distant x from the surface. Schottky assumed that $F(x)$ is the ordinary image force $e^2/4x^2$ for values of x which are large compared with the diameter of a surface atom. When the electron is near the surface, the atoms being so large in comparison with the electron, the image theory is obviously incorrect and some other form for $F(x)$ must be used. He assumed that for small values of x, the force is constant and equal to $e^2/4x^2$ up to a critical distance x_0. Langmuir[13] assumed a parabolic form for the force near the surface, starting with zero at the surface and merging into the image force at a certain small distance from the surface. Experiment cannot determine the form of $F(x)$ for small distances, but it does indicate that the force is the image force for distances greater than a few atom diameters, unless the surface is very rough, when the image law is modified.

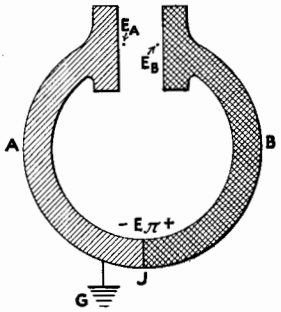


FIG. 14.

34. Contact Potential and Its Relation to the Work Function.—The work w has a very intimate connection with the *contact e.m.f.* between metals. This can easily be seen in the following way. In Fig. 14, let A and B be two metals maintained at the same temperature. Suppose that metal A is grounded. An electron which leaves metal A, arriving at a point a short distance outside its surface, has overcome, by virtue of its initial kinetic energy, practically all of the surface restraint, and an amount of kinetic energy $w = \Phi e$ has been converted into potential energy. The electron is then at a negative potential E_A with respect to ground potential, so that

$$E_A = -\Phi_A = \frac{-w_A}{e} \tag{25}$$

Similarly, an electron emerging from B does an amount of work w_B, and at a short distance from the surface of B is at a negative potential E_B with respect to the metal. The metal B

[13] LANGMUIR, *Trans. Am. Electrochem. Soc.*, **29**, 157 (1916).

is at a different potential from A because of a small Peltier *e.m.f.* E_π at junction J, so that

$$E_B = -\Phi_B \pm E_\pi = -\frac{w_B}{e} \pm E_\pi \tag{26}$$

The difference between Eqs. (25) and (26) gives

$$(E_B - E_A) = (\Phi_A - \Phi_B) \pm E_\pi \tag{27}$$

The quantity $E_B - E_A$ is the difference of potential which would exist between the two metals B and A if their surfaces were brought near together and they were not metallically connected, and is known as the *contact difference of potential*, or the *Volta potential*. Equation (27) shows that this Volta potential difference is practically equal to the difference in electron affinities, E_π usually being negligible in comparison with this difference.

When an electron escapes through the surface of an emitter, some of its kinetic energy, in amount equal to w, is converted into potential energy. The electron takes away this amount of energy, causing the emitter to be cooled. Conversely, an electron starting from rest just outside a substance, and passing inward through the same surface restraint, causes a heating equal to w ergs per electron. This heating must not be confused with the much greater heat due to the bombardment of an electrode by rapidly moving electrons. The heat of bombardment is due to the conversion of all of the kinetic energy of the electrons into heat.

35. Determination of the Work Function and Constant A.[14–32]—Most of the experimental determinations of the work

[14] DUSHMAN, *Rev. Modern Phys.*, **2**, 381 (1930).
[15] WEHNELT and JENTZSCH, *Ann. d. Physik*, **28**, 537 (1909).
[16] SCHNEIDER, *Ann. d. Physik*, **37**, 569 (1912).
[17] COOKE and RICHARDSON, *Phil. Mag.*, **25**, 624 (1913); **26**, 472 (1913).
[18] LESTER, *Phil. Mag.*, **31**, 197 (1916).
[19] WILSON, *Proc. Nat. Acad. Sci.*, **3**, 426 (1917); *Phys. Rev.*, **10**, 79 (1917).
[20] DAVISSON and GERMER, *Phys. Rev.*, **20**, 300 (1922); **24**, 666 (1924).
[21] MICHEL and SPANNER, *Zeits. f. Physik*, **35**, 395 (1925).
[22] VAN VOORHIS, *Phys. Rev.*, **30**, 318 (1928).
[23] COMPTON and VAN VOORHIS, *Proc. Nat. Acad. Sci.*, **13**, 336 (1927).
[24] HORTON, *Phil. Trans.*, A **207**, 149 (1907).
[25] LANGMUIR, *Phys. Rev.*, **2**, 450 (1913); *Phys. Zeits.*, **15**, 516 (1914); *Trans. Am. Electrochem. Soc.*, **29**, 125 (1916).
[26] SCHLICHTER, *Ann. d. Physik*, **47**, 573 (1915).
[27] KINGDON and LANGMUIR, *Phys. Rev.*, **22**, 148 (1923).

function or electron affinity Φ_0 have been made by one of two methods. The first method[15–21] involves the measurement of the cooling of the cathode due to emission, or the heating of the anode due to electrons entering its surface. The second method[20,24–32] is the direct determination of the constants of the emission equation, Eq. (23), usually by. plotting, as in Fig. 13. When the second method is used, the constant A can also be determined.

TABLE II.—EMISSION CONSTANTS[1]

Metal	Φ_0, volts	b_0, degrees Kelvin	A $\frac{\text{amp.}}{\text{cm.}^2 \text{ deg.}^2}$
Ni	2.77	32,100	26.8
C	4.00	46,500	60.2
W	**4.52**	**52,400**	**60.2**
Mo	**4.42**	**51,300**	**60.2**
Ta	**4.07**	**47,200**	**60.2**
Zr on W	3.15	36,500	5.0
Cs	1.81	21,000	16.2
Th	3.35	38,900	60.2
Th on W	2.63	30,500	3.0
U on W	2.84	33,000	3.2
Ca	2.24	26,000	60.2

$\frac{e}{\kappa}$ = 11,606 deg. per volt.

[1] International Critical Tables, vol. VI, McGraw-Hill Book Company, Inc., New York, 1929.

While most of the reliable values of the work function have been determined by the methods just described, the work function can be found also from the long-wave-length limit of the photoelectric effect, or by measuring directly the contact e.m.f. between two metals, the electron affinity of one of the metals being known. These last methods in general do not give results so accurate as the first two methods, because of the difficulty of adequately cleaning the surfaces of the metals and maintaining the surfaces free from contamination by occluded gases.

[28] KINGDON, *Phys. Rev.*, **24,** 510 (1924).

[29] SPANNER, *Ann. d. Physik*, **75,** 609 (1924).

[30] ZWIKKER, *Proc. Amst. Acad. Sci.*, **29,** 792 (1926).

[31] DUSHMAN, ROWE, EWALD, and KIDNER, *Phys. Rev.*, **25,** 338 (1925); DUSHMAN and EWALD, *Phys. Rev.*, **29,** 857 (1927).

[32] DUBRIDGE, *Phys. Rev.*, **32,** 961 (1928).

Even the first two methods are extremely difficult of execution for obtaining accurate results, owing largely to the difficulty in accurately measuring the temperatures. Because of these experimental difficulties, the values of the emission constants obtained by various observers differ considerably in some cases. The best values, taken mostly from the "International Critical Tables," are given in Table II on page 65. The values printed in boldface type are the more reliable; the other values must be considered as only approximate.

Platinum seems to be abnormal in that the emission constants for this metal, as determined by a number of observers, vary much more widely than for other metals. DuBridge[33] has tabulated the values of A, b_0, and Φ_0 for platinum. The determinations of the electron affinity vary from 2.18 to 6.71 volts, and the values for A vary from 10.7×10^{-4} to 1.45×10^{7} $\frac{\text{amp.}}{\text{deg.}^2\ \text{cm.}^2}$. This wide variation may be due to occluded oxygen which condenses readily on clean platinum, as Langmuir[35] showed.

36. Space Current vs. Voltage.—Thus far, only the variation of the total emission or saturation current from an emitter, as its temperature is varied, has been explained. The electric field intensity, which draws the electrons from the emitter to the opposed cold plate, has been assumed sufficiently large (indicated by $E = \infty$ in Fig. 12) to insure that all emitted electrons are drawn off and a true measure of emission as a function of temperature obtained. If, however, the field intensity is small, all electrons are not attracted to the plate, and the space current depends upon the strength of the electric field according to a certain law known as the *voltage law.*

A rough physical picture of the reason for the limitation of space current at low voltages may be obtained by examining the forces acting upon a single electron as it passes from the emitter to the plate of a two-electrode tube, or *diode,* shown in Fig. 15. Assume that the bulb is perfectly exhausted and that a space current is flowing, consisting of a cloud of electrons diagrammatically represented in Fig. 15. This cloud constitutes a negative *space charge,* which exerts a force on each electron according to its position. An electron just emerging from the emitter is attracted toward the plate by the electric field caused

[33] DuBridge, *Phys. Rev.*, **31**, 236 (1928).

by the voltage E_p applied between emitter and plate, but is repelled back toward the emitter by the space charge. If the former force is greater than the latter, the electron will move toward the plate with an increasing velocity. The repelling force toward the emitter, due to the space charge, decreases as the electron approaches the plate, because less repelling space charge is in front of it and more behind it. Since the resultant force on the electron increases rapidly owing to the diminution and ultimate reversal of the space-charge force, the acceleration of the electron increases. The electron cloud is most dense near the emitter. The force acting on an electron to urge it toward the plate is least when the electron is near the emitter, so that an electron that starts toward the plate is certain to arrive there. If the force due to the space charge on an electron just emerging from the emitter is equal to the attracting force due to the charged plate, the electron will be unaffected, and any further increase of electrons flowing to the plate will result in a larger net force urging the electrons back into the emitter. It is evidently impossible for a current to pass to the plate which is greater than that giving a space charge which just neutralizes the force at the emitter due to the plate. This leads to a maximum value of current known as the *saturation current* corresponding to each plate voltage.

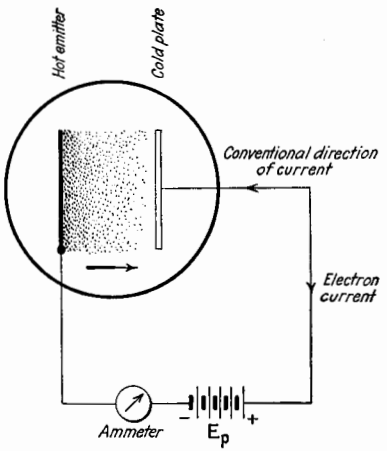


FIG. 15.—Diagrammatic illustration of space charge.

An analytical expression for the voltage law was first derived for parallel-electrode surfaces by Child[34] in 1911, and later independently derived and extended to a cylindrical plate by Langmuir.[35]

37. Voltage Law for Plane Surfaces.—The derivation of the voltage law, as obtained by Child, is here given when applied to the ideal case of two infinite parallel surfaces in a perfect vacuum,

[34] CHILD, *Phys. Rev.*, **32**, 498 (1911).

[35] LANGMUIR, *Phys. Rev.*, **2**, 450 (1913).

one surface being maintained at an infinite temperature. Further, the emitter is assumed to be an equipotential surface and the electrons are assumed to emerge from this hot surface with zero initial velocity. Figure 16 represents a portion of the two infinite surfaces. The hot surface is assumed to be at zero potential and the cold surface is maintained at potential E_p. Since the surfaces are infinite, the electric field is everywhere uniform and perpendicular to the surfaces. Imagine a small box between the surfaces, with edges of length dx, dy, dz, the direction x being perpendicular to the infinite surfaces. Gauss' theorem states that the surface integral over any closed surface of the normal component of the electric force F_n is equal to 4π times the charge q inside the surface, or

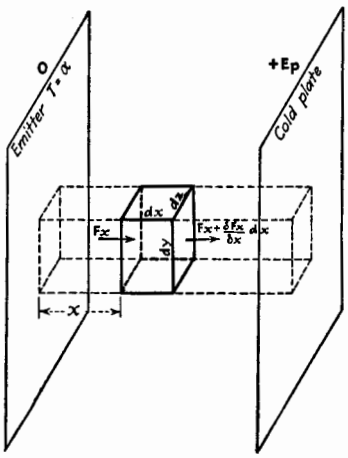


FIG. 16.—Flow of electrons between parallel plane electrodes.

$$\iint F_n dS = 4\pi q \text{ (e.s.u.)} \quad (28)$$

This equation may be applied to the small imaginary box in Fig. 16. The result is

$$-F_x dydz + \left(F_x + \frac{\partial F_x}{\partial x}dx\right)dydz = -4\pi\rho dxdydz \quad (29)$$

where ρ is the volume density of negative charge. In Eq. (29), the partial derivative may be written as a total derivative because F_x varies only with x, and the equation reduces to

$$\frac{dF_x}{dx} = -4\pi\rho \text{ (e.s.u.)} \quad (30)$$

Since $F_x = -dE/dx$, Eq. (30) becomes

$$\frac{d^2E}{dx^2} = 4\pi\rho \text{ (e.s.u.)} \quad (31)$$

where E is the potential distant x from the emitting surface.

The volume density ρ is a function of x. The charge is moving in the direction of positive x so that the charge, which is at one instant inside the box, in an interval of time dt will be displaced a distance dx. The rate of transfer of negative electricity, or the electronic current, is therefore

$$\frac{\rho dx dy dz}{dt} = I dy dz$$

or

$$I = \rho \frac{dx}{dt} = \rho v_x \tag{32}$$

I is the electron current per unit area of the surfaces and is numerically equal to the positive current flowing from anode to cathode; v_x is the velocity of the electrons at distance x from the emitting surface.

An electron which has traveled a distance x from the emitter has velocity v_x given by the equation

$$\tfrac{1}{2}mv_x^2 = Ee \tag{33}$$

Eliminating ρ and v_x from Eqs. (31), (32), and (33),

$$\frac{d^2E}{dx^2} = 2\pi\sqrt{\frac{2m}{eE}}I \text{ (e.s.u.)} \tag{34}$$

To solve the equation, we may multiply through by $2\frac{dE}{dx}$, giving

$$\frac{d}{dx}\left(\frac{dE}{dx}\right)^2 = 4\pi I\sqrt{\frac{2m}{e}}\frac{1}{E^{1/2}}\frac{dE}{dx} \text{ (e.s.u.)} \tag{35}$$

Integrating Eq. (35),

$$\left(\frac{dE}{dx}\right)^2 - \left(\frac{dE}{dx}\right)^2_{x=0} = 8\pi I\sqrt{\frac{2m}{e}}(E^{1/2} - E_0^{1/2}) \text{ (e.s.u.)} \tag{36}$$

Equilibrium is established when the current, and hence the space charge, is sufficient to reduce to zero the electric field at the emitter. The potential E_0 is zero by hypothesis. Equation (36) becomes

$$\left(\frac{dE}{dx}\right)^2 = 8\pi I\sqrt{\frac{2m}{e}}E^{1/2} \text{ (e.s.u.)} \tag{37}$$

Integrating Eq. (37),

$$I = \frac{1}{9\pi}\sqrt{\frac{2e}{m}}\frac{E^{3/2}}{x^2} \text{ (e.s.u.)} \tag{38}$$

where E is the potential at any distance x from the emitter. If x is made equal to d, the distance between emitter and plate, E becomes E_p and Eq. (38) becomes

$$I = \frac{1}{9\pi}\sqrt{\frac{2e}{m}}\frac{E_p^{3/2}}{d^2} \text{ (e.s.u.)} \tag{39}$$

Equation (39), expressed in practical units with the constants evaluated, is

$$I = 2.336 \times 10^{-6} \frac{E_p^{3/2}}{d^2} \frac{\text{amp.}}{\text{cm.}^2} \tag{40}$$

where E_p is expressed in volts and d in centimeters.

Equation (40) is the voltage law and shows that the voltage-saturation current varies as the three-halves power of the

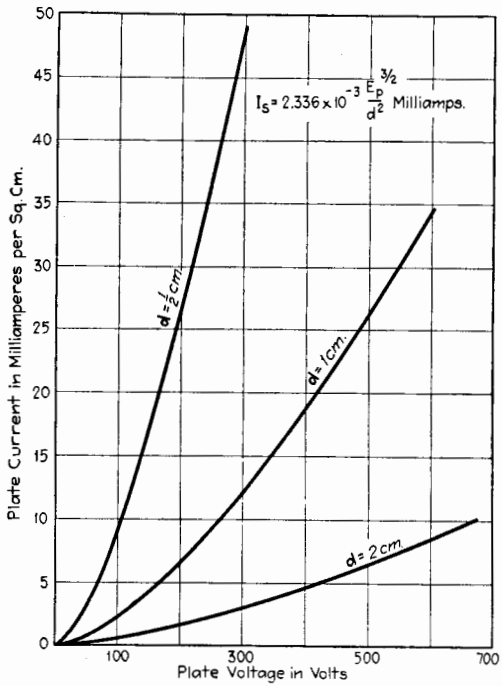


FIG. 17.—Space-charge saturation current between parallel plane electrodes.

impressed plate voltage. This equation is sometimes called Child's law.

Equation (40) also shows that with constant potential on the plate, the voltage-saturation current varies inversely as the square of the distance between a plane emitter and a plane plate. Equation (40) is plotted directly in Fig. 17, and on logarithmic paper in Fig. 18.

Since the current is the same for all values of x, Eq. (38) shows that the potential at any distance x is given by

$$E = 3\sqrt[3]{\frac{3\pi^2 m I^2}{2e}} \cdot x^{4/3}$$
$$= (\text{const.}) \cdot x^{4/3} \tag{41}$$

Therefore, the potential varies as the four-thirds power of the distance from the plane emitter.

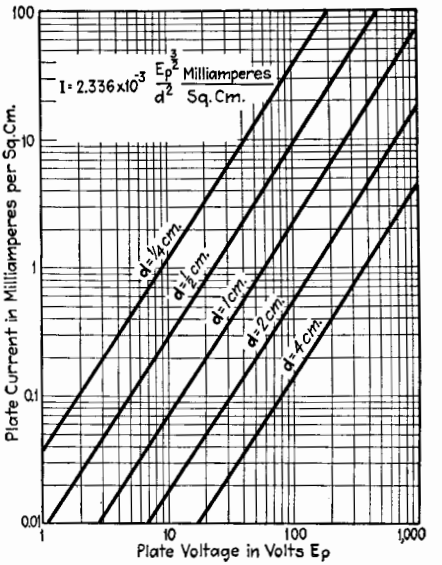


FIG. 18.—Space-charge saturation current between parallel plane electrodes.

38. Variation of Field Strength, Velocity, and Charge Density with Distance, for Plane Electrodes.—It is a simple operation to derive the following expressions:

$$F = -\frac{dE}{dx} = -(\text{const.}) \cdot x^{1/3} \tag{42}$$
$$v_x = (\text{const.}) \cdot x^{2/3} \tag{43}$$
$$\rho = (\text{const.}) \cdot x^{-2/3} \tag{44}$$

Equations (41) to (44) are shown graphically in Fig. 19. It is to be remembered that these equations all apply to the ideal case of an equipotential plane emitter, giving off electrons with no initial velocity, and placed in a perfect vacuum opposite a plane plate, both plate and emitter being so large that the field between them is everywhere uniform and perpendicular to the surfaces.

By Eq. (31), ρ is proportional to d^2E/dx^2 and hence is approximately proportional to the curvature of the graph of E.

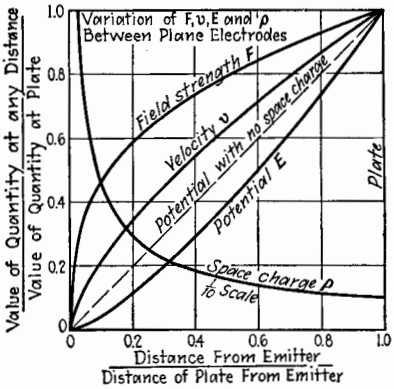


FIG. 19.—Variation of F, v, E, and ρ between parallel plane electrodes.

39. Voltage Law for Cylindrical Plate.—The law for the variation of plate current with plate voltage will now be developed for a cylindrical arrangement, consisting of a filamentary equipotential emitter of infinite length disposed along the axis of an infinite cylindrical plate. Here again the temperature of the emitter is assumed to be so large that the space current is never limited by lack of emission, and the electrons leave the filament with zero initial velocity.

Referring to Fig. 20, Gauss' theorem is first applied to the small wedge-shaped volume $rd\theta drdl$

$$-F_r rd\theta dl + \left(F_r + \frac{\partial F_r}{\partial r}dr\right)(r + dr)d\theta dl = -4\pi\rho rd\theta drdl \quad (45)$$

Reducing and dividing by $rd\theta drdl$ and replacing the partial derivative by the total derivative,

$$\frac{dF_r}{dr} + \frac{F_r}{r} = -4\pi\rho \quad (46)$$

Replacing F_r by $-dE/dr$,

$$\left.\begin{aligned} &\frac{d^2E}{dr^2} + \frac{1}{r}\frac{dE}{dr} = +4\pi\rho \\ \text{or}\quad &\frac{1}{r}\frac{d}{dr}\left(r\frac{dE}{dr}\right) = 4\pi\rho \end{aligned}\right\} \quad (47)$$

which expresses the relation among the potential E, the volume density of negative charge ρ, and the distance r from the axis of the cylindrical plate.

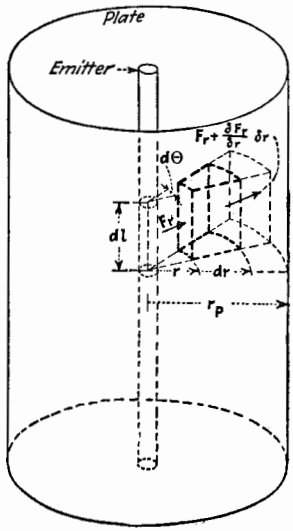


FIG. 20.—Flow of electrons between cylindrical electrodes.

The electron current perpendicular to the axis through length l of any imaginary concentric cylinder is

$$I = 2\pi r l \rho v_r \tag{48}$$

This equation, together with Eq. (33), can be combined with Eq. (47) to eliminate ρ and v_r, giving

$$r\frac{d^2E}{dr^2} + \frac{dE}{dr} = \frac{I}{l}\sqrt{\frac{2m}{Ee}} \tag{49}$$

Langmuir[36,37] has shown that the solution of this equation is

$$I = \frac{2}{9}\sqrt{\frac{2e}{m}}\frac{lE^{3/2}}{r\beta^2}\text{ (e.s.u.)} \tag{50}$$

In Eq. (50), β is a dimensionless factor given by the series

$$\beta = \log\frac{r}{a} - \frac{2}{5}\left(\log\frac{r}{a}\right)^2 + \frac{11}{120}\left(\log\frac{r}{a}\right)^3 - \frac{47}{3{,}300}\left(\log\frac{r}{a}\right)^4 + \cdots \tag{51}$$

where a is the radius of the filamentary emitter. Table III, taken from Langmuir's paper,[36,37,38] gives the values of β^2 for various values of r/a.

TABLE III[38]

r/a	β^2	r/a	β^2	r/a	β^2
1.00	0.000	6.0	0.838	30	1.091
1.50	0.116	7.0	0.887	45	1.095
2.00	0.275	8.0	0.925	67	1.089
2.50	0.405	9.0	0.955	122	1.072
3.00	0.512	10.0	0.978	221	1.053
4.00	0.665	12.0	1.012	735	1.023
5.00	0.775	16.0	1.051	2,440	1.006
		20.0	1.072	22,026	0.999

Equation (50), reduced to practical units and with constants evaluated, becomes

$$I = 14.68 \times 10^{-6}\frac{lE_p^{3/2}}{r_p\beta^2}\text{ amp.} \tag{52}$$

[36] LANGMUIR, *Phys. Rev.*, **2**, 450 (1913); *Phys. Zeits.*, **15**, 348 (1914); *Gen. Elec. Rev.*, **26**, 731 (1923).

[37] LANGMUIR and BLODGETT, *Phys. Rev.*, **22**, 347 (1923).

[38] For values of β^2 see Langmuir, *Phys. Rev.*, **21**, 435 (1923); and for table of values of β^2 for both cathode inside and cathode outside see Langmuir and Blodgett, *Phys. Rev.*, **22**, 347 (1923).

where l is the length of the filament and plate in centimeters, r_p is the radius of the plate in centimeters, and E_p is the potential between plate and filament in volts. Equation (52) is plotted logarithmically in Fig. 21.

40. Comparison of Voltage Laws for Plane and Cylindrical Plates.—Comparing Eqs. (39) and (50), it is seen that with both geometrical arrangements of filament and plate, *i.e.*, plane sur-

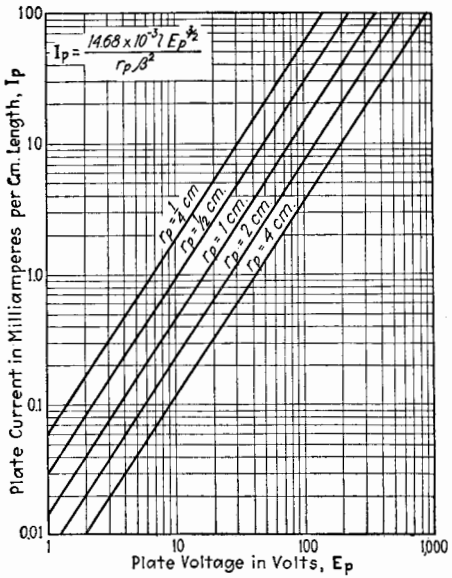


FIG. 21.—Space-charge saturation current between cylindrical electrodes. $\beta^2 = 1$.

faces and cylindrical plate with filamentary emitter, the electron current is proportional to the three-halves power of the impressed voltage. Langmuir[36] has shown theoretically that the three-halves-power law holds, no matter what the geometrical configuration of the electrodes may be.

Equation (52) shows that for a *cylindrical plate, where its radius is large compared with the radius of a filamentary emitter in the axis of the plate, the saturation current varies inversely as the first power of the radius of the plate.* This statement should be contrasted with the corresponding one for plane surfaces, in which case the saturation current varies inversely as the *square* of the distance between emitter and plate.

41. Variation of Field Strength, Velocity, and Charge Density with Distance, for Cylindrical Electrodes.—Expressions for the cylindrical diode, corresponding to those for the plane-plate diode given in Eqs. (41), (42), (43), and (44), are as follows:

$$E = (\text{const.}) \cdot (r\beta^2)^{\frac{1}{3}} \tag{53}$$

$$F = -\frac{dE}{dr} = -(\text{const.}) \cdot (r\beta^2)^{-\frac{1}{3}}\left(r\frac{d\beta^2}{dr} + \beta^2\right) \tag{54}$$

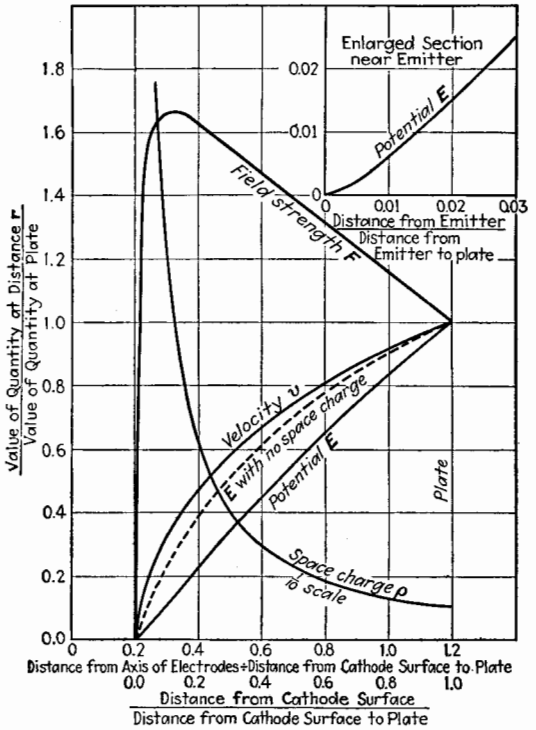


FIG. 22.—Variation of F, v, E, and ρ between cylindrical electrodes, $\frac{r_p}{a} = 6$.

$$v = (\text{const.}) \cdot (r\beta^2)^{\frac{1}{3}} \tag{55}$$

$$\rho = (\text{const.}) \cdot \frac{1}{rv} \tag{56}$$

These four equations are plotted in Fig. 22, which should be compared with Fig. 19 for the plane-plate diode.

42. Effect of Initial Velocity of Emission.—In the derivation of the voltage laws as given for the plane and cylindrical arrangements of plate and emitter, it was assumed that the electrons left

the emitter with no initial velocity. In practice this ideal condition is not realized, as was shown by the theory of emission given in the first part of this chapter. It was there pointed out that, because of the surface restraint, only those electrons inside the emitter possessing more than a certain minimum velocity can escape through the surface. This minimum velocity corresponds to a certain kinetic energy which an electron must possess in order to escape. This energy is transformed into potential energy on passing through the surface. Hence this minimum velocity is subtracted from the velocity of all electrons which emerge from the emitter. The emerging electrons therefore possess velocities ranging from zero upward. It was pointed out that the assumption made in the early theories of emission is that the electrons inside the emitter possess velocities which follow Maxwell's distribution law corresponding to the temperature of the emitter. If this assumption is true, it can be shown[39,40] from theoretical considerations that the electrons outside the emitter must possess the Maxwellian distribution of velocities corresponding to the temperature of the emitter, but with a different concentration of electrons. The new theories of emission based on statistical mechanics also give a Maxwellian distribution of velocities for the outside electrons but not for the inside electrons. The distribution of velocities of the outside electrons can be tested experimentally, and the results of several such investigations[41–50] prove that the initial velocities of emission do follow Maxwell's law for an electron atmosphere in temperature equilibrium with the hot emitter.

The space current, when limited by space charge, differs slightly from that given by the simple voltage law if the initial velocities of emission are appreciable. In studying this effect of the initial velocities of emission, examine the plot of potential

[39] RICHARDSON, *Phil. Mag.*, **18**, 695 (1909).
[40] MOTT-SMITH and LANGMUIR, *Phys. Rev.*, **28**, 727 (1926).
[41] RICHARDSON and BROWN, *Phil. Mag.*, **16**, 353 (1908).
[42] RICHARDSON, *Phil. Mag.*, **16**, 890 (1908); **18**, 681 (1909).
[43] SCHOTTKY, *Ann. Physik*, **44**, 1011 (1914).
[44] TING, *Proc. Roy. Soc.* (*London*), **98**, 374 (1920–1921).
[45] JONES, *Proc. Roy. Soc.* (*London*), **102**, 734 (1923).
[46] POTTER, *Phil. Mag.*, **46**, 768 (1923).
[47] RÖSSIGER, *Z. Physik*, **19**, 167 (1923).
[48] CONGDON, *Phil. Mag.*, **47**, 458 (1924).
[49] GERMER, *Phys. Rev.* (2), **25**, 795 (1925).
[50] DAVISSON, *Phys. Rev.* (2), **25**, 808 (1925).

at various distances from the emitter or cathode between plane parallel electrodes, when the initial velocities of emission are taken into account. The variation of potential E, when the electrons have no initial velocity, is expressed by Eq. (41), and plotted in Fig. 19 for plane surfaces. This curve of Fig. 19 is reproduced in Fig. 23.

If the emitted electrons have initial velocities of Maxwellian distribution, there is a place a short distance from the emitting surface where the potential is negative with respect to the emitter. If the potential of the plate is positive, the potential curve has a minimum E_m at a distance x_m from the cathode, as shown diagrammatically in Fig. 23. At some distance x_0, greater than x_m, the potential is zero. Since the force per unit negative charge is given by the slope of the potential curve, the force on the electrons between the emitter and x_m is negative and hence toward the emitter. At x_m the force is zero, and for values of x greater than x_m the force is toward the plate. If the plate voltage is large, the space charge is a maximum very close to x_m, and hence the average velocity of the electrons toward the plate is a minimum at this point. If the plate voltage is small, x_m is relatively large, as shown by the dotted curve of Fig. 23 for plate voltage E_p'. A curve is also shown for an insulated plate.

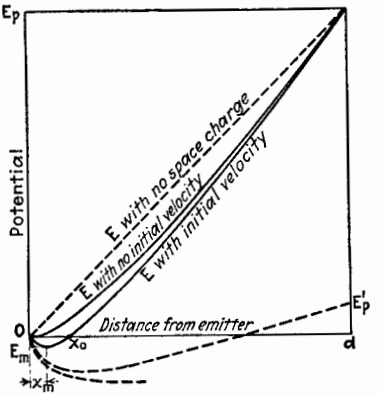


FIG. 23.—Variation of potential between plane parallel electrodes showing effect of initial velocity of emission.

Wehnelt and Bley[51] have measured the distribution of potential between the emitter and the plate under various conditions and find that x_m and E_m increase with a decrease of plate voltage and with an increase of temperature of the emitter, in accord with the theory.

Schottky,[52] Langmuir,[53] and others[54,55] have treated theoretically this effect of initial velocity and arrive at the following

[51] WEHNELT and BLEY, *Zeits. f. Physik*, **35**, 338 (1926).
[52] SCHOTTKY, *Phys. Zeits.*, **15**, 526 (1914); *Ann. d. Physik*, **44**, 1011 (1914).
[53] LANGMUIR, *Phys. Rev.*, **21**, 419 (1923).
[54] LANGE, *Jahrb. Radioakt. Elektronik*, **15**, 205 (1918).
[55] FRY, *Phys. Rev.*, **17**, 441 (1921).

approximate equation for the current when limited by space charge between plane electrodes

$$I = \frac{1}{9\pi}\sqrt{\frac{2e}{m}}\frac{(E - E_m)^{3/2}}{(x - x_m)^2}\left(1 + 2.66\sqrt{\frac{\kappa T}{e(E - E_m)}} + \cdots\right)(\text{e.s.u.}) \tag{57}$$

When the constants of Eq. (57) are evaluated and E_p is substituted for E to denote the plate voltage at a distance d from the emitter, the space current in amperes per square centimeter is

$$I = 2.336 \times 10^{-6}\frac{(E_p - E_m)^{3/2}}{(d - x_m)^2}\left(1 + 0.0229\sqrt{\frac{T}{E_p - E_m}} + \cdots\right) \tag{58}$$

Langmuir[53] gives the following table for the values of x_m and E_m for various values of E_p, when the distance d is 0.5 cm.

TABLE IV.—CURRENT BETWEEN PARALLEL PLANE ELECTRODES 0.5 CM. APART

1	2	3	4	5	6
I/I_s	I, amperes	E_p, volts	E_m, volts	x_m, centimeters	I/I_3
0.001	0.00016	2.5	−1.43	0.074	4.22
0.01	0.0016	24.4	−0.95	0.0224	1.424
0.1	0.016	131.6	−0.48	0.0062	1.134
1.0	0.16	645.0	−0.00	0.0000	1.045

The first column gives the ratio of the current I to the saturation current I_s. The sixth column gives the ratio of the current I to the current I_3 calculated by formula (40) for the ideal case.

The expression corresponding to Eq. (57) for the current per unit length of cylindrical plate of radius r, as given by Langmuir, is

$$I = \frac{2}{9}\sqrt{\frac{2e}{m}}\left[E_p - E_m + \frac{E_0}{4}\left(\log\frac{E_p}{\lambda E_0}\right)^2\right]^{3/2} \div (\beta^2 r)(\text{e.s.u.}) \tag{59}$$

where E_0 is the voltage corresponding to the average initial energy of the electrons in a radial direction. $E_0 = \frac{3}{2}\kappa T$ and is equal to 0.31 volt at $T = 2400°\text{K}$.; λ is a factor whose value lies between 1 and 2 and can, for practical purposes, be set equal to unity; β^2 in Eq. (59) has the same significance as in Eq. (50)

and is given in Table III for various values of the ratio of radius of cathode to that of plate.

The correction for the effect of initial velocity for the cylindrical arrangement is several times less than that for the plane parallel electrodes, and hence considerably less than the corrections indicated in Column 6 of Table IV.

Thus far, we have considered the electron current which flows from an emitter to the plate, only when the plate potential is positive with respect to that of the emitter. Of course, if the electrons were to emerge from the emitter with no initial velocity, the electron current to the plate would be zero when the plate potential is zero, the potential of the emitter always being taken as zero. Because of the initial velocities, some of the electrons are able to reach the plate even when the plate has a negative or retarding potential. The number that reach the plate for any retarding potential E_p depends upon the temperature of the emitter and the distribution of velocities of the emitted electrons. If we measure the number that reach the plate, or the current for various retarding potentials, we have a means of determining whether or not the initial velocities of emission follow Maxwell's distribution law.

Assuming that the emitted electrons obey Maxwell's distribution of velocities, the fraction n/n_s of the electrons which are capable of moving against a retarding potential E, is given by the Boltzmann equation[56]

$$\frac{n}{n_s} = \epsilon^{-\frac{Ee}{\kappa T}} \tag{60}$$

Equation (60) holds only for plane electrodes and only when there is no potential minimum between the two electrodes. This latter restriction means that the motion of electrons is not influenced appreciably by space charge, and that there is no secondary emission or reflection of electrons at the electrodes. Equation (60) can be written

$$\frac{I}{I_s} = \epsilon^{-\frac{Ee}{\kappa T}} \tag{61}$$

where I is the current corresponding to retarding potential E, and I_s is the saturation current corresponding to temperature T.

[56] RICHARDSON and BROWN, *Phil. Mag.*, **16**, 353 (1908); LANGMUIR and MOTT-SMITH, *Gen. Elec. Rev.*, **27**, 449 (1929); MOTT-SMITH and LANGMUIR, *Phys. Rev.*, **28**, 756 (1926).

If $\log_{10} I$ be plotted against E, retarding potentials being plotted negatively, the result is a straight line of slope $\frac{e}{\kappa T} \log_{10} \epsilon$ until I is nearly equal to I_s, when the curve bends over and becomes horizontal, as shown in Fig. 24. The two straight portions, if extended, meet on the axis for $E = 0$, provided no potential other than E is acting. If there is a Volta contact potential between the electrodes, the curve is displaced, as shown by the dotted curve of Fig. 24. The distance between the intersection point a and the axis gives the Volta potential E_v.

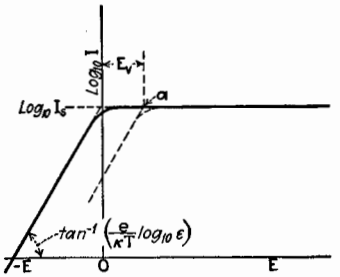


FIG. 24.—Space current between plane parallel electrodes as a function of retarding potentials.

The method outlined would provide an accurate method of determining the Volta potential difference between two metals composing the emitter and the plate, if it were possible to realize the condition of infinite plane electrodes. Since this is impossible, the cylindrical arrangement is always used, but the theory is not so simple. Schottky[57] has shown that, for the cylindrical arrangement, the current reaching the plate when the retarding potential is E, provided the emitter temperature is so small that space charge is negligible, is

$$I = I_s \frac{2}{\sqrt{\pi}} \left[\sqrt{\frac{Ee}{\kappa T}} \epsilon^{-\frac{Ee}{\kappa T}} + \int_{\frac{Ee}{\kappa T}}^{\infty} \epsilon^{-x^2} dx \right] \tag{62}$$

If E is numerically greater than about $3\frac{\kappa T}{e}$, the graph of $\log_{10} I$ is approximately a straight line. When extended it does not, however, intersect the horizontal line corresponding to I_s at a point which gives the correct value of the Volta potential. The intersection point gives a value of the Volta potential which is in error by the amount.

$$\text{Error in Volta potential} = -\frac{0.7\kappa T}{e} \text{ (approx.)} \tag{63}$$

A better method of testing the validity of Eq. (62), as well as to determine the Volta potential, is as follows:

[57] SCHOTTKY, *Ann. Physik*, **44**, 1011 (1914).

The values of $\log_{10} \frac{I_s}{I}$ from Eq. (62) for various values of $Ee/\kappa T$ have been calculated by Germer[58] and are given in the following table. If the retarding potential E_p is plotted against the value of $Ee/\kappa T$ as determined from Table V for the value of

TABLE V

$\frac{Ee}{\kappa T}$	$\log_{10} \frac{I_s}{I}$	$\frac{Ee}{\kappa T}$	$\log_{10} \frac{I_s}{I}$
1	0.2423	10	3.7698
2	0.5827	11	4.1850
3	0.9523	12	4.6024
4	1.3371	14	5.4398
5	1.7312	16	6.2812
6	2.1318	18	7.1245
7	2.5369	20	7.9714
8	2.9455	25	10.0978
9	3.3567		

$\log_{10} \frac{I_s}{I}$ obtained with voltage E_p, a straight line will be obtained, provided the velocities follow Maxwell's law. The intercept on the E_p axis gives the Volta potential, and the slope of the line gives the absolute temperature T of the emitter. Such a plot is shown in Fig. 25.

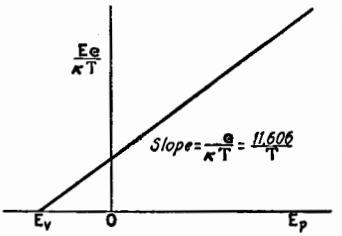


FIG. 25.—Determination of the Volta potential.

Many investigations[59] have been made, using methods similar to those outlined, which prove that the velocities of the electrons emitted from various pure and coated metals follow Maxwell's distribution law. In performing these experiments, it is to be noted that one of the requirements is a constant potential emitter. This is usually attained by the use of a commu-

[58] GERMER, *Phys. Rev.*, **25**, 795 (1925).

[59] RICHARDSON and BROWN, *Phil. Mag.*, **16**, 353 (1908); RICHARDSON, *Phil. Mag.*, **16**, 890 (1908); **18**, 681 (1909); SCHOTTKY, *Ann. d. Physik*, **44**, 1011 (1914); TING, *Proc. Roy. Soc. (London)*, **98**, 374 (1920–1921); JONES, *Proc. Roy. Soc. (London)*, **102**, 734 (1923); POTTER, *Phil. Mag.*, **46**, 768 (1923); RÖSSIGER, *Zeits. f. Physik*, **19**, 167 (1923); CONGDEN, *Phil. Mag.*, **47**, 458 (1924); GERMER, *Phys. Rev.*, **25**, 795 (1925); KOLLER, *Phys. Rev.*, **25**, 671 (1925); ROTHE, *Zeits. f. Physik*, **37**, 414 (1926).

tator, or the electrical equivalent, whereby the filament current is periodically interrupted, and the emission current is measured only during the time when the filament current is zero. Before leaving the subject of Volta potential, consider an ingenious scheme by Lange[60] for measuring the Volta potential difference between the grid and plate of a three-electrode tube. The method is based on the assumption that the division of current between the grid and plate, *i.e.*, the ratio I_g/I_p, depends only upon the ratio of potentials of grid and plate with respect to the emitter. Therefore, if E_{vg} and E_{vp} are the Volta potentials of grid and plate with respect to the filament, the ratio I_g/I_p is constant if $\frac{E_g + E_{vg}}{E_p + E_{vp}}$ is constant. In order to eliminate the effects of initial velocities of the electrons and the potential drop along the filament, the ratio $\frac{E_g + E_{vg}}{E_p + E_{vp}}$ is made equal to unity. Hence

$$E_g + E_{vg} = E_p + E_{vp} \text{ or } E_g = E_p + E_{v(p-g)} \tag{64}$$

where $E_{v(p-g)}$ is the Volta potential between plate and grid. The method consists in finding a value of $E_{v(p-g)}$ such that E_p and E_g can be varied in accordance with Eq. (64), while maintaining the ratio I_g/I_p constant.

43. Characteristic Surface of Diode.—Let us now investigate the shape of the characteristic curves of a two-electrode tube, or diode. The plate current is a function of both the plate potential and the temperature of the emitter. This can be expressed by the following equation:

$$I_p = f(E_p, T) \tag{65}$$

We have already examined the shape of the plate-current curve when E_p is infinite and T is varied, *i.e.*, Richardson's law, and the shape of the plate-current curve when T is infinite and E_p is varied, *i.e.*, the voltage law. It is now pertinent to study the shape the curves assume when both E_p and T are finite. The functional relation of Eq. (65), being an equation in three variables, can be plotted in a three-dimensional model, as shown in Fig. 26. Plate current I_p, the dependent variable, is plotted vertically, and plate voltage E_p and absolute temperature T are plotted on the horizontal plane. The model, of course, extends to infinity in the positive direction of all three variables, but the

[60] LANGE, *Diss. Dresden* (1927).

drawing shows only a small portion of the characteristic surface. Theoretically, the front edge, or the intersection of the two curved surfaces, is sharp, but, for reasons to be explained later, it is actually somewhat rounded as indicated in Fig. 26.

44. Characteristic Curves of Diode.—If the model is viewed in a direction perpendicular to the T-axis, the several sections of the model for various values of T give the family of curves shown in Fig. 27. If, on the other hand, the model is viewed in a direction perpendicular to the E_p-axis, another family of curves is obtained and is plotted in Fig. 28. Referring again to Fig. 26, every point on the surface which has its lower edge along the E_p-axis is independent of E_p and hence is determined only by the temperature T. Every such current is limited, therefore, by emission or what we may call *temperature saturation.* Similarly, every point on the other surface which starts from the T-axis represents a current determined only by E_p and hence a current limited by *space-charge saturation.* Those points on the rounded edge of the intersection of the two surfaces are limited simultaneously by both types of saturation.

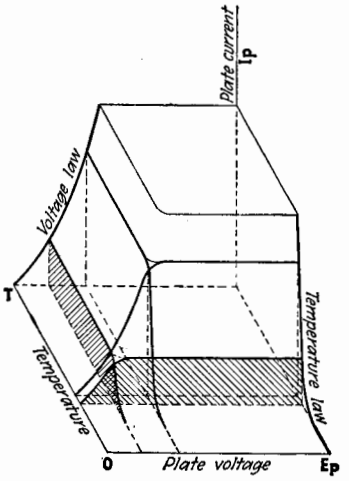


FIG. 26.—Characteristic surface of the diode.

The model of Fig. 26 is ideal and is often departed from in practice. Very often the edge is much more rounded and one or both of the surfaces may be inclined to the axis, so that saturation is less marked and the horizontal portions of the curves of Figs. 27 and 28 may be inclined with positive slopes.

45. Curves of I_p vs. I_f.—Figure 29 is a graph of the observed plate current plotted against filament current instead of against temperature. The envelope curve for a very large E_p is some complex function of the filament current.

The expression

$$I_p = (\text{constant}) \cdot \epsilon^{-\frac{g}{I_f}} \tag{66}$$

where g is a constant, fits very well the experimental observations, although there is no more theoretical justification for the formula

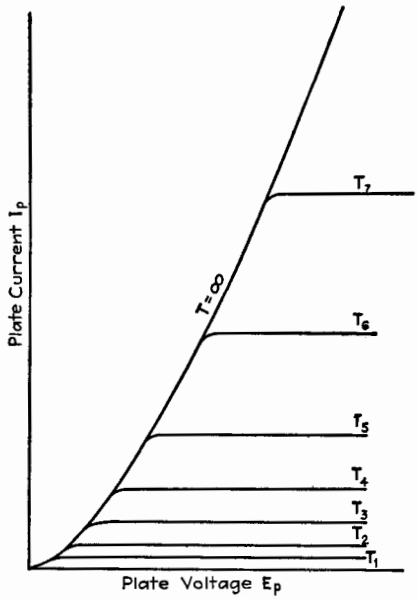


FIG. 27.—Plate current vs. plate voltage for the diode.

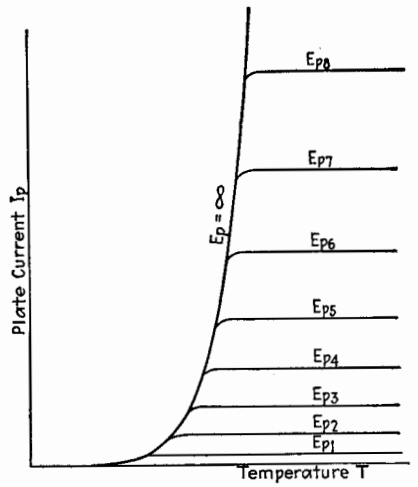


Plate current vs temperature for diode

FIG. 28.—Plate current vs. temperature for the diode.

than its similarity of form to Richardson's equation. The degree to which the expression fits experimental observation is indicated in Fig. 30, where $\log_{10} I_p$ is plotted against $1/I_f$.

Another expression which agrees with the experimental results with fair accuracy over a portion of the range of I_p is

$$I_p = CI_f^m \tag{67}$$

where C and m are constants. This expression has even less theoretical justification than Eq. (66) but does give a better

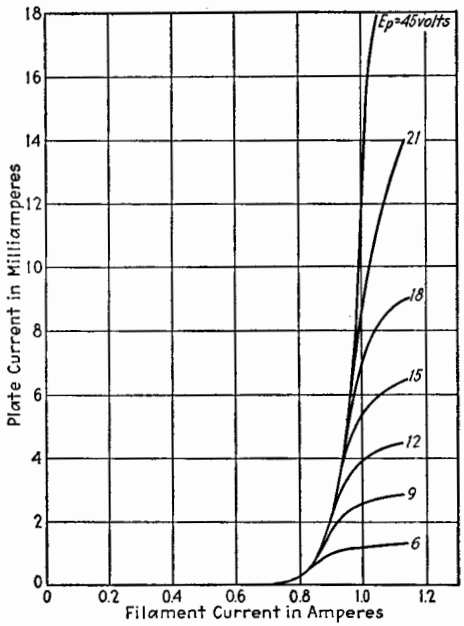


FIG. 29.—Plate current vs. filament current for a diode.

idea of how rapidly I_p changes with I_f, because m, as is seen from the logarithmic plot of Eq. (67) given in Fig. 31, is about 17.4.

The saturation plate current can be expressed in still another way which has several very practical advantages. This is only briefly referred to here, being described in more detail in Chap. V. The saturation current is a function of the heating power P dissipated in the filament, and the relation between I_p and P can be plotted conveniently on a special kind of coordinate paper. The emission current per watt can also be plotted as a function of P and is known as the *emission efficiency*.

Practical details concerning various types of emitters in common use and the effects on emission caused by potential drop along the filament, cooling of the end, etc., are foreign to this chapter, which is confined to the strictly ideal theoretical aspects of emission. The practical details are considered in the next chapter.

46. Cold-cathode Emission.—Reference was made in Sec. 30 to the idea, suggested by Schottky,[61] that the work function w

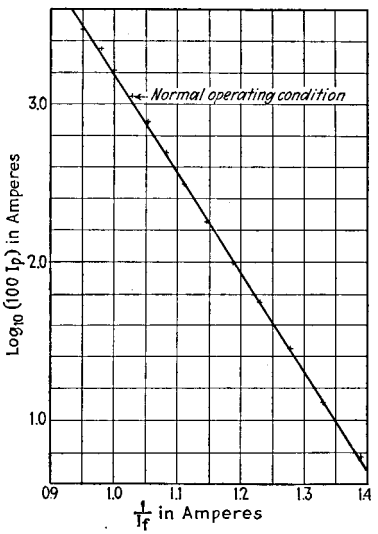


FIG. 30.—Emission of a tungsten filament (UV = 201).

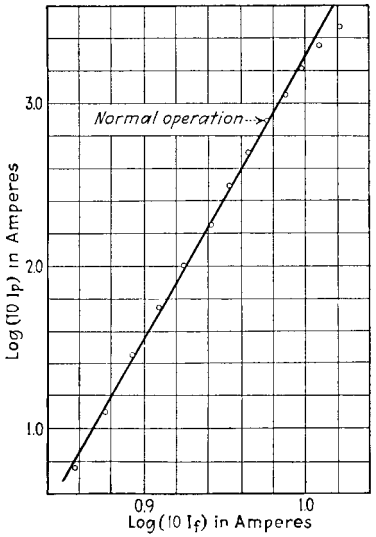


FIG. 31.—Emission of a tungsten filament (UV = 201).

is the work necessary to drag an electron from a metal in opposition to the electric force which pulls the electron back toward the surface. This force is generally assumed to be the ordinary electrostatic image force for distances from the surface greater than a distance x_0, which is of the order of a few diameters of the surface atoms, and Schottky assumed that for distances less than x_0 the force is constant and equal to the image force at x_0. If we assume that the temperature of the emitter is low, *i.e.*, of the order of room temperature, the velocities of agitation of the electrons can be neglected and the effect producing a minimum of potential outside an emitter, due to the initial velocities of emission, is practically out of the picture. Considering

[61] SCHOTTKY, *Phys. Zeits.*, **15**, 872 (1914); **20**, 220 (1919); *Ann. d. Physik*, **44**, 1011 (1914); *Zeits. f. Physik*, **14**, 63 (1923).

only this surface restraint at a cold electrode, we may represent diagrammatically the course of the force and potential as in Fig. 32. If $x = \infty$, the potential E becomes the electron affinity for the metal. Then,

$$\Phi = \frac{e}{2x_0} \tag{68}$$

Substituting in Eq. (68) the value of Φ for tungsten, *i.e.*, 4.52 volts, x_0 is about 1.6×10^{-8} cm.

When an outside field is impressed, tending to draw the electrons away from the electrode, we may represent the resultant potential as in Fig. 33. This resultant is the sum of the potential shown in Fig. 32 and the potential E' due to the outside field, where E' is given in terms of the rate of change of potential at the surface of the electrode. The curve for E' is not necessarily a straight line, although it is so shown in Fig. 33 for plane electrodes.

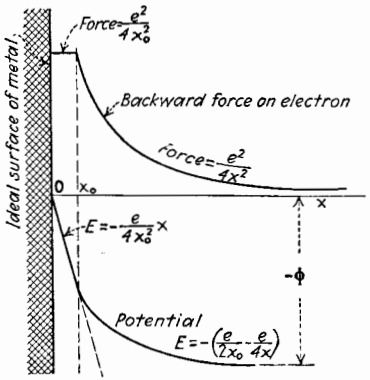


FIG. 32.—Schottky's picture of the force and potential near a surface of metal. No external impressed force.

The resultant potential has a minimum E_m at a distance x_m from the surface. At this point the force on the electron is zero and changes sign. If an electron reaches x_m, it is sure to leave the surface because of the outside field. The resultant force is also shown in Fig. 33. The work to detach an electron from the electrode surface against the forces holding it is represented by the area under the resultant-force curve up to x_m. Hence, the work function is *decreased* by the external field by the amount $(\Phi - E_m)e$.

It should be noted that the E_m and x_m as used here have no direct relation to the same letters as used in Fig 23.

E_m and x_m can be found by putting the derivative of the resultant potential with respect to x equal to zero. Thus,

$$x_m = \frac{1}{2}\sqrt{\frac{e}{dE'/dx}} \tag{69}$$

whence

$$\Phi - E_m = \sqrt{e\frac{dE'}{dx}} \text{ (e.s.u.)} \tag{70}$$

Equation (70), expressed in volts, becomes

$$\Phi - E_m = 3.78\sqrt{\frac{dE'}{dx}} \cdot 10^{-4} \text{ volt} \tag{71}$$

The apparent electron affinity E_m would be reduced to zero if

$$\Phi = 3.78\sqrt{\frac{dE'}{dx}} \cdot 10^{-4} \text{ volt} \tag{72}$$

For tungsten, for which $\Phi = 4.52$ volts, the value of dE'/dx would have to be $1.43 \cdot 10^8$ volts/cm. If such a field could be

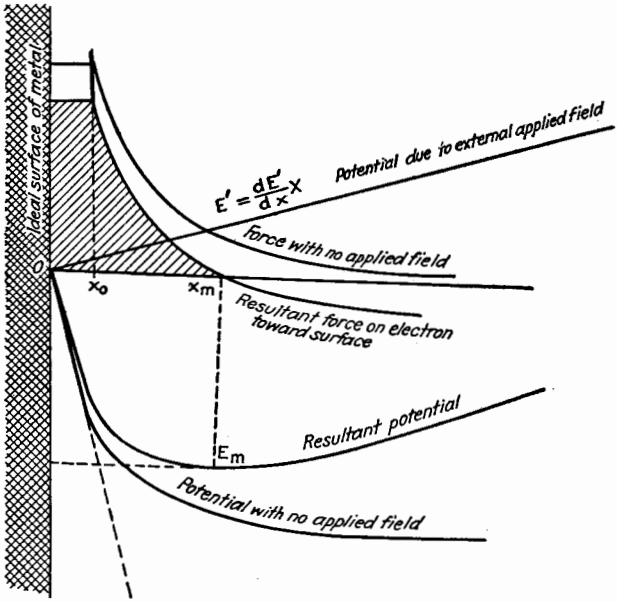


FIG. 33.—Force and potential near a metal surface when an external field is impressed. (*Schottky.*)

impressed, the saturation current, by Eq. (23), would become $I_s = AT^2$. Although such high fields cannot be obtained experimentally, the reduction of the work function by external fields is easily observable. In place of Eq. (23) we may write

$$I_s = AT^2\epsilon^{-\frac{e}{\kappa T}\left(\Phi_0 - \sqrt{e\frac{dE'}{dx}}\right)} \tag{73}$$

$$= AT^2\epsilon^{\xi}\epsilon^{-\frac{e\Phi_0}{\kappa T}} \tag{74}$$

where

$$\xi = \epsilon^{\frac{e}{\kappa T}\sqrt{e\frac{dE'}{dx}}} = \epsilon^{\frac{4.389}{T}\sqrt{\frac{dE'(\text{volts})}{dx}}} \tag{75}$$

is a factor which depends upon the field strength and temperature and in effect increases the value of the saturation current I_s as defined by Eq. (23).

The experimental values of log ξ plotted against $(dE'/dx)^{\frac{1}{2}}$ are found[62] to give a straight line for clean surfaces of pure metals when the field varies from a few volts per centimeter up to the highest fields attainable (10^6 volts per centimeter). Generally the slope of the logarithmic plot is larger than that calculated from the factors in Eq. (75), owing probably to microscopic roughness of the surface. With coated filaments, the decrease in work function due to an outside field is often very marked.

The Schottky effect, as just discussed, often causes the saturation portions of the curves of Fig. 27 to slope upward as the plate voltage is increased, especially when high voltages are used. Appreciable currents have been observed[63] and studied even when the emitting surface is at ordinary room temperature. Lilienfeld has succeeded in constructing very satisfactory X-ray tubes which depend upon this cold-cathode emission. The effect is of considerable importance in high-power high-voltage thermionic tubes.

II. SECONDARY EMISSION

47. Essential Facts Concerning Secondary Emission.—The second type of emission referred to in the first paragraph of this chapter is known as secondary emission and has been briefly described in Chap. II. A rapidly moving electron or ion on striking a solid substance may have sufficient energy to dislodge one or more electrons from the solid. These dislodged electrons are known as secondary electrons, although it is sometimes very difficult experimentally to distinguish the secondary electrons from the reflected primary electrons. A great deal

[62] SCHOTTKY, *Ann. d. Physik*, **44**, 1011 (1914); DUSHMAN *et al.*, *Phys. Rev.*, **25**, 338 (1925); LAURISTON and MACKEOWN, *Phys. Rev.*, **32**, 326 (1928); PFORTE, *Zeits. f. Physik*, **49**, 46 (1928); DE BRUYNE, *Proc. Roy. Soc. (London)*, **A 120**, 423 (1928); REYNOLDS, *Phys. Rev.*, **35**, 158 (1930).

[63] WOOD, *Phys. Rev.*, **5**, 1 (1897); ROTHER, *Ann. d. Physik*, **81**, 317 (1926); MILLIKAN and EYRING, *Phys. Rev.*, **27**, 51 (1926); **31**, 900 (1928); DE BRUYNE, *Proc. Cambridge Phil. Soc.*, **120**, 423 (1928); GOSSLING, *Phil. Mag.*, (7), **1**, 609 (1926); MILLIKAN and LAURITSEN, *Proc. Nat. Acad. Sci.*, **14**, 45 (1928).

of experimental work has been directed to the study of secondary emission.[64]

The essential facts concerning secondary emission, as summarized by Compton and Langmuir,[65] are as follows:

a. Secondary electrons may be derived from insulators as well as from conductors.

b. The number of secondary electrons per primary bombarding electron depends greatly upon the physical characteristics of the surface, thoroughly clean and degassed surfaces usually yielding less than contaminated surfaces, and a film of electropositive metal increasing the secondary emission.

c. The number of secondary electrons per primary electron increases up to a maximum, as the velocity of the primary electron is increased up to that corresponding to a few hundred volts, and then decreases as the velocity of the primary electron increases.

d. The maximum number of secondary electrons per primary electron reaches a maximum of from 1 to 1.5 for degassed surfaces, 3 to 4 for untreated surfaces, and 8 to 10 for metals coated with electropositive metals.

e. The velocities of the secondary electrons are low and of the order corresponding to only a few volts.[66]

f. The secondary electrons leave the surface in all directions, the angular distribution curve depending somewhat upon the characteristic of the surface.

g. Bombarding positive ions compared with electrons produce very few secondary electrons. The number per positive ion is at most only a few per cent of that for electrons for velocities corresponding to voltages less than a few hundred volts and is of the order of 10 to 20 per cent for velocities corresponding to 1,000 volts or more.

h. Metastable atoms meeting a surface result in the emission of secondary electrons in greater numbers than are produced by bombarding positive ions.

When secondary electrons are emitted by an electrode of a discharge tube, they return to the electrode unless there is in the vicinity a second electrode more positive in potential than the emitting electrode. Under this condition the emitting electrode may lose more electrons than it receives as primary electrons, so

[64] HYATT, *Phys. Rev.*, **32**, 922, and SMITH, 929 (1928). References to early work are given in the second reference.

[65] COMPTON and LANGMUIR, *Rev. Modern Phys.*, **2**, 123 (1930).

[66] FARNSWORTH, *Phys. Rev.*, **25**, 41 (1925).

that the electrode becomes positively charged if it is insulated. If the electrode is connected to the other electrodes through a battery which maintains the potential of the electrode at a positive value, the current to it may reverse on account of the secondary emission and the current may flow against the potential of the battery. This action may be made clearer by an example illustrated in Fig. 34.

The filament F supplies the primary electrons, some of which are drawn to electrode P_1 by the positive potential E_{B_1}. This electron stream is denoted by I_1. Other primary electrons go to the more positively charged plate P_2, the potential of which with respect to the filament is $E_{B_1} + E_{B_2}$. The primary electron current I_1 causes the emission of secondary electrons from plate P_1, and these are attracted to the second plate P_2 forming the electron stream I_2. Evidently the electron current to P_1 is $I_2 - I_1$. Secondary emission also takes place at P_2, but the secondary electrons return to P_2 since it is the most positive conductor in the system.

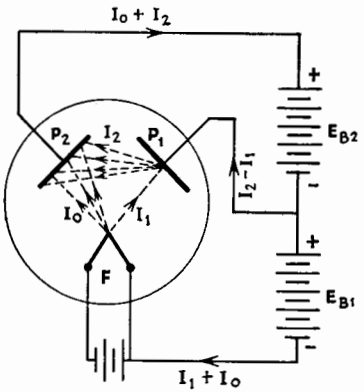


FIG. 34.—Secondary emission.

This same action may take place if P_1 is isolated or is an insulator, as, for example, a portion of the glass tube. When a steady state is reached, I_1 must equal I_2; but when the discharge is started, I_2 is momentarily greater than I_1, so as to impart to P_1 the necessary positive potential to maintain the electron stream I_1 to P_1. Langmuir[67] calls attention to this phenomenon in high-vacuum discharge tubes in which P_1 is a small portion of the glass envelope. The author has often observed the bombardment of the glass envelope due to secondary emission, which resulted in one case in melting a hole through the tube.

Some types of vacuum tubes making use of secondary emission have been invented and are known as *dynatrons*. These will be referred to in Chap. XXII. Secondary emission is often present in ordinary types of vacuum tubes and very much modifies the mode of operation. Such effects may be harmful and even

[67] LANGMUIR, *Gen. Elec. Rev.*, **23**, 513 (1920).

disastrous, as, for instance, when secondary emission causes the grid voltage of an oscillator tube suddenly to become positive, a phenomenon commonly known as "blocking."

III. PHOTOELECTRIC EMISSION

48. Essential Facts Concerning Photoelectric Emission.— Photoelectric emission is the third type of emission referred to in the early part of this chapter and was briefly described in Chap. II. When radiation, such as light of sufficiently short wave length, or X-rays, falls on the surface of a metal, electrons are dislodged. Since the radiant energy is absorbed by the surface atoms only in quanta, and since energy must be imparted to the electron sufficient to carry it through the surface restraint which is represented by the electron affinity Φ, electrons will be dislodged only if the frequency of light is greater than ν_0 where

$$h\nu_0 = \Phi e. \tag{76}$$

In Eq. (76), h is Planck's constant and $h\nu_0$ is the quantum of energy corresponding to ν_0. If the frequency of incident light is greater than ν_0, the quantum of energy is $h\nu$. The excess of absorbed energy over and above the amount Φe appears as kinetic energy of the emerging electron according to Einstein's equation[68]

$$h\nu = \Phi e + \tfrac{1}{2}mv^2_{\max.} \tag{77}$$

This holds provided the electron has no encounters with atoms after it is detached from the atom by the radiation. Such encounters cause a loss of kinetic energy. Some of the electrons, before emerging, do have these encounters, and the velocities of the emerging electrons range from zero to $v_{\max.}$ as given by Eq. (77). The most probable energy, except when photoelectric emission is caused by X-rays, is about one-half the maximum energy.

From the above discussion it is clear why the velocities of the emitted electrons, due to photoelectric action, are dependent only upon the wave length of incident radiation and the electron affinity Φ, whereas the number of emitted electrons depends upon the intensity of the incident radiation and upon the metal. Those metals having a low electron affinity give photoelectrons for longer wave lengths, but the efficiency of the metal as a photoelectric emitter is not directly a function of Φ. The

[68] EINSTEIN, *Ann. d. Physik*, **17**, 145 (1905).

photoelectric-emission curves of a few metals are given in Fig. 35, taken from "Photoelectric Cells" by Campbell and Ritchie. For other curves of photoelectric emission see a paper by Eleanor Seiler.[69] The alkali metals, sodium, potassium, rubidium and their alloys, are most active as photoelectric emitters and are used in photoelectric cells.

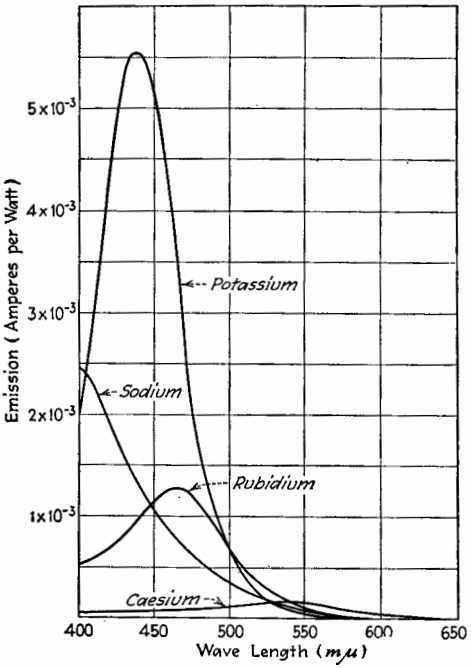


FIG. 35.—Average photoelectric emission curves of alkali metals. (*Campbell and Ritchie, Pitman.*)

General References

ALLEN: "Photoelectricity," Longmans Green & Co., New York, 1913.

RICHARDSON: "Emission of Electrons from Hot Bodies," Longmans, Green & Co., New York, 1916.

BLOCH: "Les Phénomènes Thermioniques," A. Blanchard, Paris (1923).

HEWLETT: "Electron Currents through a Vacuum and through Gaseous Atmospheres," Johns Hopkins Press, Baltimore, 1925.

SCHOTTKY, ROTHE, SIMON: "Handbuch der Experimental Physik," Wien-Harms, Vol. XIII, Akademische Verlagsgesellschaft M.B.H., Leipzig, 1928.

COMPTON and LANGMUIR: Electrical Discharges in Gases, *Rev. Modern Phys.*, **2**, 123 (1930).

[69] SEILER, *Astrophys. J.*, **52**, 129 (1920).

DUSHMAN: Thermionic Emission, *Rev. Modern Phys.*, **2**, 381 (1930).
CAMPBELL and RITCHIE: "Photoelectric Cells," Sir Isaac Pitman & Sons, New York, 1930.
ZWORYKIN and WILSON: "Photo Cells and Their Applications," John Wiley & Sons, Inc., New York, 1930.
HUGHES and DUBRIDGE: "Photoelectric Phenomena," McGraw-Hill Book Company, Inc., New York, 1932.

Special References

KIMBALL: Space Charge and Thermionic Currents, *Phil. Mag.*, **49**, 695 (1925).
GILL: Distribution of Electric Forces in Space Traversed by Electrons, *Phil. Mag.*, **10**, 134 (1930).

CHAPTER V

PRACTICAL SOURCES OF EMISSION AND SOME GENERAL PHYSICAL ASPECTS OF VACUUM TUBES

This chapter deals with practical information and data concerning thermionic emitters or cathodes as commonly used.

49. Common Types of Emitters.—At the present time there are three types of thermionic emitters or cathodes in use. Any of these types may take on different shapes and be heated in different ways. At present, however, we are concerned only with the physical and electrical characteristics of the three types of emitters, leaving until later a discussion of the structure and the method of heating the cathode.

The three types of emitters are:

I. *Filaments of pure metals.*

II. *Oxide-coated cathodes.*

III. *Metal filaments with an adsorbed monatomic film of one of the electropositive metals.*

I. FILAMENTS OF PURE METALS

Every metal when heated to a sufficiently high temperature emits electrons in accordance with the emission Eq. (23), page 59. But every metal at the same temperature does not emit the same number of electrons per second because of the significant difference in the values of the constants A and b_0 in Eq. (23). Only those metals are useful, as emitters of this particular class, which emit copiously at temperatures well below their melting and vaporizing points. Tungsten is the outstanding metal for filamentary emitters of pure metal. It has the highest melting point (3655°K.) of any ductile metal and emits electrons in satisfactory amounts when operated at temperatures at which evaporation is not impracticably great. Tantalum is also a highly satisfactory metal for emitters. Although its melting point (3120°K.) is lower than that of tungsten, its lower work function makes it a satisfactory emitter when operated at temperatures which are lower than the normal operating temperature

of tungsten. A comparison of the emission of tungsten and tantalum was shown in Fig. 13, Chap. IV.

Although emitters of the second and third types have replaced tungsten filaments in low-power vacuum tubes because of the lower heating power required for the same emission, these high-efficiency emitters do not stand up well in most high-power high-voltage vacuum tubes. Hence, cathodes of the first type, made of pure metals, are almost universally used in such high-power high-voltage tubes.

50. Tungsten Filaments.—The first tungsten filaments were made by Just and Hanaman about 1903 or 1904, but these filaments were brittle and non-ductile. In 1908 W. D. Coolidge found that tungsten can be made ductile by working and becomes brittle again if heated to a high temperature. Ductile tungsten is fibrous in structure owing to the deformation of the crystals and a breaking up of the crystalline structure. When the temperature of the ductile tungsten is sufficiently elevated, recrystallization takes place and the wire becomes brittle. After 1910 or early 1911 drawn tungsten wire came into general use for the filaments of incandescent lamps.

A great deal of work[1] has been done to determine for tungsten the constant A in the emission equation, Eq. (23), page 59, and also to determine the work function w_0 or electron affinity Φ_0. In consequence, these constants are probably better known for tungsten than for any other metal or substance. For pure tungsten, Eq. (23) is

$$I_s = 60.2T^2\epsilon^{-\frac{52,400}{T}} \frac{\text{amp.}}{\text{cm.}^2} \tag{78}$$

Since $\Phi_0 = 8.62 \times 10^{-5}b_0$, the value of Φ_0 for tungsten is 4.52 volts. Equation (78) is plotted in Fig. 12, page 61 of Chap. IV.

The operating temperature of any filamentary emitter is chosen to be as high as possible consistent with reasonable life. The life of a pure-metal filament is determined by the rate of evaporation of the metal and the consequent reduction in cross section. Following the practice of incandescent-lamp design, a life of 2,000 hr. is often used as a basis of reference. If the decrease in cross section is limited to 10 per cent, the tempera-

[1] DAVISSON and GERMER, *Phys. Rev.*, **20**, 300 (1922); DUSHMAN, ROWE, EWALD, and KIDNER, *Phys. Rev.*, **25**, 338 (1925); ZWIKKER, *Proc. Amst. Acad. Sci.*, **29**, 792 (1926).

ture of operation is that temperature which gives this decrease in 2,000 hr.

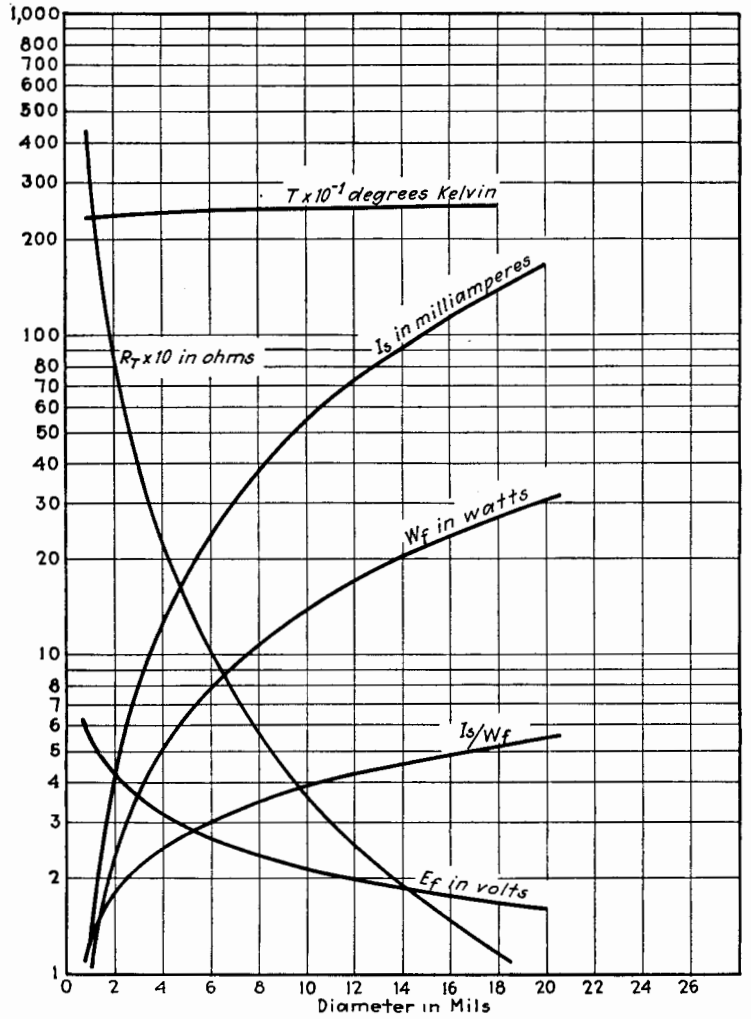


FIG. 36.—Characteristics of filament of pure tungsten 1 in. long operating at temperature to give 10 per cent decrease in cross section in 2,000 hr.

Since rate of evaporation is a function of both temperature and surface area, whereas the permissible life is dependent upon the volume of the filament, the operating temperature is different for filaments of different sizes.

Langmuir published in 1916[2] some of the properties of tungsten filaments, and in 1927 Jones and Langmuir[5] presented more complete data, including the rate of evaporation as a function of the temperature. From these data the curves of Fig. 36 were calculated giving the temperature in degrees Kelvin, the resistance in ohms, the heating power in watts, the voltage across the filament, and the emission current in milliamperes, for filaments 1 in. long and of various diameters in mils, operating at a temperature to give a 10 per cent decrease in cross section in 2,000 hr. A curve is also shown in Fig. 36 for the emission current in milliamperes per watt of heating power. Later in this chapter the emission efficiency of tungsten will be discussed in comparison with that of other emitters.

The emission and life of a tungsten filament are very much affected by the presence of certain gases. All of the gases, such as oxygen, nitrogen, carbon dioxide, water vapor, and hydrocarbons, which can combine with the filament, so change the surface of the emitter that the emission current is reduced. At a relatively low temperature oxygen forms an oxide over the surface of the filament, which very greatly reduces the emission. At high temperatures this oxide is evaporated. For nitrogen, Langmuir[3] found that the reduction of emission or "poisoning action" occurs only when the filament is bombarded by positive nitrogen ions. It is believed that nitrogen when ionized is more active and can then combine with tungsten to form WN_2. Hydrogen, however, can reduce the surface compounds and restore emission in some degree. Water vapor has a particularly powerful effect in reducing emission and at a pressure of only 10^{-6} mm. of mercury produces very great changes in space current. The hydrocarbons and carbon dioxide are reduced at the high temperature of the filament. The carbon combines with the filament forming tungsten carbide (W_2C) and, in the presence of a large amount of carbon, WC is formed. This results in a considerable increase in resistance of the filament, as shown by the work of Andrews.[4] Inert gases, such as argon,

[2] LANGMUIR, *Phys. Rev.*, **7**, 302 (1916).

[3] LANGMUIR, *Phys. Rev.*, **2**, 461 (1913).

[4] ANDREWS, *J. Phys. Chem.*, **27**, 270 (1923).

[5] JONES and LANGMUIR, *Gen. Elec. Rev.*, **30**, 310, 354 (1927); FORSYTHE and WORTHING, *Astrophys. J.*, **61**, 146 (1925); *Phys. Rev.*, **18**, 144 (1921); ZWIKKER, *Physica*, **5**, 249 (1925); *Arch. Néerland. sci.*, **9**, 207 (1925).

helium, neon, and mercury vapor, have practically no direct effect upon the emission from tungsten.

The presence of these inert gases, although having no direct effect upon the emission, may have a great effect upon the life of the filament for the following reason. Gas is ionized in a vacuum tube if the plate voltage is sufficiently high. The positive ions are driven to the filament and strike it with a velocity dependent upon the plate potential. This bombardment may cause a tearing off of particles of metal, perhaps of molecular size, which results in a disintegration of the filament and a rapid reduction of size with concurrent blackening of the glass walls of the tube. This "sputtering" of the filament is especially marked in argon. There is, however, a critical voltage for each gas below which this sputtering does not take place. The critical voltage for air is of the order of 25 volts.

Attention has already been directed to the fact that when used for filaments the physical properties of tungsten undergo some change. Drawn tungsten wire is tough, strong, and springy, can be bent without breaking when cold, and can be drawn, rolled, and otherwise mechanically worked when red hot. The drawn or swedged tungsten has a fibrous structure. If a large wire is bent near an end, the wire usually splits into a number of fibers. After the tungsten filament has been heated to incandescence for a short time, recrystallization takes place and the tungsten becomes very brittle. In fact, the purer the tungsten, the larger are the crystals and the more apt the filament is to fracture. It was discovered by accident in the laboratories of the General Electric Company that the addition of small amounts of certain foreign substances, such as thoria, to the tungsten reduces this tendency to crystallization, and all lamp manufacturers now use tungsten with a small percentage of thoria or some other equally effective substance. The presence of thoria, or any one of a number of other foreign substances, makes enormous changes in the emission characteristics of tungsten.

The electrical properties of tungsten have been determined independently by several investigators.[5] Table VI gives results obtained by Jones and Langmuir. The characteristics in Table VI are expressed for a wire 1 cm. long and 1 cm. in diameter. Column 2 gives the watts radiated from the surface of a wire of these dimensions, *i.e.*, from a surface of π sq. cm. Column 3 gives the resistance in microhms. The current in amperes

TABLE VI.—SPECIFIC CHARACTERISTICS OF IDEAL TUNGSTEN FILAMENTS[1]
(For wire 1 cm. long and 1 cm. in diameter)

1	2	3	4	5	6	7	8	9
	W'	$R' \times 10^6$	A'	$V' \times 10^3$	C'	L'	I'	R'_T/R'_{293}
T, °K.	W/ld	$(Rd^2/l) \times 10^6$	$A/d^{3/2}$	$(V\sqrt{d/l}) \times 10^3$	C/ld	$1.02C'\pi^2$	I/ld	R_T/R_{293}
	watts cm.$^{-2}$	ohms cm.	amp. cm.$^{-3/2}$	volts cm.$^{-1/2}$	int. candles cm.$^{-2}$	lumens cm.$^{-2}$	amp. cm.$^{-2}$	
273	.....	6.37	.....	.....	.....	.....	.....	0.911
293	0.0	6.99	0.0	0.0	.....	.....	.....	1.00
300	0.000100	7.20	3.73	0.027	.....	.....	.....	1.03
400	0.00624	10.26	24.67	0.253	.....	.....	.....	1.467
500	0.0305	13.45	47.62	0.640	.....	.....	.....	1.924
600	0.0954	16.85	75.25	1.268	.....	.....	.....	2.41
700	0.240	20.49	108.2	2.218	.....	.....	.....	2.93
800	0.530	24.19	148.0	3.581	.....	.....	.....	3.46
900	1.041	27.94	193.1	5.393	.....	.....	.....	4.00
1000	1.891	31.74	244.1	7.749	0.00013	0.00131	3.36×10^{-15}	4.54
1100	3.223	35.58	301.0	10.71	0.0011	0.0111	4.77×10^{-13}	5.08
1200	5.210	39.46	363.4	14.34	0.0065	0.0655	3.06×10^{-11}	5.65
1300	8.060	43.40	430.9	18.70	0.0285	0.286	1.01×10^{-9}	6.22
1400	12.01	47.37	503.5	23.85	0.107	1.08	2.08×10^{-8}	6.78
1500	17.33	51.40	580.6	29.85	0.343	3.44	2.87×10^{-7}	7.36
1600	24.32	55.46	662.2	36.73	0.956	9.60	2.91×10^{-6}	7.93
1700	33.28	59.58	747.3	44.52	2.40	24.1	2.22×10^{-5}	8.52
1800	44.54	63.74	836.0	53.28	5.27	53.0	1.40×10^{-4}	9.12
1900	58.45	67.94	927.4	63.02	11.27	113.3	7.15×10^{-4}	9.72

[1] Jones and Langmuir, *General Electric Rev.*, **30**, 310, 354 (1927).

TABLE VI.—SPECIFIC CHARACTERISTICS OF IDEAL TUNGSTEN FILAMENTS.[1]—(*Continued*)

1	2	3	4	5	6	7	8	9
	W'	$R' \times 10^6$	A'	$V' \times 10^3$	C'	L'	I'	R'_T/R'_{293}
T, °K.	W/ld	$(Rd^2/l) \times 10^6$	$A/d^{3/2}$	$(V\sqrt{d/l}) \times 10^3$	C/ld	$1.02C'\pi^2$	I/ld	R_T/R_{293}
	watts cm.$^{-2}$	ohms cm.	amp. cm.$^{-3/2}$	volts cm.$^{-1/2}$	int. candles cm.$^{-2}$	lumens cm.$^{-2}$	amp. cm.$^{-2}$	
2000	75.37	72.19	1022	73.75	21.3	214.0	3.15×10^{-3}	10.33
2100	95.69	76.49	1119	85.57	38.9	391	1.23×10^{-2}	10.93
2200	119.8	80.83	1217	98.40	65.9	662	4.17×10^{-2}	11.57
2300	148.2	85.22	1319	112.4	106.8	1073	1.28×10^{-1}	12.19
2400	181.2	89.65	1422	127.5	169.4	1702	0.364	12.83
2500	219.3	94.13	1526	143.6	255.5	2567	0.935	13.47
2600	263.0	98.66	1632	161.1	375.0	3770	2.25	14.12
2700	312.7	103.22	1741	179.7	548	5510	5.12	14.76
2800	368.9	107.85	1849	199.5	754	7575	11.11	15.43
2900	432.4	112.51	1961	220.6	1017	10220	22.95	16.10
3000	503.5	117.21	2072	243.0	1364	13720	44.40	16.77
3100	583.0	121.95	2187	266.7	1798	18070	83.0	17.46
3200	671.5	126.76	2301	291.7	2320	23300	150.2	18.15
3300	769.7	131.60	2418	318.3	2980	29950	265.2	18.83
3400	878.3	136.49	2537	346.2	3770	37880	446.0	19.53
3500	998.0	141.42	2657	375.7	4680	47000	732.0	20.24
3600	1130	146.40	2777	406.7	5700	57250	1173	20.95
3655	1202	149.15	2838	423.4	6350	63800	1505	21.34

and the voltage in millivolts are given in Columns 4 and 5, respectively. Column 6 gives the light radiation in international candles, and Column 7 the light flux in lumens. The electron emission in Column 8 is the current in amperes from π sq. cm.

The power W in watts radiated from 1 sq. cm. of tungsten as a function of the absolute temperature T is of some interest and is given by the following expression as determined by Worthing and Forsythe.[5]

$$\log_{10} W = 3.680(\log_{10} T - 3.3) - \frac{1{,}040}{T} + 1.900 \qquad (79)$$

This is of value in calculating the size of the plates of a power tube to operate below the temperature at which appreciable emission occurs.

51. Tantalum Filaments.—Tantalum filaments in incandescent lamps were first introduced about 1902 by von Bolton, but were soon displaced by the more refractory metal tungsten. Tantalum when pure is hard but ductile, so that it can be drawn into fine wire.

The melting point of tantalum is about 3120°K., much lower than that of tungsten. Its electron affinity is about 4.07 volts as compared to 4.52 for tungsten. In consequence, as shown by Fig. 12, page 61, the emission of tantalum is roughly ten times that of tungsten if both are at the same operating temperature within the range from 2000 to 2500°K. Tantalum filaments are used to some extent in high-voltage power tubes.

Tantalum, like tungsten, may become brittle after being heated to a high temperature for some time. This brittleness is due to recrystallization and the development of large crystals.

Tantalum is especially sensitive to residual gases. Oxygen and water vapor cause the formation of tantalum pentoxide with consequent great reduction in emission. Tantalum absorbs large volumes of hydrogen, which causes the emission to decrease and also makes the tantalum very brittle. Only by melting the tantalum in a vacuum can all of the hydrogen be driven off.

II. OXIDE-COATED CATHODES

52. The Early Oxide-coated Cathode.—The second type of emitter was developed from a discovery made by A. Wehnelt,[6]

[6] WEHNELT, *Ann. d. Physik*, **14**, 425 (1904).

in 1904, that a coating of certain oxides on a metal core gives a copious emission of electrons even at comparatively low temperatures. Such an emitter is known as the oxide-coated or *Wehnelt cathode*. Wehnelt, in his experiments, coated a platinum-strip heater with oxides of various elements and found that the oxides of the rare-earth metals barium, strontium, and calcium were especially effective in permitting large currents to pass through the discharge tube. In practically all of Wehnelt's early experiments considerable gas was present in the tube so that the path of the beam of electrons was made luminous by the intense ionization of the gas. Wehnelt did not have pure electron emission because the presence of large numbers of positive ions neutralized the space charge, thus allowing large currents to pass, and the bombardment of the cathode by the positive ions presumably influenced the emission from the cathode.

The core metal used by Wehnelt and others was platinum or platinum-iridium alloy, the iridium being added to give greater tensile strength. Because of the expense of platinum, other materials for the core were sought and now various metals and alloys are used. The most common are pure nickel, alloys of nickel with platinum, with silicon, and with cobalt, iron, and titanium, this last alloy being known as "Konel" metal.

For the coating of modern oxide-coated emitters, only compounds of barium and strontium are generally used. Except for the ease of preparation of the emitter, it makes little difference which of the many chemical compounds of barium and strontium is chosen as the raw material.

53. Mechanism of Emission from Oxide-coated Cathodes.— For a few years after Wehnelt's discovery there was considerable controversy, as shown by the writings of Wehnelt,[7] Soddy,[8] and others,[9] as to whether or not the residual gas caused the large currents to flow by some chemical effect at the cathode.

It is now generally believed that the emission of the oxide-coated cathode, when made with oxides of the rare-earth metals, is pure thermionic emission from particles of the metals reduced from the oxides, the oxide serving only as a reservoir for the

[7] WEHNELT, *Phil. Mag.*, **10,** 80 (1905).
[8] SODDY, *Nature*, p. 53, November, 1907.
[9] FREDENHAGEN, *Leipziger Ber.*, **65,** 42 (1913).
LANGMUIR, *Phys. Rev.*, **11,** 484 (1913); *Proc. I.R.E.*, **3,** 261 (1915).

metals. The high emission at low temperatures can be attributed to the low work function of these metals and the effects of single atomic layers which aid the escape of electrons.

There are still differences of opinion as to the *location of the emitting layer or islands of metal,* as to the *mechanism of the initial reduction of the metal,* and as to the *maintenance of the active metal during the life of the filament.*

54. Location of Active Metal.—The ideas concerning the location of the emitting centers fall naturally into two schools, and each has convincing experimental evidence to support its views. One school, supported by Koller,[10] Rothe,[11] Espe,[12] Detels,[13] Becker,[14] and others, holds that the emitting metal is on the *outside surface* of the oxide coating and that the core metal, upon which the oxides are coated, plays *no other rôle* than that of a support and an electrical conductor.

The other school, including Lowry,[15] Wagner,[16] Reimann and Murgoci,[17] believes that the active emitting layer is a monatomic layer of the electropositive metal which is adsorbed on the surface of the core metal under the oxide coating. For example, Lowry,[15] to explain the great increase in emission observed by him when the core metal was changed from platinum to an alloy of nickel, cobalt, iron, and titanium known as "Konel" alloy, holds that, since the base metal makes so much difference, the emitting layer of reduced metal must reside at the interface between core and oxide. Increased knowledge of emitters of the third type has to some extent given support to the view held by Lowry, Wagner, *et al.* Wagner[16] believes that the active metal is at the interface between core and coating. His argument for this belief is that, if electrolysis of the coating by the emission current takes place, the positively charged ions of strontium and barium are transported toward the core. Wagner holds that some chemical tests of coated cathodes which have been activated show the presence of active metal on the surface of the core metal.

[10] Koller, *Phys. Rev.*, **25,** 671 (1925).
[11] Rothe, *Zeits. f. Physik,* **36,** 737 (1926).
[12] Espe, *Wiss. Veröffentlichungen Siemens Konzern,* **5,** 29 (1927).
[13] Detels, *Jahrb. Drahtlos. Tel. Tel.*, **30,** 10, 52 (1927).
[14] Becker, *Phys. Rev.*, **34,** 1323 (1929).
[15] Lowry, *Phys. Rev.*, **35,** 1367 (1930).
[16] Wagner, *Electronics,* **1,** 178 (1930).
[17] Reimann and Murgoci, *Phil. Mag.*, **8,** 440 (1930).

55. Activation or the Initial Reduction of the Active Metal.—The oxide-coated cathode, made by coating a core metal with compounds of strontium and barium and mounted in a tube which is exhausted to a high vacuum, usually requires treatment known as *activation* or "breakdown" to establish substantial electron emission. Before discussing the various methods of activation, it is well to point out that there are two kinds of oxide-coated cathode resulting from differences in the methods of coating. The methods of preparation will be given more in detail later, but for the present it is sufficient merely to point out the essential differences in preparation which lead to the two kinds of cathode.

The two kinds are known as (1) the *uncombined kind* and (2) the *combined kind.* The *first* or uncombined kind, as the name implies, has a coating of compounds of strontium and barium, usually the carbonates or nitrates, which are only mechanically held on the core metal with no chemical combination therewith. This kind of coating is made by drying on the coating at relatively low temperatures and usually in an atmosphere of carbon dioxide. When this kind of cathode is finally heated to a high temperature in the exhausted tube, the carbonates or nitrates are broken down into oxides.

The *second* or combined kind is made by heating the coated cathode in air to a temperature of about 1000°K. At this temperature the compounds of strontium and barium which are in contact with the core metal combine with it to form platinates or nickelates according to the metal used for the core. These coatings are usually dark colored as contrasted with the white coatings of the first kind of cathode.

Both kinds of cathode, when finally mounted in the exhausted tube, are coated with oxides, and in the second kind with more complex compounds as well. These oxides and compounds do not emit electrons and the activation process consists in the reduction of some of the compounds to metallic barium and strontium. There are differences of opinion as to the mechanism of this reduction process. The opinions may be classified under the following three heads: (1) *reduction by electrolysis,* (2) *reduction by positive-ion bombardment,* (3) *thermal reduction.*

The observations on which these several opinions are based are as follows: The cathode, after being heated and the tube exhausted, may or may not have initial activity. The combined

kind of cathode usually has practically no initial activity. If the core metal of the uncombined kind of cathode is nickel, siliconickel, or Konel metal, and if the temperature of the cathode has been properly controlled during preparation, the cathode is initially active. If, on the other hand, the core is of chromium nickel, the cathode has no initial activity. The cathodes which have practically no initial activity can be activated by applying a voltage of one or two hundred volts between a plate and the heated cathode so as to draw emission from the cathode. The emission is small at first but can be increased by elevating the temperature. The emission builds up as activation proceeds and the temperature of the cathode can then be progressively reduced. This activation process may take only a few minutes or it may require several hours. Usually, gas is evolved during the activation, and the vacuum tube is filled with the blue glow of ionized gas.

Espe[12] suggested that electrolysis, due to the conduction of the space current through the hot oxide layer, brings about the reduction of the oxide to metal. The metal in this case would be transported and freed at the surface of the core, but diffusion of the metal through the layer to the outer surface would then take place at the high temperature of the layer of oxide. That such electrolysis of the oxide layer does actually take place is believed by Rothe,[11] Detels,[13] Becker,[14] Lowry,[15] and others. Detels found that oxygen was given off when a space current was drawn from an oxide-coated cathode, but Becker showed that oxygen was liberated only when the space current was not limited by space charge. Becker states that the rate of liberation of oxygen increases with the temperature, with the space current drawn from the cathode, with the plate potential, and depends somewhat on the composition of the oxide layer. If any electrolysis takes place, oxygen is liberated at the outer surface of the oxide coating. Some of it then diffuses into the layer where it combines with barium and strontium metal. That which does not diffuse into the layer may combine with the active barium and strontium on the surface, and some very small portion of the oxygen may be liberated and be driven into the glass walls or combine with other substances in the tube. We might picture a continuous cyclic electrolysis of the oxide and recombination of constituents to form oxides, a film of active metal being maintained on the surface.

Since gas is usually present in the activation process, some scientists hold that positive-ion bombardment accounts for the reduction of the oxides. McNabb[18] believes that this bombardment is necessary to activation. Wagner[16] has calculated that positive ions of CO_2, driven by a plate potential of 100 volts, have sufficient energy to break down BaO and SrO into metal and oxygen. Although activation may be hastened by positive-ion bombardment, such bombardment in high vacuum can hardly explain the maintenance of the supply of active reduced metal, which continually evaporates from an active oxide-coated cathode, as Becker,[14] Davisson,[19] and Eglin[20] have shown to occur.

When the base metal is nickel or certain nickel alloys, or even platinum, electrolysis and positive-ion bombardment cannot account for the initial activity of the uncombined coated filaments. It is possible that the barium and strontium oxides can be dissociated by heat, or by displacement of the barium or strontium by the base metal when platinum or nickel is used. Wagner[16] has calculated from the heats of formation of the various oxides that iron, chromium, and titanium are sufficiently active chemically to liberate barium and strontium from their oxides. Yet there is no initial activity when a nickel-chromium base metal is used. At present there is no satisfactory complete explanation of the initial activity sometimes observed, unless it is due to thermal reduction.

In some cases there is a possibility of misjudgment concerning initial activity. It is common practice to degas the cold electrodes of vacuum tubes by heating them to a high temperature by means of eddy currents. These currents are induced in the electrodes by a powerful high-frequency current traversing a coil of a few turns placed around the tube. During this degassing process, intense electric fields are induced in the space within the tube and ionization usually takes place, as evidenced by the blue glow within the tube. Capacitance currents surge to and from the electrodes, and positive-ion bombardment exists. If the cathode is heated, electrolysis of the oxide coating may take place. Although, under the proper conditions, activation may take place during this high-frequency treatment, initial activity

[18] McNabb, *J. Optical Soc. Am. Rev. Sci. Inst.*, **19**, 33 (1929).
[19] Davisson, *Phys. Rev.*, **34**, 1323 (1929). See p. 1332.
[20] Eglin, *Phys. Rev.*, **31**, 1127 (1928).

even without this treatment has sometimes been observed. Such initial activity has been reported by Lowry[15] when the only act was the heating of the cathode for a few minutes to an abnormally high temperature with a heating power of about twice the normal value.

A summary of the most plausible theory of action of an oxide-coated filament follows. First, we may be reasonably sure that the chemical compounds forming the coatings are the storage source from which the active metals barium and strontium are reduced by electrolysis and sometimes, in part, by positive-ion bombardment. Also, thermal dissociation may take place, assisted in some way, perhaps by catalytic action or by the chemical activity of the vapor of the core metal. The reduced metals diffuse outward through the hot oxide layer and at the same time oxygen diffuses inward. In spite of some recombination within the oxide, some metal reaches the outer surface and adheres thereto, perhaps in little islands, as a monatomic film. The greater part of the electron emission takes place from this adhered film, although some electron emission may come from free metal within the coating, or even from free metal on the surface of the base metal. The forces of adhesion reduce evaporation of the metals. Perhaps an under layer of adhered oxygen may play a part in reducing the work function. Presumably the metals evaporate and need replenishing, which is accomplished largely by continuous electrolysis of the coating by the space current and diffusion outward of the reduced metal.

With so complex an action and structure, it is not surprising that the total emission of a coated filament changes with time and depends in a complicated way upon its previous history. These cathodes show no marked saturation and this might be expected if some of the emission comes from the deeper layers, being drawn out by intense fields.

56. Emission Constants of Oxide-coated Cathodes.—The work function of the various oxides used for oxide-coated filaments has been measured by many scientists, beginning with Wehnelt. There is very little agreement among the results of the various experimenters, but there is a marked progressive decrease in magnitude from Wehnelt in 1904 to Espe in 1926. Detels[13] showed that both Φ_0 and A decrease as activation proceeds. This progressive change, together with the fact that

oxide-coated cathodes show no marked saturation, makes the determination of the emission constants very difficult, and it is not surprising that there should be a wide variation in values. Espe[12] offers as the most probable values of Φ_0 those given in Table VII. As is seen from the table, BaO has the lowest electron affinity. The value of the constant A varies greatly

TABLE VII

Oxide	Φ_0, volts	A,* amp./cm.^{2} deg.^{2}
CaO	1.77	1.29×10^2 or 2.49×10^2
SrO	1.27	4.07 or 2.58×10^2
BaO	0.99	2.88 or 2.72×10^2
CaO + SrO + BaO	1.24	8.3×10^{-3}

* From "International Critical Tables," McGraw-Hill Book Company, Inc., New York.

for different filaments and as measured by different experimenters, but is far from the value found for pure metal emitters.

57. Preparation of Oxide-coated Cathodes.—There are various ways of manufacturing oxide-coated cathodes. The earliest method of applying the coating is known as the *candle method* and is described by Arnold.[21] The salt of the alkali-earth metal, usually the carbonate, is ground to a very fine powder in melted paraffin or resin and then the suspension is cast in the form of a stick resembling a candle, from which the process derives its name. Sometimes the salts of two or more of the alkali-earth metals are mixed or separate candles for the separate salts may be used. The candle is applied to the core metal while hot and the paraffin or wax carrier allowed to evaporate. Several layers are applied and may be alternated if separate salts in separate candles are used. The cathode is then baked at about 1200°C. for an hour or two. During this baking, the salts are broken down to oxides which to some extent combine with the core metal forming dark-blue or black compounds, and the so-called "combined" kind of cathode results. In the early tubes, the core was made by rolling a wire of platinum, or an alloy of 10 per cent of iridium in platinum, to form a ribbon having larger surface than a round wire. It was later discovered that cathodes give emission at lower temperatures if the core is

[21] ARNOLD, *Phys. Rev.*, **16**, 70 (1920).

made of an alloy of 5 to 10 per cent of nickel in platinum. Usually this core is slightly oxidized before the coating is applied.

The uncombined kind of oxide-coated filament is used in practically all of the commercial radio tubes of today. The most common method of manufacturing this kind of emitter is to pass the ribbon-shaped core, made of nickel or nickel alloy, over grooved wheels, which direct the ribbon successively through a trough containing a water suspension of finely ground barium and strontium carbonates, then through a drying oven in which an atmosphere of CO_2 is maintained, then through another coating trough and drying oven, and so on for a number of such passes. The finished cathode is pure white in color because the coating consists only of the dried carbonates. In the preparation of the water suspensions of the carbonates of barium and strontium, it is very essential that the salts be pure and that they be ground to very fine powder. McNabb[18] finds that the use of a colloidal suspension of the carbonates results in greater emission, probably due to the finer particles. The colloidal carbonates are flocculated at a lower temperature than that which causes decomposition of the carbonates.

If the coating applied by the candle method is not strongly heated in air, the binder is evaporated and an uncombined coating of carbonates is left, which is not inferior to the combined kind in emitting efficiency. Except perhaps as to the mechanical structure and adherence to the core, it matters little how the uncombined coating is applied.

Superior filaments can be made by evaporating the metals barium and strontium in a vacuum so that some of the metal condenses on a preoxidized filament, usually of tungsten, which during the formation process is maintained at a red heat. The condensed barium and strontium displace the oxygen of the oxidized base metal, forming a thin coating of barium and strontium oxides. The vapor of the alkali-earth metals can be produced in the vacuum tube by heating a plate on which is attached a small piece of the metal, or more conveniently by gradually heating the azide of the metal[22] until it is decomposed and finally evaporated. This heating must be done cautiously at a temperature of about 150°C., as the azide readily explodes

[22] German patent 443,323 (1923).

at higher temperatures. Many of the coated cathodes used in Europe are made by the azide process, the azide often being painted on the plate from which it is evaporated when the plate is heated by eddy currents.

The coating of a metal base is only a part of the process of the production of an active cathode. The cathode is mounted in the tube, and the tube exhausted and baked in an oven at a temperature near the softening point of the glass in order to drive off the gas occluded on the walls of the tube. For soft glass, the baking temperature is from 350 to 420°C., and for hard glass from 500 to 700°C. The gases in the metal parts of the tube are driven off by heating these parts by means of high-frequency currents induced in them by an external coil, or later in the process by electronic bombardment. The cathode is then slowly heated to from 1400 to 1600°C., which serves to drive off occluded gases from the combined kind of cathode, and to convert the carbonates to the oxides for the uncombined kind. In the latter case, a large amount of CO_2 is liberated and is drawn off by the pump. Heating the cathode for a short time to a higher temperature of about 1750° to 1800°C. helps considerably in the formation process. The cathode, if of an uncombined kind, now may be active but further activation or *breakdown* is usually necessary. This is accomplished by impressing from 100 to 200 volts on the plate through a protective resistance, usually a tungsten lamp of a size which is lighted to nearly full brilliance by the normal emission current from the finished cathode. The cathode may require momentary heating to a high temperature to start the emission. Bombardment of the cathode by positive ions of gas and electrolysis of the coating take place. The emission current builds up usually quite rapidly and the filament temperature may then be reduced. If the coating contains much combined salts or hydroxides, the activation process may take considerable time. Becker[14] suggests that, after considerable emission has been established, the temperature and voltage should be adjusted until the emission is just limited by space charge. The tungsten lamp, commonly placed in series with the vacuum tube whose cathode is being activated, automatically reduces the voltage applied to the plate as the emission increases. This reduction of plate voltage during activation may be very desirable, as it

was shown by Hull and Winters[23] that positive ions driven against a cathode by a potential difference greater than 20 or 25 volts cause disintegration of the active-metal deposit. There are many variations of the process of activation some of which are described by Arnold,[21] McNabb,[18] Wagner,[16] etc.

58. Electrical and Radiation Characteristics of Oxide-coated Cathodes.—The emission of electrons from the coated filament has been shown to follow the emission equation, Eq. (23), page 59, Chap. IV. The most important characteristic of oxide-coated filaments is the low temperature at which copious emission is obtained. The normal operating temperature is about 1000°C., which may be roughly described as a cherry red, although with some coated cathodes considerable emission is obtained at temperatures so low that no visible light is emitted. The dark-colored combined kind of cathodes are much better radiators of heat and hence often require from two to three times the amount of power to maintain them at the same temperature as the white uncombined kind of exactly the same dimensions. Even some of the uncombined cathodes which have oxidized cores require more power, owing to the better radiating effect of the underlying oxidized surface. The emission per watt of heating power therefore depends upon the radiation coefficient as well as upon the activity of the coating.

Davisson[24] gives the following expression for the power radiated per square centimeter from the combined kind of oxide-coated cathode.

$$\left.\begin{aligned} W &= 5.735 \times 10^{-12}(0.4 + 2.5 \cdot 10^{-4}T)T^4 \frac{\text{watts}}{\text{cm.}^2} \\ &= 3.73 \frac{\text{watts}}{\text{cm.}^2}, \text{ at } T = 1000°\text{K.} \end{aligned}\right\} \tag{80}$$

Dushman[25] gives data, calculated from Eq. (80), for the power radiated per square centimeter from the combined kind of cathode used by the Western Electric Company, and also from cathodes made with Konel metal as core. The data for the latter kind of cathode Dushman obtained from results given by Lowry.[15] These data are given in Table VIII. The ratio

[23] HULL and WINTERS, *Phys. Rev.*, **21**, 211 (1923).

[24] DAVISSON, "International Critical Tables," vol. VI, p. 53, McGraw-Hill Book Company, Inc., New York, 1929.

[25] DUSHMAN, *Rev. Modern Phys.*, **2**, 381 (1930).

of hot to cold resistance of these filaments is also given in Table VIII.

TABLE VIII.—RESISTIVITY AND POWER RADIATED FROM COATED FILAMENTS OF BaO + SrO

T, °K.	Power radiated, W, watts per square centimeter		$\frac{\text{Hot resistance}}{\text{Cold resistance}}$	
	W.E. or combined-type cathode	Konel	W.E. or combined-type cathode	Konel
300	0.022	....	1.00	1.00
700	0.79	1.66	1.80	1.11
800	1.41	3.01	1.96	1.13
900	2.35	4.72	2.12	1.16
1000	3.73	6.57	2.26	1.19
1100	5.67	8.91	2.40	1.21
1200	10.70	....	2.53	1.24
1300	15.15	....	2.65	1.27
1400	20.93	....	2.75	1.29
1500	28.31	....	2.85	1.32

The emission efficiency is discussed later and the values for oxide-coated cathodes are given in Fig. 44, page 125.

The coated cathodes have one outstanding weakness, namely, the development of so-called "hot-spots," evidenced by intense local heating at one or more points on the cathode, by the flying off of hot particles of oxide coating, and usually by the melting and burning out of filamentary cathodes. These hot-spots are much more apt to form when high plate voltages are used, particularly if the space current is limited by underheating of the cathode. Thick oxide coatings are much more subject to this difficulty. The reason for the formation of these hot-spots, as given by Arnold,[21] is as follows: When the temperature is low, the resistance of the coating is high and a thick portion with a higher resistance is more heated by the space current than a thin portion. This higher temperature increases the emission, which then causes more heat. Thus the temperature builds up to the vaporizing point of the oxide and the melting point of the core metal. Such local increase in current does not usually take place if the space current is limited by space charge. This weakness of oxide-coated cathodes has prevented up to the present time their use in high-voltage high-vacuum tubes.

The useful life of an oxide-coated filament is terminated either by burn-out of the filament or by loss in emission due to operation at too high a temperature. If the coating is sintered by excessive temperature, the emission cannot be restored in full by any process of activation. The formation of hot-spots or the flaking off of the coating from any cause may so much reduce the coating that the cathode becomes useless. Some oxide-coated cathodes have operated for thousands of hours without appreciable decrease in emission.

III. METAL FILAMENTS HAVING AN ADSORBED MONATOMIC FILM OF AN ELECTROPOSITIVE METAL

59. Discovery and Mechanism of Emission of Adsorbed Monatomic Films.—Langmuir[26] discovered in 1914 that thoria, which is often admixed with the tungsten in the manufacture of tungsten filaments to reduce the recrystallization of the filaments, can under proper conditions increase tremendously the thermionic emission of the filament. Further investigation[27] showed that metallic thorium, produced by the reduction of the thoria at high temperatures, gradually diffuses to the surface of the filament under proper conditions of temperature and vacuum and forms there an active layer one molecule thick. Thorium has a low electron affinity and emits electrons abundantly at low temperatures, yet the work function of a monatomic layer of thorium on tungsten is lower than that of pure thorium and the emission at a given temperature is greater than for a pure thorium wire. Furthermore, the so-called *thoriated filament* can be operated at a much higher temperature than can a pure filament of thorium, owing to the strong force of adhesion which retards the evaporation of the monatomic layer.

The high emission of a thoriated wire can be explained in part by the high electric field that results from the electrical double layer. This double layer is due to the adsorbed layer of electropositive thorium atoms, many of which, as Becker[28] has shown, are adsorbed as positive ions. The intense electric field in the surface layer is in a direction to assist an electron to escape,

[26] LANGMUIR and ROGERS, *Phys. Rev.*, **4**, 544 (1914).

[27] LANGMUIR, *Phys. Rev.*, **22**, 357 (1923); U. S. patents, 1,244,216, 1,244,217 (1917).

[28] BECKER, *Bell. Tel. Lab. Reprint B*-412, or *Trans. Am. Electrochem. Soc.*, **55**, 153 (1929).

and the force function near the surface is very different from that near a surface of pure metal.

If oxygen, which is electronegative, is adsorbed on a surface of tungsten, the work function is greatly increased owing to the retarding field of the adsorbed negative ions of oxygen. If, however, a monatomic layer of barium is adsorbed on the layer of adsorbed oxygen, the intense field between the negative oxygen layer and the positive barium layer is in a direction to assist the escape of an electron. This field may extend many atom diameters. The work function of such a composite surface is less than that of a monatomic layer of barium on tungsten, and the emission is many times greater at the same temperature.

60. Activation of Thoriated Tungsten Filaments.—In the manufacture of a thoriated cathode, a few per cent (1 to 2 per cent) of thorium oxide is added to the tungsten powder before the sintering process. In the finished wire there may be left only about 0.7 per cent of thorium oxide. This wire is mounted in a vacuum tube, the tube is exhausted, and the glass and all metal parts freed of occluded gas by baking in an oven and heating the electrodes by means of an induction furnace. Since the thorium film is chemically very active, oxygen and water vapor, if present, attack and destroy the film. Furthermore, positive-ion bombardment, resulting from the ionization of residual gas, quickly destroys the sensitive thorium film. For these reasons it is more important in this case than in the case of other types of emitters to procure the best vacuum possible. Usually it is necessary to vaporize in the tube some active chemical substance such as magnesium, calcium, phosphorus, or some one of the several other suitable chemicals, which absorbs the residual gas and produces and maintains a high vacuum. These substances are popularly known as "getters" and their use dates from 1884 when Malignani[29] patented the use of "vaporizable reagents" to take up residual gases in an incandescent lamp. He used arsenic and iodine as vaporizable reagents. Soddy,[30] in 1906, vaporized calcium to clean up the gases in a vacuum chamber and obtained thereby a very high vacuum.

When the best vacuum procurable has been obtained, the tungsten cathode is heated almost to its melting point, to about

[29] MALIGNANI, U. S. patent 537,693 (1884).

[30] SODDY, U. S. patent 859,021 (1906).

2500°C., for from ½ to 2 min. This high temperature does two things. First, the surface is cleaned owing to the vaporization of any oxide film, and, second, some of the thoria inside the metal of the cathode is reduced to metallic thorium. Since all thorium which might have existed on the surface is evaporated at this high temperature, the cathode immediately after this temperature treatment possesses only the emission of pure tungsten.

The activation process consists in heating the cathode to a temperature within the range of from 1800 to 2000°C. Within this temperature range the thorium, which was reduced by the intense heating, diffuses rapidly outward and a monatomic layer of thorium on clean tungsten builds up. The thorium apparently moves between the tungsten crystals rather than through the crystals, so that the finer the crystals the more readily does this diffusion process progress.

The work function of the cathode is reduced from that of tungsten by an amount dependent on the fraction θ of the surface covered by the film of thorium. Langmuir[27] assumed that the work function for any value of θ is a linear function of θ as given by the equation

$$b_\theta = b_w + C\theta \qquad (81)$$

where b_θ is the equivalent value of b_0 of Eq. (23), b_w is the value of b_0 for pure tungsten, and C is a constant and equal to $b_{th} - b_w$, b_{th} being the value of b_θ for $\theta = 1$. Figure 37 is a plot of values of ϕ_0, *i.e.*, $\frac{b_{\theta K}}{e}$, obtained by Dushman and Ewald[31] for thorium on tungsten. These values were obtained from emission data, corrected for the Schottky effect by Eq. (74), page 88, in order to obtain the emission for zero field outside the emitter. This correction as given by Eq. (74) is undoubtedly inaccurate, as pointed out by Nottingham.[32] This is due to the fact that the force field assumed by Schottky (see Chap. IV) at the surface of a pure metal is, as Becker has shown, very different near a surface

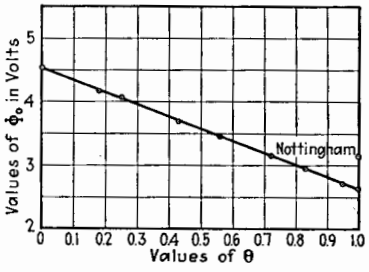


FIG. 37.—Φ_0 vs. θ for thorium on tungsten. (*Dushman and Ewald.*)

[31] DUSHMAN and EWALD, *Phys. Rev.*, **29**, 857 (1927).

[32] NOTTINGHAM, *Phys. Rev.*, **36**, 386 (1930).

covered by a monatomic film of electropositive atoms. Nottingham deduces that, for $\theta = 1$, b_0 is 36,500 or Φ_0 is 3.15 volts.

The value of the constant A of the emission equation, Eq. (23), is also affected by the film of thorium, as shown by the experiments of Dushman and Ewald[31] and by those of Kingdon.[33] In Fig. 38, the observation points marked by a circle were obtained by Dush-

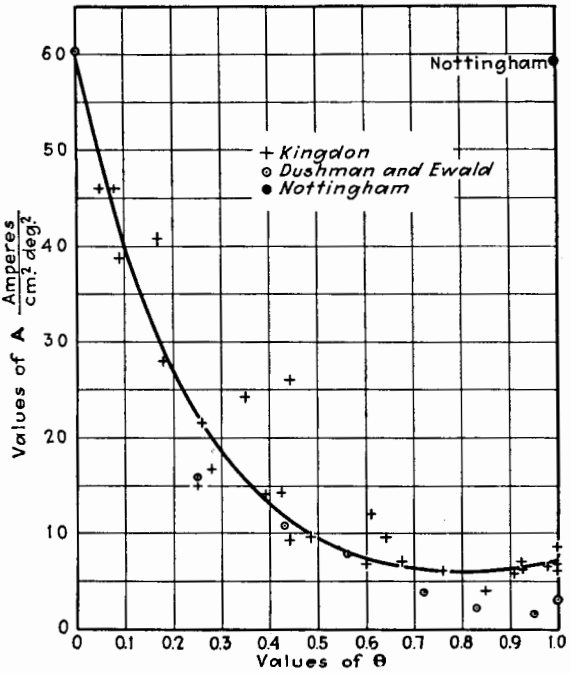


FIG. 38.—A vs. θ for a monatomic film of thorium on tungsten.

man and Ewald, while the observation points marked by a cross were obtained by Kingdon. The curve represents the following empirical equation, given by Kingdon:

$$A_\theta = (a_1^\theta + a_2^{1-\theta} - 1)A_0 \tag{82}$$

In Eq. (82), A_0 has the numerical value of unity, but has the dimensions of amp./cm.^{2} deg.^{2} The constant a_1 has the numerical value A for a surface of tungsten totally covered with a monatomic film of thorium. The numerical value of a_2 for a tungsten surface

[33] KINGDON, *Phys. Rev.*, **24**, 510 (1924).

is 60.2. The same criticism applies here to the correction as given by Eq. (74) for the Schottky effect. Nottingham,[32] using a different method for the calculation of A for zero field, obtains the value of 59 when $\theta = 1$. This value of A is shown by the solid circle in Fig. 38.

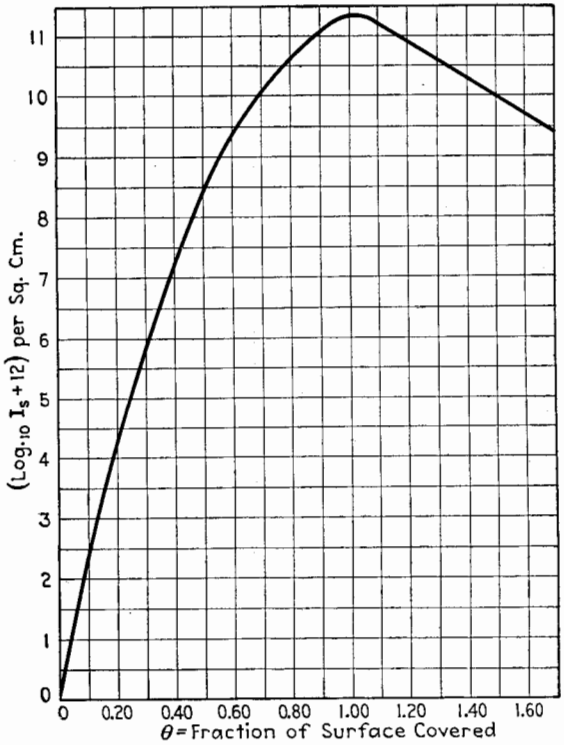


FIG. 39.—Log I_s vs. θ for a film of barium on tungsten. (*Becker.*)

The value of θ may be obtained from experimental data by use of Eq. (81).

$$\theta = \frac{b_\theta - b_w}{b_{th} - b_w} \tag{83}$$

Since $\log A_\theta$ varies in such a way as to be nearly linearly related to b_θ, as shown by experiments of Kingdon,[33] a more convenient form of Eq. (83) to be used in determining θ is

$$\theta = \frac{\log I_\theta - \log I_0}{\log I_1 - \log I_0} \tag{84}$$

where I_θ is the saturation current for any value of θ, I_0 is the saturation current for $\theta = 0$, and I_1 is the saturation current for $\theta = 1$, all at the same temperature.

Becker[28] claims that when the film of adsorbed metal becomes more than one atom thick, the emission decreases, as shown by Fig. 39 taken from Becker's paper. Langmuir has recently expressed the opinion that it is impossible, under the conditions of these experiments, to build a layer of alkali metal more than one atom thick because of the rapid induced evaporation. If this is true, then θ is less than unity at maximum activity.

Langmuir has shown that, at constant temperature, θ increases with time according to curve A, Fig. 40a. This curve is plotted

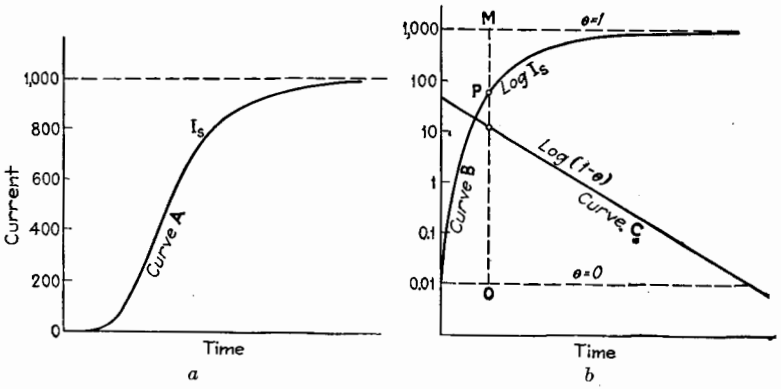


FIG. 40.—Rate of activation of thoriated-tungsten filaments. (*Langmuir.*)

logarithmically in curve B, Fig. 40b. From Eq. (83) the value of θ is OP/OM. According to Langmuir the value of θ may be expressed by the equation

$$\theta = 1 - \epsilon^{-Kt} \tag{85}$$

or

$$\log_\epsilon (1 - \theta) = -Kt$$

The plot of log $(1 - \theta)$ against the time t is a straight line, as shown by curve C, Fig. 40b, the slope of which is the value of the constant K, which measures the rate of activation. The reciprocal of K is known as the *activation time* T_a.

Figure 41, from Langmuir's paper,[27] shows the activation curves for a number of temperatures. The higher the temperature, the more rapidly the thorium diffuses outward to form the film. Evaporation of thorium from the film also takes place

and is more rapid the higher the temperature. If thorium atoms pile on top of each other to form a film more than one atom thick, evaporation of the outer atoms takes place more readily because the adhesive forces are less. This tends to prevent the formation of a layer thicker than a monatomic layer.

During the activation process, thorium is rapidly diffusing from the inside and being evaporated from the film. The life

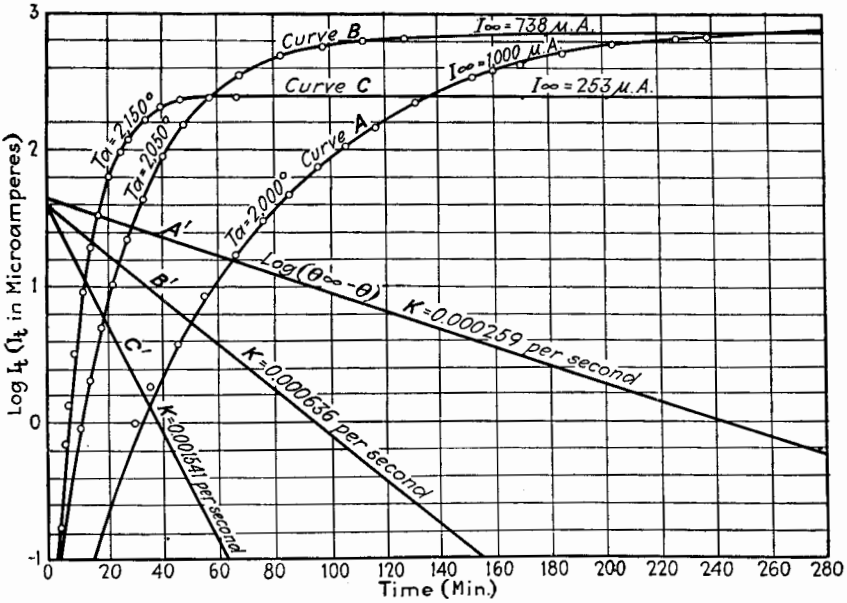


FIG. 41.—Rate of activation of a thoriated-tungsten filament at temperature T_a after flashing at 2800° K. (*Langmuir.*)

of a thoriated filament is limited by the supply of thorium in the wire and hence the life is greater the lower the temperature. Furthermore, the operating temperature must be chosen so that diffusion outward is sufficiently rapid to compensate for the evaporation of the thorium from the surface. The best operating temperature is about 1900°K., although it depends somewhat upon the size and thorium content of the wire. Table IX gives the equilibrium values of emission current, life, and activation time T_a, at different temperatures, for a tungsten filament containing 1 per cent of thoria.

A thoriated cathode which has lost its emission can often be reactivated, provided the vacuum is still high and the thorium

TABLE IX.—EQUILIBRIUM CHARACTERISTICS AT VARIOUS TEMPERATURES OF A TUNGSTEN FILAMENT CONTAINING 1 PER CENT OF THORIA[1]

T, °K.	θ	I, amperes per square centimeter	Life of thoria, hours	Activation time T_a seconds
1300	0.99997	$4.14 \cdot 10^{-4}$		
1400	0.99975	$3.12 \cdot 10^{-3}$	.....	$1.43 \cdot 10^{11}$
1500	0.99878	0.0179	.....	$5.23 \cdot 10^{9}$
1600	0.99528	0.0812	.....	$2.88 \cdot 10^{8}$
1700	0.9848	0.287	.....	$2.15 \cdot 10^{7}$
1800	0.9605	0.772	720,000	$2.02 \cdot 10^{6}$
1900	0.9191	1.59	94,000	$2.18 \cdot 10^{5}$
2000	0.8713	2.89	15,100	$2.48 \cdot 10^{4}$
2100	0.781	3.43	2,897	$9.48 \cdot 10^{3}$
2200	0.551	1.24	643	$2.10 \cdot 10^{3}$
2300	0.139	0.114	164	531
2400	0.0601	0.168	47	7.0
2500	0.0355	0.357	14.6	1.28
2600	0.0207	0.774	5.01	0.268
2800	0.0088	3.48	0.74	0.0161
3000	0.0041	13.5	0.14	0.0014

[1] "Handbuch der Experimental Physik," Wien and Harms, Akademische Verlagsgesellschaft M.B.H., Leipzig, 1928.

content of the metal has not been exhausted. The process of reactivation consists in flashing the filament for 10 to 20 sec. at a high temperature by impressing across the filament about three and one-half times normal operating voltage. The cathode is then operated at about one and one-half times normal operating voltage, with no plate voltage, for from 1 to 2 hr. If this treatment does not restore emission, the thorium content of the metal has been exhausted.

A schedule showing the marked physical changes of a thoriated tungsten cathode within various temperature ranges, as given by Schottky and Rothe,[34] is given below. The figures in parentheses are given by Dushman.

$T < 1900$°K.	*Operating Temperatures*
(1800)	Reduction, evaporation, and diffusion practically absent. Stable condition.

[34] "Handbuch der Experimental Physik," Wien and Harms, p. 169.

	Activation Temperatures
T = 1900–2100°K. (1800–2000)	Practically, only diffusion and induced evaporation exist. $\theta = 1$. Activation velocity increases with temperature.
T = 2100–2400°K. (2000–2300)	Diffusion and spontaneous evaporation exist. $\theta < 1$. High activation velocity.
	Deactivation Temperatures
T = 2400–2600°K. (2300–2600)	Spontaneous evaporation exceeds diffusion. $\theta < 1$. Only small amount of reduction.
	Reduction Temperatures
$T > 2600$°K.	Intense reduction, diffusion, and evaporation.

All that has been said regarding thoriated wire or cathodes is also true, in principle, for base metals other than tungsten, as tantalum and molybdenum, and also for monatomic films of electropositive metals, such as barium, caesium, and lanthanum.[35]

The emission characteristics for monatomic films of thorium and caesium on tungsten, as given by Langmuir,[36] are shown in Fig. 42. Langmuir and Kingdon[36] have shown that, if tungsten is first slightly oxidized and then covered with a monatomic film of caesium, the emission current is much greater than from a film of thorium directly on tungsten, as shown by the curve of Fig. 42 and also by Fig. 13, page 62.

The emission efficiencies of filaments with monatomic films are given in Fig. 45 and are discussed in the next section.

61. Emission Efficiency.—The practical value of a source of electrons for use in vacuum tubes depends largely upon its emission efficiency. The thermionic emitter requires heat to maintain its temperature, and the necessary heat energy is almost invariably supplied electrically. If P is the electrical power supplied to heat the emitter, there must be some functional relation between the total emission current I_s and the total power P, or

$$I_s = F(P) \tag{86}$$

[35] BECKER, *Phys. Rev.*, **28,** 341 (1926); **34,** 1323 (1929); *Trans. Am. Electrochem. Soc.*, **55,** 153 (1929); DUSHMAN, DENNISON, and REYNOLDS, *Phys. Rev.*, **29,** 903 (1927); DUSHMAN and EWALD, *Phys. Rev.*, **29,** 857 (1927); KINGDON, *Phys. Rev.*, **24,** 510 (1924).

[36] LANGMUIR, *Ind. Eng. Chem.*, **22,** 390 (1930).

It has been found experimentally that, for the operating range of temperature, Eq. (86) can be expressed with fair accuracy in the form

$$I_s = CP^n \tag{87}$$

where n is a constant dependent upon the kind of emitter, C is a different constant for each particular filament because, as will be seen, C is dependent upon the surface area of the emitter. The degree to which Eq. (87) represents the experimental results over the operating range from 2400 to 2700°K. can be

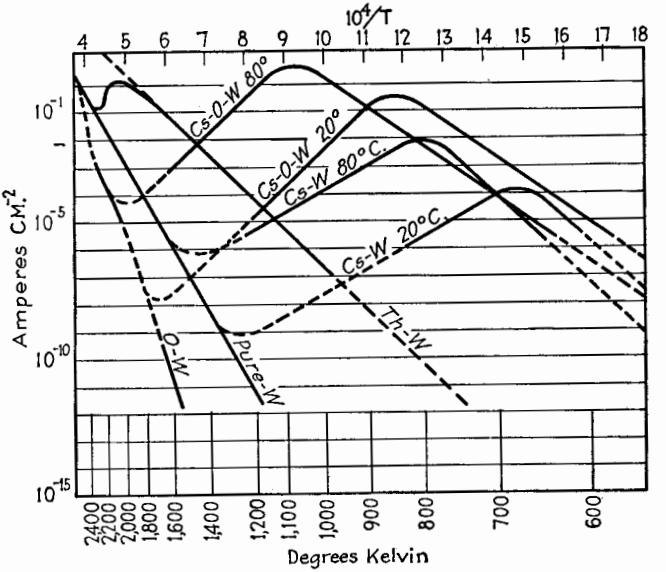


FIG. 42.—Emission characteristics of various monatomic films on tungsten. (*Langmuir.*)

seen by examining Fig. 43, which shows the logarithm of emission current per square centimeter plotted against the logarithm of the power in watts per square centimeter for tungsten, taken from data by Jones and Langmuir.[37] For tungsten, the exponent n obtained from Fig. 43 is 4.87.

The plot of emission current against filament heating power is only approximately represented by Eq. (87), as is shown by the deviation of the points in Fig. 43 from the straight line, both above and below the operating range. Davisson, by slightly

[37] JONES and LANGMUIR, *Gen. Elec. Rev.*, **30**, 312, 354 (1927).

distorting the coordinates of the double logarithmic paper, obtained a chart on which the emission-power curve plots as a straight line, provided the radiation from the filament follows the Stefan-Boltzmann law. Figure 44 shows the emission of tungsten, from data by Jones and Langmuir given in Table VI, plotted on this special paper. There is some deviation from a

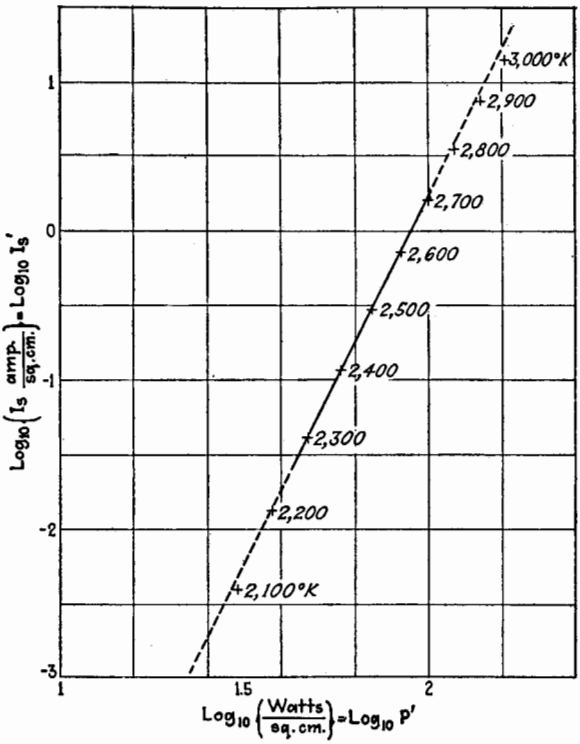


FIG. 43.—Emission-power curve for tungsten.

straight line at the upper part of the graph. Figure 44 also shows a line for thoriated tungsten, plotted from data given by Dushman and Ewald,[38] and two lines for oxide-coated platinum-nickel alloy, from data given by King.[39] Since it is often unsafe to use the normal heating power in obtaining data for full emission, the advantage of a chart of the type shown in Fig. 44

[38] DUSHMAN and EWALD, *Gen. Elec. Rev.*, **26**, 154 (1923).
[39] KING, *Bell System Tech. J.*, **2**, No. 4 (1923).

is that the normal emission can be obtained by extending the straight-line plot through data points obtained for low power.

In Eqs. (86) and (87), the emission current I_s is proportional to the surface area A of the emitter. Also, since the power P

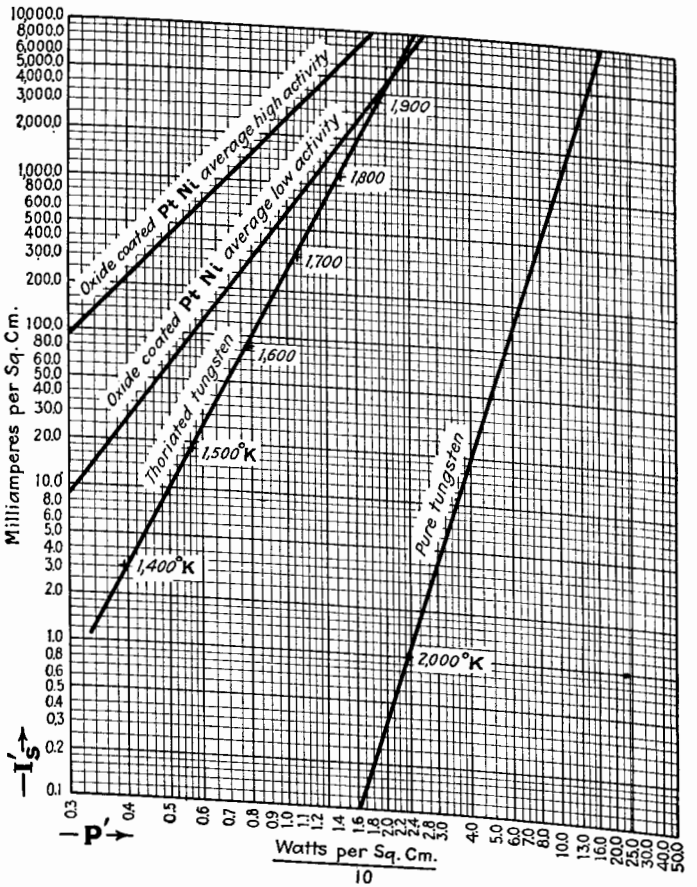


FIG. 44.—Emission-power chart.

is radiated from the surface of the filament, it too must be proportional to the surface area A. Dividing Eq. (87) by A gives

$$\frac{I_s}{A} = CA^{n-1}\left(\frac{P}{A}\right)^n \tag{88}$$

or

$$I'_s = C'P'^n \tag{89}$$

where I'_s is the emission current per square centimeter, P' is the power per square centimeter, and C' is a constant applying to 1 sq. cm. and hence is dependent only upon the material of the emitter. It is apparent that

$$C' = CA^{n-1} \qquad (90)$$

If we divide Eq. (89) by P', we have the emission efficiency expressed as emission current per square centimeter per watt used in heating. This is denoted by S and from Eq. (89),

$$S = \frac{I'_s}{P'} = C'P'^{(n-1)} \qquad (91)$$

Expressed for the total filament, Eq. (91) becomes

$$S = \frac{I'_s A}{P'A} = \frac{I_s}{P} = CP^{n-1} \qquad (92)$$

Obviously, from Eqs. (91) and (92), S plotted against P or P' on the special-coordinate chart of Fig. 44 is a straight line, and is approximately a straight line on ordinary logarithmic paper. Of more practical interest is the value of S under normal operating conditions chosen to give a life of the order of 2,000 hr., since this value gives an approximate comparison of the relative emission efficiencies of the three types of emitters. Values of S thus obtained for a number of tubes for each type of emitter are plotted in Fig. 45. A wide variation would naturally be expected in the value of S for both the oxide-coated and thoriated filaments because of different physical conditions and different degrees of activation which are inevitably met with in commercial tubes. Under operating conditions, it is evident that on the average the emission efficiency of the thoriated filament is about ten times that of tungsten, and the efficiency of the oxide-coated filament

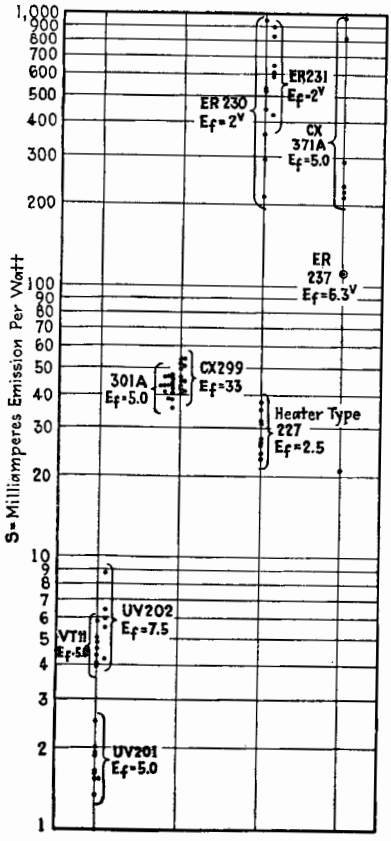


FIG. 45.—Emission efficiency of commercial tubes.

about one hundred times that of tungsten. It is to be noted that the upper group of observations for oxide-coated emitters is for filamentary cathodes, while the lower group is for separately heated cathodes. The latter group would be expected to show a lower efficiency than the former group.

62. Cooling of Ends of Filament.—In an actual tube the whole of the filament is not at the same temperature, owing to the conduction of heat away from the filament by the end connections and supporting wires. The most marked effect produced by this variation in temperature is a reduction in the emission, since those parts of the filament which are cooled are less active in emitting electrons. The effect is more pronounced for a short filament, and the emission efficiency of such filaments is reduced by a considerable amount.

Another effect of the cooling of the ends of the filament is to round off the knee of the curve of emission current plotted against plate voltage (see Fig. 27, page 84). The reason for this may be made clear by considering that the resultant filament-saturation curve is made up of the sum of many curves, each being for an infinitesimal length of the filament and for a certain temperature, the temperatures for most of the elemental curves being different.

63. Voltage Drop along Filament.—If the emitter is also the heating element, there is a voltage drop along the filament of from 2 to 5 volts for receiving tubes, according to the design. The voltage drop along the filament of a power tube is usually much greater and may be as much as 50 volts. This potential gradient along the filament has several effects. The difference of potential between the positive plate and the negative end of the filament is greater than that to the positive end of the filament. Consequently, when the space current is limited by space charge, the space current is greater from the negative end than from the positive end of the filament. If the plate voltage is low, the filament drop may be an appreciable fraction of the plate voltage, so that the unequal distribution of space current along the filament is greatly accentuated. On the other hand, in transmitting tubes, this inequality in distribution of the space current is usually negligible because of the large plate voltage in comparison with the filament voltage.

Another effect of the voltage drop along the filament is to round off the knee of the space-charge-saturation curves of Fig. 28. The reason for this is similar to that already given for the

rounding of the knee of the temperature-saturation curves due to a temperature gradient along the filament. In this case the resultant curve of space current plotted against temperature can be thought of as the sum of a large number of curves, each for a small length of filament and a particular plate voltage, the plate voltage being different for the various elemental lengths of filament.

The third effect of the potential gradient along the filament is to alter the shape of the space-current curve plotted against plate voltage at its lower bend, changing it from the three-halves-power law to one of another exponent which Langmuir has shown theoretically to be five halves. This five-halves-power law can easily be derived by simple integration of the three-halves-power law along the filament. Van der Bijl,[40] in his book, states that actually the lower bend of the curve of receiving tubes often follows quite accurately a square law, which is more convenient mathematically.

64. Effect of Space Current on Filament Current.—As yet no account has been taken of the fact that the space current entering all along the filament (electrons leaving) is superimposed upon the heating current, making the actual current in the filament different at different points along the filament. Since the heating at any point of the filament is proportional to the square of the total filament current at that point, the space current, if appreciable compared with the filament current, may cause a considerable variation of temperature along the filament. This, in turn, alters the emission and alters the resistance and consequently the potential gradient along the filament. All these various effects are mutually dependent and may conspire to accentuate the temperature difference between the two ends of the filament, even to the extent of causing one end to burn out. It is desirable, therefore, to study the effect of the superimposed space current upon the total filament current.

It is very difficult to take account of all the effects which are due to the superimposition of the space current on the filament current. The danger of burning out the filament is greatest in power tubes where the space current is usually a larger portion of the filament current than in receiving tubes. In power tubes, the plate voltage is usually so high that the voltage drop along the

[40] VAN DER BIJL, "The Thermionic Vacuum Tube and Its Applications," p. 64, McGraw-Hill Book Company, Inc., New York, 1920.

filament can be neglected in obtaining an approximate solution. Furthermore, since in the normal operation of a tube, space current is for most of the time limited by space charge, the inequality in emission due to the temperature gradient along the filament can be neglected. Consider a filament receiving space current *uniformly* along its length. In Fig. 46, let L be the length of the filament, and x be the distance from the positive end to any point of the filament. If I_s is the total space current, that portion of it flowing to a length dx anywhere along the filament is $dI_s = \frac{I_s}{L}dx$.

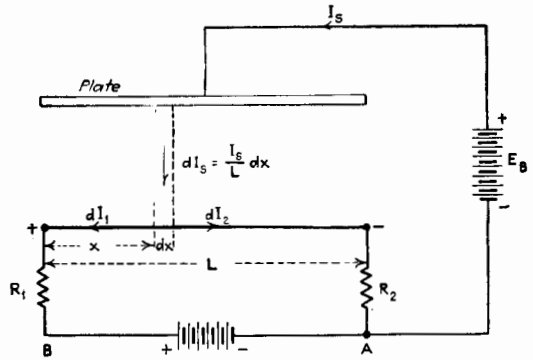


FIG. 46.—Space current in the filament.

Let R_1 and R_2 be two resistances in the heating circuit, which may represent controlling resistances. The heating battery is assumed to have negligible internal resistance, although the battery resistance may be considered as a part of R_1 if the plate circuit is connected to point A, or as part of R_2 if the plate circuit is connected to point B. The current dI_s divides at the filament into two portions whose magnitudes are inversely as the resistances of the two paths. Therefore, dI_1, that portion of dI_s going to the left, is related to dI_2, the portion going to the right, as

$$\frac{dI_1}{dI_2} = \frac{(L - x)r + R_2}{xr + R_1} \tag{93}$$

where r is the resistance of unit length of the filament. The e.m.f. of the heating battery makes no difference in this division of current; and if the battery has negligible internal resistance, it makes no difference whether the plate circuit is connected to point A or to point B.

From Eq. (93), the following expressions for dI_1 and dI_2 are easily obtained:

$$\left.\begin{aligned} dI_1 &= \frac{(L-x)r + R_2}{Lr + R_1 + R_2} \cdot \frac{I_s}{L} dx \\ &= \frac{(L-x)r + R_2}{R} \cdot \frac{I_s}{L} dx \end{aligned}\right\} \tag{94}$$

$$\left.\begin{aligned} dI_2 &= \frac{xr + R_1}{Lr + R_1 + R_2} \cdot \frac{I_s}{L} dx \\ &= \frac{xr + R_1}{R} \cdot \frac{I_s}{L} dx \end{aligned}\right\} \tag{95}$$

where R is the total resistance of the filament circuit. At any point x on the filament, that portion of the space current in the filament which is flowing toward the positive end is the sum of all the di_1's from x to L, and that portion flowing toward the negative end of the filament is the sum of all the di_2's from the portion of the filament from zero to x. At point x, the actual space current in the filament which is flowing toward the negative end is

$$\begin{aligned} I_{sx} &= \int_0^x dI_2 - \int_x^L dI_1 \\ &= \int_0^x \frac{xr + R_1}{R} \cdot \frac{I_s}{L} dx - \int_x^L \frac{(L-x)r + R_2}{R} \cdot \frac{I_s}{L} dx \\ &= I_s \left[\frac{x}{L} - \frac{\frac{Lr}{2} + R_2}{R} \right] \end{aligned} \tag{96}$$

The total current at any point is obtained by adding the filament current I_a supplied by the filament battery to the space current given by Eq. (96). Therefore, the total heating current is

$$I = I_a + I_s \left[\frac{x}{L} - \frac{\frac{Lr}{2} + R_2}{R} \right] \tag{97}$$

At the positive end of the filament the heating current is

$$I_+ = I_a - I_s \cdot \frac{\frac{R_f}{2} + R_2}{R} \tag{98}$$

where R_f is the resistance of the filament. At the negative end the heating current is

$$I_- = I_a + I_s \cdot \frac{\frac{R_f}{2} + R_1}{R} \tag{99}$$

Examining Eqs. (98) and (99), it is evident that the temperature of the positive end of the filament is reduced, while that of the negative end of the filament is elevated by the space current. In order that the increase in temperature of the negative end may be small, R_1 should be small. Increasing R_2, since it increases R, helps to reduce the excessive heating of the negative end of the filament. We may make a general statement that, *when the space current is appreciable compared with the filament current and there is danger of excessive heating of the negative end of the filament, any external resistance in the filament circuit should be connected between the negative end of the filament and the plate-circuit connection to the filament circuit.*

When alternating current is used to heat a filamentary emitter, the plate return is connected to a mid-point tap on the secondary winding of the transformer which supplies the heating current, or to the mid-point of a resistance connected directly across the filament. Assuming that the heating current I_a in Eqs. (97), (98), and (99) is a sinusoidal current, $i_a = \sqrt{2}I_a \sin \omega t$, the total instantaneous heating current at any point of the filament from Eq. (97), is

$$i = \sqrt{2}I_a \sin \omega t + I_s\left[\frac{x}{L} - \frac{\frac{R_f}{2} + R_2}{R}\right] \tag{100}$$

The square of the effective value of current given by Eq. (100) is

$$I_x^2 = I_a^2 + I_s^2\left[\frac{x}{L} - \frac{\frac{R_f}{2} + R_2}{R}\right]^2 \tag{101}$$

If the transformer mid-point is used, R_2 and R_1 are practically zero and the effective heating at any point of the filament is proportional to

$$I_x^2 = I_a^2 + I_s^2\left(\frac{x}{L} - \frac{1}{2}\right)^2 \tag{102}$$

Obviously, from Eq. (102), both ends of the filament are slightly overheated, but the amount is usually negligible. The heat capacity of the filament is usually great enough to smooth the temperature fluctuation due to the alternating current.

65. Equipotential Cathodes.—It has become general practice to heat cathodes by alternating current, thus obviating either the annoying frequent running down of the dry cell or the troublesome storage battery and charging outfit. When alternating current is used for heating a filamentary emitter, the filament is usually made short and heavy in order to reduce the alternating voltage across the filament and thus reduce the fluctuations in the plate current which cause hum in repeaters and receiving sets. A better way of eliminating any effects upon the grid and plate circuits due to the alternating heating currents is the use of equipotential cathodes or separately heated emitters. This type of cathode, now used in many of the commercial receiving tubes, usually consists of a cylinder of nickel or nickel alloy coated on the outside with an oxide-emitting coating and heated by a filament within, which carries the alternating heating current. Sometimes this heating filament is coiled and supported by insulating plugs inserted in the ends of the cylinder, or the filament may pass through two longitudinal holes in a long insulating piece that fits inside the cylinder. The former arrangement is quick heating, but is more apt to short-circuit than the latter arrangement which has the principal fault of requiring a minute or more to attain operating temperature owing to the poor heat conductivity of the insulating tube. With both schemes, emission from the heater to the grid and plate must be prevented by shielding, and emission from the heater to the cathode cylinder is usually prevented by an opposing field produced by making the heater positive in potential with respect to the cylindrical cathode.

The emission efficiency S of a separately heated cathode is generally less than that of a directly heated cathode of the same type, as shown in Fig. 45 for measurements on commercial (type 27) tubes.

66. Magnetic Effect of Current through Filament.—The current through the filament produces a magnetic field surrounding the filament and this field acts to deflect the emitted electrons in a direction perpendicular both to the field and to the direction of motion of the electrons. In practice, however, the electron

has very small velocity near the filament where the magnetic field is strongest, and hence the magnitude of the deflecting force is negligible in comparison with the electrostatic pull of the positively charged plate.

67. Effect of Gaseous Ionization on Plate Current.—In Chap. IV, some of the effects of gas upon the emission of a cathode were discussed. There it was pointed out that the saturation emission of tungsten is reduced by chemical reaction of the

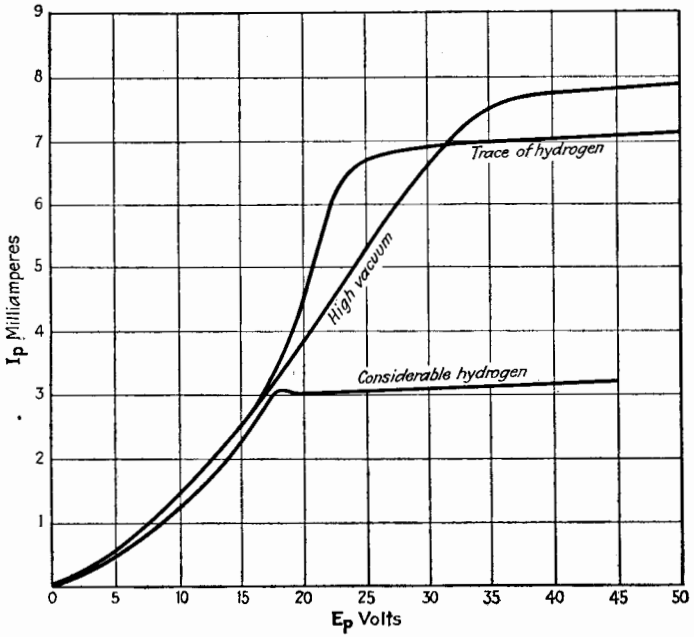


FIG. 47.—Effect of gas on the current through a diode.

tungsten with small amounts of oxygen, nitrogen, or water vapor. Oxygen and water vapor were also especially detrimental to the emission from oxide-coated and thoriated cathodes.

Besides its effect upon the emission, gas when ionized may greatly affect the plate current, if it is limited by space charge. This effect upon the plate current is due to the partial neutralization of the space charge by the positive ions. Because of the low velocity of the positive ions, a single ion can neutralize the electrostatic effect of hundreds or even thousands of electrons. Therefore, a slight trace of gas can have a marked effect upon the plate current, as shown in Fig. 47.

The curve marked "high vacuum" was taken when the bulb was connected to the pumps and the vacuum was as high as possible. The curve marked "trace of hydrogen" was taken after a very small amount of hydrogen had been admitted. Undoubtedly some oxygen and some water vapor were present which considerably reduced the saturation current. The filament current in this case was greater than that in taking the high-vacuum curve. The two curves are practically coincident up to about 17 volts, where ionization begins. The curve with gas rises rapidly for higher voltages, owing to the partial neutralization of space charge by the positive ions. The third curve marked "considerable hydrogen" was taken for the same filament current as for the high-vacuum curve and shows the effect of the gas upon the saturation current. It also shows the lower current before ionization due to frequent collisions of the electrons with the molecules of gas.

A very convenient method of testing for a trace of gas in a vacuum tube will be described in a later chapter.

68. Schottky Effect or "Schroteffekt."—Schottky,[41] in 1918, predicted theoretically that, since the electron stream in a thermionic vacuum tube consists of a statistical stream of discrete charges, there should be fortuitous fluctuations in the plate current, and hence disturbances should be set up in any tuned circuit connected in the plate circuit. Schottky called this phenomenon the *Schroteffekt*, or "small shot effect." That this small random fluctuation of the electron current does exist has been shown by various investigators,[42] who have measured the electrical variations in tuned circuits at various frequencies. The theory developed by Schottky and others[41,43] applies only to electron currents not limited by space charge. Such theory, although checked approximately by experiment, does not help in the calculation of the *Schroteffekt* in amplifiers because the effect is much less when the current is limited by space charge.

[41] SCHOTTKY, *Ann. d. Physik*, **57**, 541 (1918); **68**, 157 (1922).

[42] HULL and WILLIAMS, *Phys. Rev.*, **25**, 147 (1925); HARTMAN, *Ann. d. Physik*, **65**, 51 (1921); *Phys. Zeits.*, **23**, 436 (1922); JOHNSON, *Phys. Rev.*, **26**, 71 (1925); WILLIAMS and VINCENT, *Phys. Rev.*, **28**, 1250 (1926).

[43] JOHNSON, *Ann. d. Physik*, **67**, 154 (1922); FRY, *J. Franklin Inst.*, **199**, 203 (1925); BALLANTINE, *Radio Freq. Lab. Reprint* **7**, 1928, or *J. Franklin Inst.*, **206**, 159 (1928).

Llewellyn[44] has deduced the magnitude of the shot effect when the space current is limited by space charge. The expression obtained by Llewellyn for the mean-square voltage produced by the shot effect across an impedance Z in series with the filament-to-plate path, both with and without space charge, is

$$\bar{E}^2 = \left(\frac{\partial I}{\partial I_s}\right)\frac{eI_s}{\pi}\int_0^\infty (Z)_n^2 dn \tag{103}$$

In Eq. (103), e is the electronic charge, $(Z)_n$ is the impedance of Z at frequency n, I_s is the saturation current corresponding

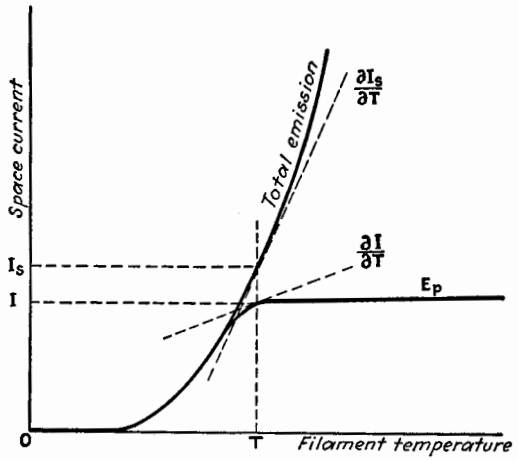


FIG. 48.—Space current vs. temperature of cathode.

to temperature T, and I is the space-charge-limited current. The meaning of I and I_s may be made clearer by referring to Fig. 48. The ratio $\partial I/\partial I_s$ can be written

$$\frac{\partial I}{\partial I_s} = \frac{\partial I/\partial T}{\partial I_s/\partial T} \tag{104}$$

Expression (104) is easily evaluated as indicated in Fig. 48. When the space current is not limited by space charge, the ratio given in Eq. (104) reduces to unity, and Eq. (103) reduces to the usual equation for no space charge.

If an amplifier giving a voltage ratio $(G)_n$ is used to increase the voltage across the impedance, Eq. (103) becomes

[44] LLEWELLYN, *Proc. I.R.E.*, **18**, 243 (1930).

$$\bar{E}^2 = \left(\frac{\partial I}{\partial I_s}\right)\frac{eI_s}{\pi}\int_0^\infty (Z)_n^2(G)_n^2 dn \qquad (105)$$

Equations (103) and (105) show that the shot effect is zero when the space-current curve is parallel to the temperature axis in Fig. 48. If a certain plate voltage is impressed on a vacuum tube and the filament current is increased, the shot effect increases with I_s until the ratio $\partial I/\partial I_s$ begins to depart appreciably from unity. Beyond this point the shot effect decreases and approaches zero.

Although the shot effect is practically zero when the space current is limited by space charge, other effects exist to give fluctuations in the space current and thus cause noise in high-gain amplifiers. Llewellyn[44] and Ballantine[45] have discussed these various effects. A brief reference is here made to one of the most important causes of fluctuation noises in amplifiers, namely, the *thermal effect*. The heat agitation of electrons in a conductor sets up across the conductor fluctuations in potential dependent upon the impedance of the conductor and its temperature. Schottky[41] first called attention to this cause of fluctuations in current in a conductor. Others[44,46] have verified experimentally the existence of such fluctuations, both in metal conductors and in the space between a filamentary emitter and the plate.

69. The Cold Electrodes.—Thus far in this chapter we have confined our attention to the cathode. A few remarks are now appended concerning the cold electrodes of the vacuum tube.

Certain metals[47] are better than others for use in vacuum tubes, the best metals in general being those that can easily be denuded of their gas content. The metals nickel, tungsten, molybdenum, and tantalum are commonly used. Nickel is inexpensive and is easily denuded of its occluded gases when heated to redness during the exhausting process. Tungsten, molybdenum, and tantalum have an advantage over nickel because of their higher melting points and are often used in high-power tubes in which the plates may operate normally at high temperatures. Tantalum absorbs enormous volumes of gas when heated and for this reason is very useful in

[45] BALLANTINE, *Radio Freq. Lab. Reprint* 21, 1930.

[46] JOHNSON, *Phys. Rev.*, **32**, 97 (1928); NYQUIST, *Phys. Rev.*, **32**, 110 (1928).

[47] The properties of some of the rare metals are well described in a booklet "Rare Metals" published by the Fansteel Products Company, Inc., North Chicago, Ill.

maintaining the vacuum in a tube. Often a small piece of tantalum is welded to the plate for the purpose of absorbing gases. Tantalum and molybdenum can be worked when cold, while tungsten must be worked when hot.

The energy possessed by the rapidly moving electrons constituting the space current is converted into heat when the electrons strike the plate. If the plate voltage is high, the amount of heat may be sufficient to raise the temperature of the plate to incandescence. This heat is dissipated almost entirely by radiation. The capacity of power tubes is limited by the rate at which the plate loses its heat. The power in watts radiated from an exposed surface of tungsten is given by Eq. (79). Table X gives the watts radiated per square centimeter calculated by Eq. (79), page 102, and the emission current calculated by Eq. (78) for certain values of the absolute temperature T.

TABLE X.—POWER RADIATED AND EMISSION CURRENT PER SQUARE CENTIMETER FROM TUNGSTEN

T, °K.	P, watts per square centimeter	I_s, milliamperes per square centimeter
1000	0.5701	1.054×10^{-12}
1200	1.663	9.42×10^{-9}
1400	3.898	6.56×10^{-6}
1600	7.89	5.82×10^{-4}
1800	14.37	4.56×10^{-2}
2000	24.01	1.008
2200	38.31	13.2
2400	57.78	114.3
2600	83.80	719

According to the data of Table X it is permissible to dissipate about 12 to 14 watts per square centimeter. A higher temperature would cause too great an emission from the plate.

It is customary to blacken the surfaces of the plates of some tubes by carbonizing or by oxidizing in order to increase the radiating efficiency of the plates. When the plates are thus treated, the dissipation may be several times greater than that given by Eq. (79).

CHAPTER VI

NOMENCLATURE AND LETTER SYMBOLS

Before developing further the theory of vacuum tubes, a suitable system of notations or letter symbols must be adopted in order to denote easily the many different quantities involved. Although several schemes of notation have been suggested, no one system as yet has come into general use. A satisfactory system of symbols is urgently needed because of the exceptionally large number of different components and aspects of the same quantity necessary to be distinguished in this field. For example, the potential between two points may consist of a steady value on which is superimposed a periodically changing value. We have then to deal with the steady value, average value, peak value, and maximum value of a harmonic component, total instantaneous value, instantaneous value measured from the average value and also from the steady value, etc. The symbols should be such as to suggest, as far as possible, the quantities represented and to burden the memory of the reader as little as possible. Furthermore, the symbols should be convenient for blackboard use, easy to write on a typewriter if possible, and, most important of all, the symbols should be easily set up in type so that the printing of theoretical papers is not excessively expensive. This last requirement has been the principal stumbling block to the acceptance of any system of special symbols, and yet it appears to be impossible to represent all of the necessary quantities by ordinary type unless awkward combinations of letters and subscripts are used. The author believes that free expression in this new and important field of vacuum tubes should not be hampered by shortsighted limitations placed upon the choice of symbols, but that special symbols should be accepted, as has been done in other branches of science, such as mathematics, astronomy, and botany.

The system of symbols here presented, except for a few minor changes, was suggested in 1927[1] and is convenient and adequate,

[1] CHAFFEE, Vacuum Tube Nomenclature, *Proc. I.R.E.*, **15**, 182 (1927).

as proved by several years of use by the author. Special symbols are required, although more awkward alternatives to the special symbols are suggested, to be used when the special type is not available.

70. Fundamental Scheme of System of Symbols.—Before describing in detail the various symbols, a few general principles are presented.

1. *Capital letters are used only for quantities which are not directly functions of time, current, or potential, but which may be functions of frequency.*

2. *Small letters are used for instantaneous values of quantities which are directly functions of time or of some other independent variable such as current, potential, temperature.*

3. *Boldface type is used to denote complex or vector quantities.*

4. *Subscripts immediately following a letter symbol denote, in general, location rather than kind of quantity.*

a. Subscripts f, p, and g immediately following a letter denote quantities directly associated with the filament, plate, and grid of a vacuum tube.

b. Subscripts a, b, and c immediately following a letter denote quantities which are associated with the portions of the filament, plate, and grid circuits external to the vacuum tube.

c. Capital letter subscripts A, B, and C denote sources of power, such as batteries and generators, in the filament, plate, and grid circuits.

5. *Subscripts immediately following parentheses enclosing a quantity indicate the frequency of the alternating quantity or the frequency at which the quantity is evaluated.*

Subscripts h, i, and l immediately following parentheses enclosing a quantity are used to indicate high, intermediate, and low frequencies.

6. *Those electrical quantities most used, such as total instantaneous values and effective or r.m.s. values, are denoted simply by small and capital letters with no attached supersigns.*

NOTE.—The symbols used in the first five chapters are not in strict accordance with this proposed system. Capital letters I and E were there used to denote steady values instead of r.m.s. values. It seemed unwise to introduce before this point a more or less complicated system of symbols. A small letter for potential might have been confused with the symbol for electronic charge which is now so firmly established by usage that it seemed unwise to change the letter for it.

From this point on, static characteristic curves will be expressed in terms of small letters.

71. Special Symbols.—Some of the special symbols, together with their alternatives, are described below.

7. *Dash over a letter which denotes an electrical quantity* indicates the *steady value* of that quantity.

Example: $\bar{E}$. Alternatives: $(E)_0$ or $({}_sE)$.

8. *Dash over a letter which denotes a circuit characteristic* indicates the value of that characteristic for a *steady current.*

Example: $\bar{R}$. Alternatives: $(R)_0$ or $({}_sR)$.

9. *Two dashes over a letter which denotes an electrical quantity* indicate the *time average* of that quantity.

Example: $\bar{\bar{E}}$. Alternatives: E_{av} or $({}_aE)$.

10. *Circumflex over a capital letter which denotes an electrical quantity* indicates the *maximum value* of a sinusoidally varying quantity.

Example: $\hat{E}$. Alternatives: $E_{max.}$ or $({}_mE)$.

The maximum value may also be denoted by the square root of two times the effective or r.m.s. value, as $\sqrt{2}E$.

11. *Caret sign over a letter which denotes an electrical quantity* indicates the *peak value* of the quantity.

Example: $\hat{\bar{E}}$. Alternatives: E_{pk} or $({}_pE)$.

12. *Tilde sign over a small letter which denotes an electrical quantity* indicates that the quantity *varies harmonically.*

Example: $\tilde{e}$. Alternative: $({}_he)$.

13. *Dash over a small letter which denotes an electrical quantity* indicates a component *measured from the steady value.*

Example: $\bar{e}$. Alternative: $({}_se)$.

14. *Two dashes over a small letter which denotes an electrical quantity* indicate a component *measured from the average value.*

Example: $\bar{\bar{e}}$. Alternative: $({}_ae)$.

15. *A* Δ *sign preceding a letter which denotes an electrical quantity* indicates that the quantity is *small but finite.*

Example: ΔE or Δe.

16. *An inverted circumflex over, or square brackets about, a letter or symbol* indicates the *equivalent value* or indicates that the quantity is *fictitious.*

Examples: $\breve{E}$ or $[E]$.

72. List of Symbols.—For easy reference, the various quantities, with their distinguishing symbols as described above, are listed below:

LIST OF QUANTITIES WITH SYMBOLS

Quantity	Preferred Symbol	Alternative Symbol
Effective or r.m.s. potential	E	
Maximum of harmonic potential	$\sqrt{2}E$ or $\widehat{\mathrm{E}}$	$E_{max.}$ or $({}_mE)$
Steady component of potential	$\bar{E}$	$(E)_0$ or $({}_sE)$
Average value of potential	$\overline{\mathrm{E}}$	E_{av} or $({}_aE)$
Peak value of potential	$\hat{\mathrm{E}}$	$\hat{E}_{pk}$ or $({}_pE)$
Total instantaneous potential	e	
Instantaneous potential measured from steady value	$\tilde{e}$	$({}_se)$
Instantaneous potential measured from average value	$\tilde{\tilde{e}}$	$({}_ae)$
Small potential	ΔE or Δe	
Equivalent potential	$\breve{E}$	$[E]$
Fictitious potential	$\breve{E}$	$[E]$
Resistance to steady current	$\bar{R}$	$(R)_0$
Resistance to current of high or radio frequency	$(R)_h$	
Resistance to alternating current without specifying frequency	$\tilde{R}$	
Variational resistance	r	

NOTE.—Although the symbols are illustrated as applied to potential and resistance, they are similarly applied to other electrical quantities and circuit characteristics.

73. Illustrations and Examples.—The use of the symbols is further illustrated.

A harmonically varying potential superimposed upon a steady value is illustrated in Fig. 49*a*. In the case shown, the steady and average values are identical. Obviously,

$$e = \bar{E} + \tilde{e} = \bar{E} + \hat{E} \sin \omega t,$$
$$= \bar{E} + \sqrt{2}E \sin \omega t.$$

A current or potential often varies nonsinusoidally. An example of this is illustrated in Fig. 49*b*. We may then write

$$e = \bar{\bar{E}} + \tilde{e}, \text{ or, developing } e \text{ in a Fourier series,}$$
$$e = \bar{\bar{E}} + (\hat{E})_1 \sin \omega_1 t + (\hat{E})_2 \sin (\omega_2 t + \Phi_2)$$
$$+ (\hat{E})_3 \sin (\omega_3 t + \Phi_3) + \cdots$$

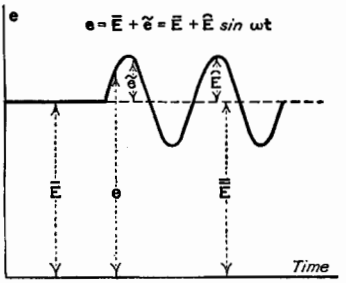


FIG. 49*a*.

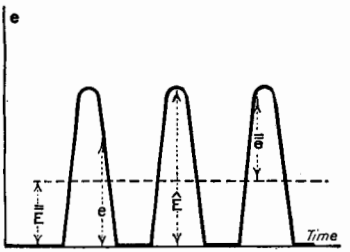


FIG. 49*b*.

where the subscripts attached to ω and Φ indicate the several harmonics.

In Fig. 50 curve *Oab* represents the characteristic curve of a conductor whose resistance to steady current is a function of the impressed voltage. The resistance should be denoted by a small letter with a dash over it to indicate that it is a steady-current resistance. Therefore, the resistance $\bar{E}/\bar{I}$ is denoted by $\bar{r}$. The *variational resistance*, which is equal to $\text{limit}_{\Delta e \doteq 0}\left(\frac{\Delta e}{\Delta i}\right)$, is denoted by r. Its reciprocal, the *variational conductance*, is denoted by k.

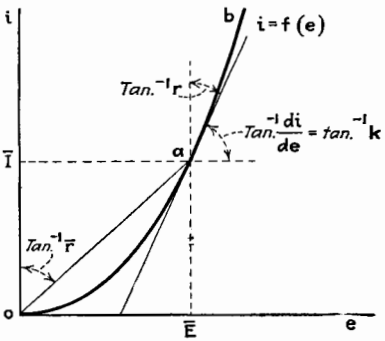


FIG. 50.

In Fig. 51 the use of symbols with subscripts is illustrated as applied to a vacuum-tube circuit. The circuit L_1C_2 is assumed to be carrying a sinusoidal current induced from a circuit not shown. The total instantaneous plate voltage e_p may be made up of a steady component $\bar{E}_p$ and an alternating component $\Delta\tilde{e}_p$, so that

$$e_p = \bar{E}_p + \Delta\tilde{e}_p$$

Similarly, the voltages e_b, e_c, and e_g may be made up of steady and sinusoidal components. The currents to plate and grid may also be made up of components as shown in the figure.

The system of symbols described is adequate to distinguish the various quantities which are usually dealt with in theoretical discussions of the operation of vacuum tubes. When the theory

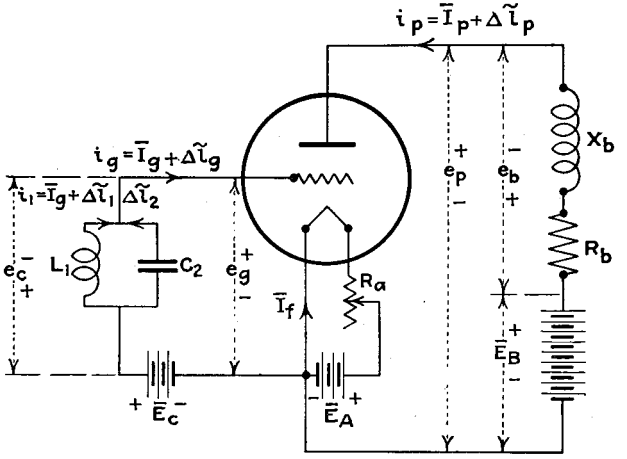


FIG. 51.—Illustration of use of symbols.

is simple and, in consequence, few symbols are necessary, some of the special signs may be omitted. In problems dealing only with direct currents, it would be absurd to place a dash over every letter denoting current and potential. The tilde sign may be omitted if only sinusoidal variations are being considered.

In later chapters, other special letters and symbols will be used to represent special quantities then defined. For a complete list of symbols see page v.

CHAPTER VII

THREE-ELECTRODE TUBE OR TRIODE

Previous to the year 1907, practically the only use to which the two-electrode tube was put was the rectification of alternating currents, either for power purposes or as a Fleming valve detector in the reception of radio messages. These applications will be described in later chapters. Lee de Forest[1] in 1907 made a most important and valuable improvement in the vacuum tube, which very much expanded its field of usefulness. His invention consisted in the addition of a third electrode adapted to the electrostatic control of the electron current to the plate, and the vacuum tube as thus constructed he called the "audion." Usually, the control electrode is in the form of a wire grid or mesh, situated *between* the filament and plate so that the electron stream passes through the spaces of this control electrode. This third electrode is usually called the *grid.*

The use of a third electrode, as introduced by de Forest, was not entirely novel. Lenard[2] in 1902 used a third electrode in a vacuum tube for the purpose of studying the photoelectric effect. De Forest, however, deserves the credit of making the vacuum tube a more practical device by the insertion of the grid.

It should be remarked that at about the time de Forest made his invention, von Baeyer[3] in Germany added a grid in the form of a wire-gauze cylinder, placed between filament and plate.

The object of this chapter is to study the current-potential characteristics of the triode, from both the mathematical and the graphical points of view.

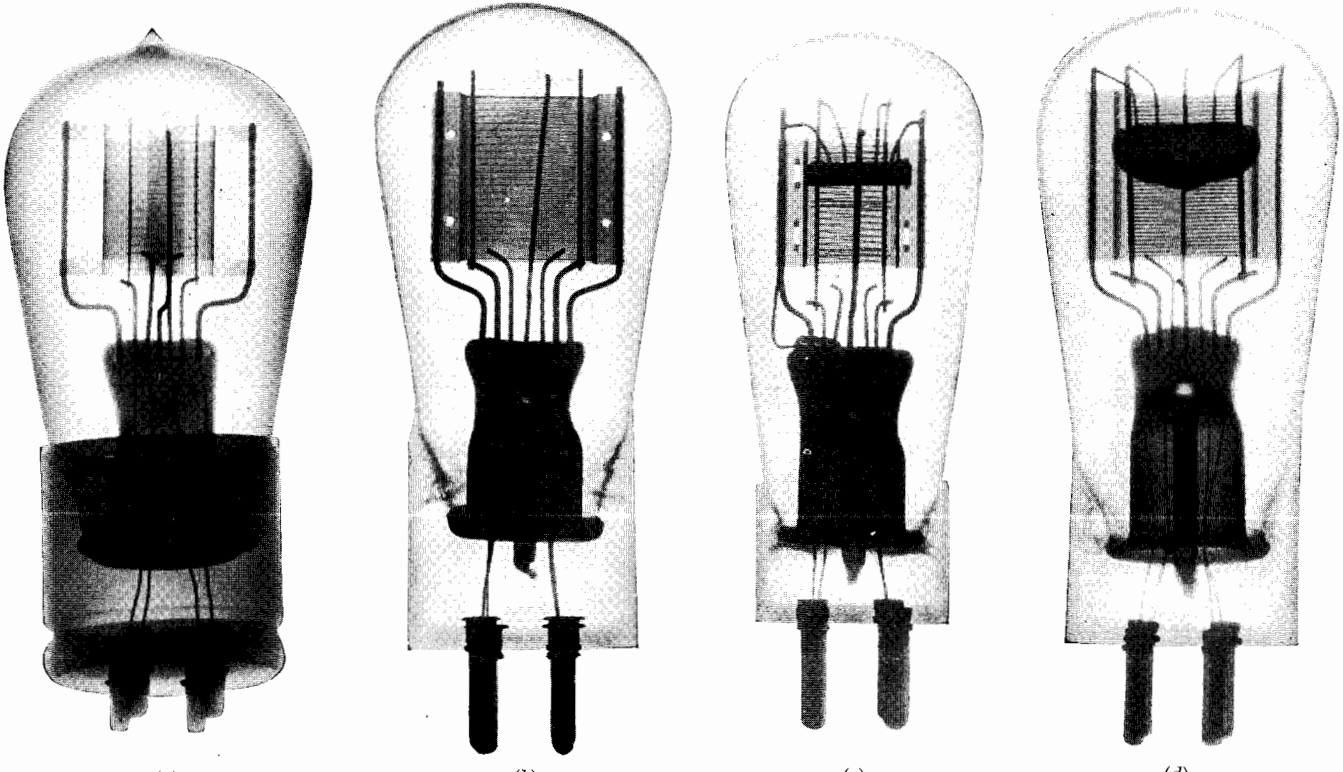
74. The Equivalent Diode.—Van der Bijl[4] in 1913 was one of the first to give a quantitative discussion of the effect of the third electrode when situated between filament and plate. The following treatment is substantially his.

[1] De Forest, U. S. patent 841,387 (1907); U. S. patent 879,532 (1908).

[2] Lenard, *Verh. D. Phys. Ges.*, **8**, 149 (1902).

[3] Von Baeyer, *Verh. D. Phys. Ges.*, **7**, 109 (1908).

[4] Van der Bijl, *Verh. D. Phys. Ges.*, **15**, 338 (1913); *Phys. Rev.*, **12**, 180 (1918).



(*a*) (*b*) (*c*) (*d*)

PLATE II.—X-ray photographs of vacuum tubes. (*a*) Triode, UV-201; (*b*) triode, type 201A, (*c*) triode, type 230; (*d*) triode, type 240.

(*Facing page* 144)

In Fig. 52, F, G, and P represent the filament, grid, and plate of a triode. The directions of the positive currents (opposite to direction of travel of the electrons) $\bar{I}_f$, i_g, and i_p are indicated by arrows. The potentials of the grid and plate with respect to the filament are determined by the grid and plate batteries $\bar{E}_C$ and $\bar{E}_B$. In the arrangement shown, $e_g = \bar{E}_C$ and $e_p = \bar{E}_B$.

Observe the use of the notation for steady current to indicate the filament current, because in general it is held constant. The plate and grid currents and the corresponding plate and grid voltages are indicated by small letters, because in general in the operation of the tube they vary with time. Although the characteristics may be obtained experimentally by using steady values of current and voltage, these characteristics of the tube are valid for varying current and voltage and are expressed, therefore, in terms of instantaneous values.

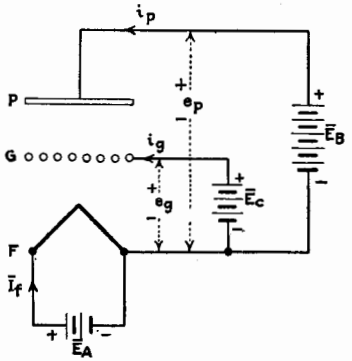


FIG. 52.—The triode.

We shall first examine the electrostatic fields in the vacuum tube before the filament is heated. The intensity of the electrostatic field at the filament due to the grid voltage alone, the plate voltage being zero, is proportional to e_g. Further, if e_g be assumed zero and e_p be applied, the intensity of the field at the filament due to e_p is the same as though some smaller voltage De_p, where D is a constant, were impressed between filament and grid in place of e_p between filament and plate.

The constant factor D takes account of the shielding effect of the grid mesh and the greater distance from the filament to the plate than from the filament to the grid. Van der Bijl used the Greek letter γ to denote the shielding factor D, whereas in German literature the letter D is used and is called the *Durchgriff*. Obviously D has a value less than unity.

If e_p and e_g are simultaneously impressed, the resulting field at the filament is equal to the sum of the fields due to e_p and e_g acting separately and is proportional to $e_g + De_p$. Therefore, $e_g + De_p$ is an *equivalent grid voltage* which would produce the same field intensity at the filament as is actually produced by the

joint action of both grid and plate potentials. This equivalent grid voltage is denoted by $[e_g]$.

It was shown in Chap. IV that the space current, if limited by space charge, is determined by the electrostatic field at the filament produced by the charged plate. When the space current flows, the field at or very near the filament is reduced to zero by the opposing electric field set up by the space charge. The electrostatic field at the filament is the same as if $[e_g]$ alone were impressed on the grid, the plate being at filament potential. By analogy with Eq. (39), page 69, for the diode, we may say that the total space current in the triode is given by

$$i = i_p + i_g = B[e_g]^{3/2} = B(e_g + De_p)^{3/2} \tag{106}$$

where B is a constant. Equation (106), like Eq. (39) for the diode, holds only for the lower part of the plate-current curve before filament saturation begins to limit the current.

The total space current of a triode can thus be calculated by assuming the triode to be replaced by a special kind of diode; the cathode and the plate of the triode being the negative electrode of the diode, and the grid of the triode being the positive electrode of the diode. Such a diode is not the kind of diode for which Eq. (39) applies; hence constant B in Eq. (106) is not the same as the constant of Eq. (39). If a standard type of diode were made having a solid plate in the same position as the grid of the triode, the space current in this diode would be the same as the space current of the special diode if a voltage $\lambda[e_g]$ were impressed across the standard type of diode, where λ is a correction factor less than unity. Therefore, using Eq. (40) instead of Eq. (39),

$$i = B[e_g]^{3/2} = 2.336 \times 10^{-6}\frac{(\lambda e_g)^{3/2}}{d^2{}_g}\ \frac{\text{amp.}}{\text{cm.}^2} \tag{107}$$

The correction factor λ can be found as follows: The space current from the cathode is determined by the intensity of the electrostatic field at the cathode due to the neighboring charged electrodes although this field is reduced to zero by the opposing field from the space charge when the space current flows. This electrostatic field at the cathode is directly proportional to the induced charge q per unit area of the cathode, assuming the cathode grounded and the other electrodes charged. For example, in the triode

$$q = c_g e_g + c_p e_p \tag{108}$$

where c_g and c_p are the coefficients of capacitance. Since the space current is the same if $[e_g]$ alone acts on the grid, the plate being grounded,

$$q = c_g[e_g] = c_g(e_g + De_p) \tag{109}$$

From Eqs. (108) and (109)

$$c_p = Dc_g \tag{110}$$

Again, since the same current flows in the standard type of diode when λe_g acts on its plate,

$$q = c(\lambda e_g) \tag{111}$$

where c is the coefficient of capacitance for the standard diode.

Consider now the triode. If e_p is equal to e_g, the space between grid and plate is a constant-potential region, and the plane of the grid is practically an equipotential surface. The field between grid and cathode is then practically the same as if the grid were replaced by a solid plate as in the standard diode. Consequently from Eqs. (108) and (111),

$$c = c_g + c_p = c_g(1 + D) \tag{112}$$

Since the special diode and the standard diode are assumed to pass the same current, Eqs. (109) and (111) can be equated, $[e_g]$ and e_g being numerically the same. Hence

$$\lambda = \frac{1}{1 + D} \tag{113}$$

Equation (107) for the triode with plane electrodes can now be written

$$i = i_p + i_g = \frac{2.336 \times 10^{-6}}{d_g^2}\left(\frac{e_g}{1 + D} + \frac{De_p}{1 + D}\right)^{3/2} \frac{\text{amp.}}{\text{cm.}^2} \tag{114}$$

From Eqs. (52) and (113) the total space current per unit length for a triode with cylindrical electrodes is

$$i_p + i_g = \frac{14.68 \times 10^{-6}}{r_g\beta^2}\left(\frac{e_g}{1 + D} + \frac{De_p}{1 + D}\right)^{3/2} \text{amp.} \tag{115}$$

Evidently the correction factor which must be multiplied into B of Eq. (106) to give the constant for the standard type of diode is $\left(\frac{1}{1 + D}\right)^{3/2}$. In other words, the plane-electrode form of triode with plate connected to cathode and a voltage e_g impressed

on the grid is equivalent to a standard diode having a plate at a distance $d_g(1 + D)^{\frac{3}{4}}$ from the cathode. If the triode has cylindrical electrodes and the cathode has a small radius compared to that of the grid so that β^2 is practically unity, the plate of the equivalent diode has a radius $r_g(1 + D)^{\frac{3}{4}}$.

As explained in Chap. IV, it is more exact to add to the equivalent voltages in Eqs. (106), (107), (114), and (115) a small voltage e', which is an approximate correction to allow for the initial velocity of emission of the electrons and for any contact-potential difference which may exist. The more exact form of Eq. (114), for example, is

$$i_p + i_g = \frac{2.336 \times 10^{-6}}{d_g^2(1 + D)^{\frac{3}{2}}}(e_g + De_p + e')^{\frac{3}{2}} \frac{\text{amp.}}{\text{cm.}^2} \qquad (116)$$

75. Curves of Constant Total Space Current.—If D is actually a constant, as assumed, graphs of constant $i_p + i_g$ plotted on the $e_p - e_g$ plane should be parallel straight lines. That this is true in some cases is shown by the graphs of Fig. 53. As shown by the figure, D departs from its constant value for small positive values of e_p. It also departs from constancy for a range of negative values of e_p which is greater, the greater the current. Since the triode is practically never used with values of e_p within these limits, this departure of D does not at all reduce the practical value of the assumption of constant value of D. This departure from linearity of the constant-current curves for small values of e_p is due to a space charge built up near the plate. When e_p is small, the electrons which pass through the grid and move toward the plate are in a retarding field. The electrons lose the velocity acquired in moving from cathode to grid, and near the plate they have their minimum velocity. If the plate voltage is zero, the electrons have zero velocity at the plate. If the plate voltage is negative, the electrons have zero velocity somewhere between the plate and grid, depending upon the value of e_p. Where the velocity of the electrons is low, their density is great and therefore a negative space charge exists between the grid and plate which alters the effect of the plate voltage in producing an electrostatic field at the cathode. For small positive values of e_p, the space charge decreases the effect of e_p, and hence e_p must be increased in order to maintain the same current from the cathode. For small negative values of e_p, the space charge opposes the effect of the grid to produce an accelerat-

ing field at the cathode. Under this condition, the plate voltage must be less negative to maintain constant current from the cathode.

Figure 53 also shows that when the space current is reduced to values of the order of 0.001 milliamp., the value of D departs from its constant value. D becomes considerably greater than the usual value for positive values of e_p, and slightly less than the usual value for negative values of e_p. This change in the value of

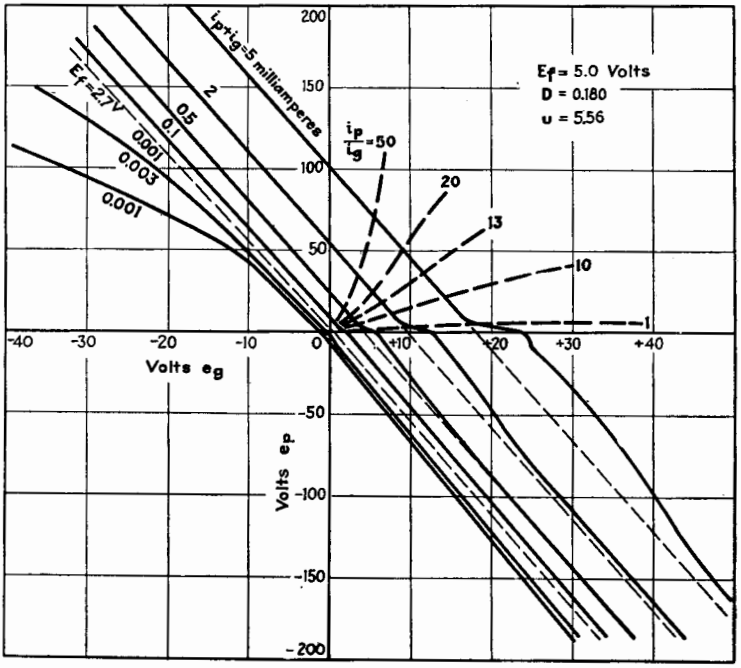


FIG. 53.—Static characteristic curves of constant total space current of a triode having a tungsten filament (UV 201).

D is mostly due to the change in location of the region of minimum potential outside the cathode. In Chap. IV, Sec. 42, it was explained that the point of minimum potential outside the cathode, caused by the initial velocity of emission of the electrons, is ordinarily very close to the cathode. But when the space current limited by space charge is a very small fraction of the total emission of the cathode, this region of minimum potential, which may roughly be considered as the position of a virtual cathode, moves out to a very appreciable distance from the

cathode surface. In Fig. 53, when the space current is 0.001 milliamp. and the total emission current is greater than 5 milliamp., the position of the virtual cathode is much nearer the grid than normally, and hence the effective shielding of the grid is less. That this is the correct explanation is shown by the graph in dashes of Fig. 53 taken for 0.001-milliamp. space current, but for a

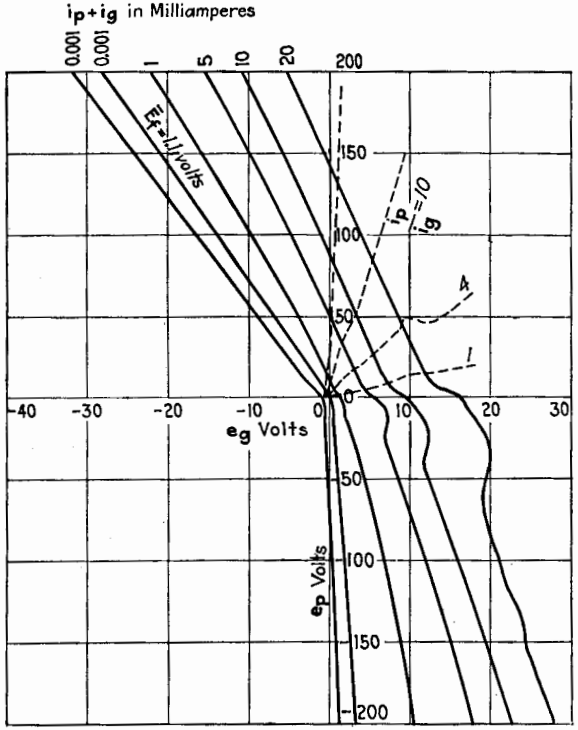


FIG. 54.—Static characteristic curves of constant total space current. $E_f = 2.5$ volts. Equipotential cathode (UY 227).

much reduced filament temperature corresponding to a filament voltage of 2.7 instead of 5.0 volts. This reduction in filament temperature much reduces the total emission and hence reduces the effect of initial velocity of the electrons.

In Fig. 53, e_g is negative in the second quadrant and hence all the space current goes *to the plate*. In the fourth quadrant, e_p is negative and hence all the space current goes *to the grid*. In the first quadrant are both a grid current and a plate current.

The dotted lines are loci of constant ratio of plate to grid current, the ratio being noted on the curves.

The constant total-current curves are not always so regular and straight as those shown in Fig. 53. For example, the curves of Fig. 54 were taken for a tube of cylindrical plate and grid and

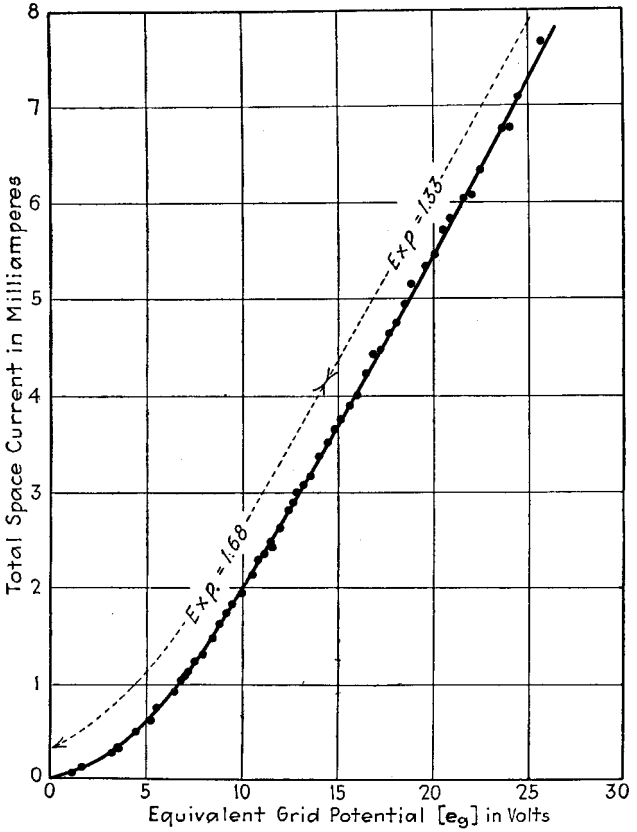


FIG. 55.—Total space current vs. equivalent grid voltage. Tungsten filament (UV 201). $D = 0.181$.

a separately heated, oxide-coated cathode. For positive values of e_p, the value of D is practically constant for a constant value of total space current. D varies somewhat as the space current varies, being about 0.090 for 20 milliamp. and 0.105 for 5 milliamp. The same effect described in the case of Fig. 53 operates to increase D for small values of current. The variation of D for large values of current in this tube is due to irregularity of the

geometry of the tube and will be treated more in detail shortly. Note particularly the large change in D as e_p becomes negative, especially for small values of e_g and small space currents. For example, for a current of 0.001 milliamp., D is approximately 0.014.

In Fig. 55, the total space current is shown plotted against the equivalent grid voltage $[e_g]$ for the tube from which the data for Fig. 53 were taken. The observed points were taken for values of e_p ranging from 20 to 120 volts, and, except for experimental error, lie well on a single curve for the equivalent diode. The value of the contact potential e' in Eq. (116) is difficult to determine from these data but is apparently not much greater than 0.5 volt.

As explained in Chap. V for the diode, the potential drop along the filament and other effects may cause the exponent of Eq. (116) to be different from $\frac{3}{2}$. The exponent for the curve of Fig. 55 is about 1.68 over the lower part of the curve but becomes 1.33 at the upper part.

If D in Eq. (116) is taken outside the parentheses,

$$i_p + i_g = D^{\frac{3}{2}}B[e_p + u(e_g + e')]^{\frac{3}{2}} \tag{117}$$

where $u = 1/D$.† Evidently u is the ratio of the voltage change applied to the plate to that applied to the grid for the same change in plate current. The factor u is known as the *voltage amplification factor* or, preferably, the *voltage ratio* of the tube and is a very important quantity. For the tube used in obtaining the curves of Figs. 53 and 55, u is 5.52.

This treatment for the triode is valuable in giving a simple physical picture and in pointing out the constancy of D under certain conditions. It is limited in its application because the three-halves-power law of Eq. (116) does not accurately give the true course of the current curve. Furthermore, it does not express the current when temperature saturation affects it or when secondary emission or gas ionization is present. A mathematical formula, such as Eq. (116), for example, is seldom of value in dealing with problems concerning vacuum tubes. A general

† The voltage ratio of a triode is denoted by the Greek letter μ in much of the literature on vacuum tubes written in English. The author, however, prefers the letter u for this quantity as being the letter nearest in form to μ, and in order that the three coefficients, u, k, and s, of a triode may be denoted by letters of the Roman alphabet.

theory will now be given taking into account all effects which determine the flow of the grid and plate currents in a triode.

76. General Treatment of the Triode.—The plate and grid currents of a triode are the two dependent variables and are different functions of the three independent variables, namely, temperature of the cathode, plate voltage, and grid voltage. Expressed in mathematical language,

$$i_p = F_1(T, e_p, e_g) \tag{118}$$

$$i_g = f_1(T, e_p, e_g) \tag{119}$$

The manner in which both i_p and i_g vary, as the temperature T of the emitter is changed, is similar to the variation of the space current of a diode with changes in temperature, as discussed in Chap. IV. To simplify the problem, the temperature will be assumed constant at the rated value. Equations (118) and (119) may be written

$$i_p = F(e_p, e_g) \tag{120}$$

$$i_g = f(e_p, e_g) \tag{121}$$

77. Static Characteristic Curves of the Triode.—The functional relations of Eqs. (120) and (121) cannot be expressed by any mathematical equations which hold over the entire range of values of the variables as used in practice. We must then resort to graphical methods to give a picture of the variations of i_p and i_g with changes in e_p and e_g. Since we have reduced each equation to one of three variables, each relation can be represented by a surface in three dimensions. In each case we may plot dependent variable, i_p or i_g, vertically, and the independent variables e_p and e_g horizontally. Plate III shows photographs of two plaster models of the plate and grid currents of a particular triode. The surfaces for different triodes are different in scale and in some details but the general characteristics of all triode surfaces are the same. Those of Plate III may be considered as typical.

Sections of these surfaces, made by planes perpendicular to the three axes, give for each surface three families of curves, known as the *static characteristic curves* of the tube. Usually, in solving a given problem, one family of plate-current curves and one family of grid-current curves give all the necessary information, but the particular sets of curves used depend on the problem. One should become familiar with the general shapes of all six families of curves. These curves are illustrated for a particular tube in the following six figures taken from a paper by

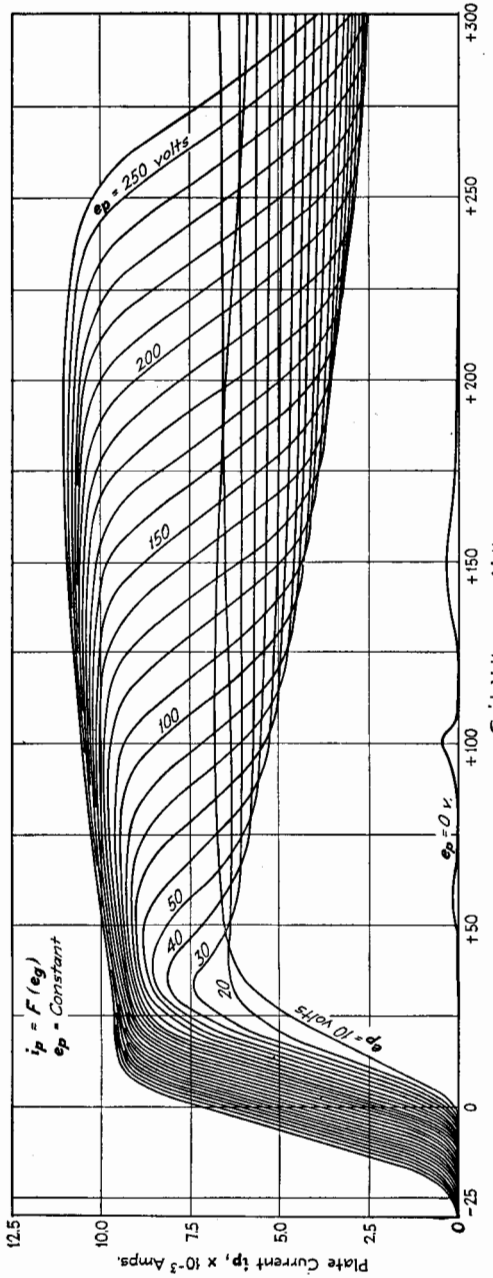
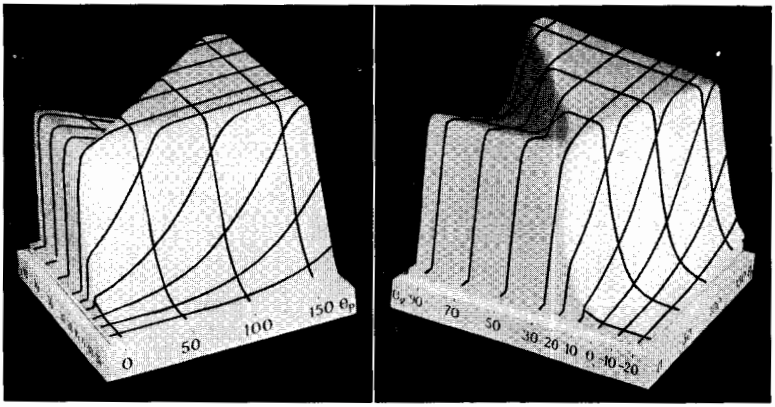
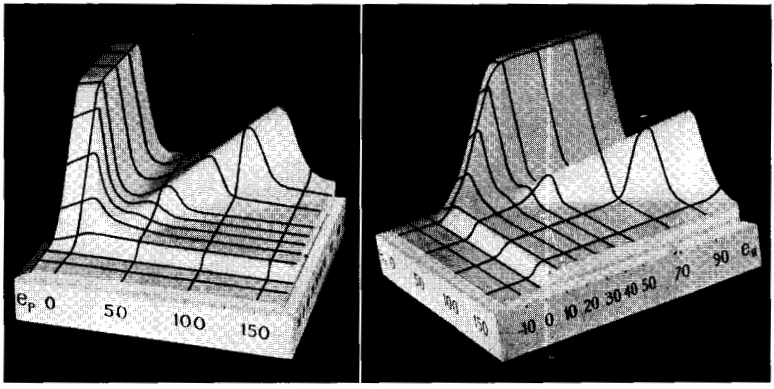


FIG. 56.—Static characteristic curves of the plate current of a triode. Curves of constant plate voltage. (*van der Pol.*)



(*a*) (*b*)



(*c*) (*d*)

PLATE III.—Plate-current and grid-current surfaces of a triode. (*a*) and (*b*) plate-current surface; (*c*) and (*d*) grid-current surface.

(*Facing page* 154)

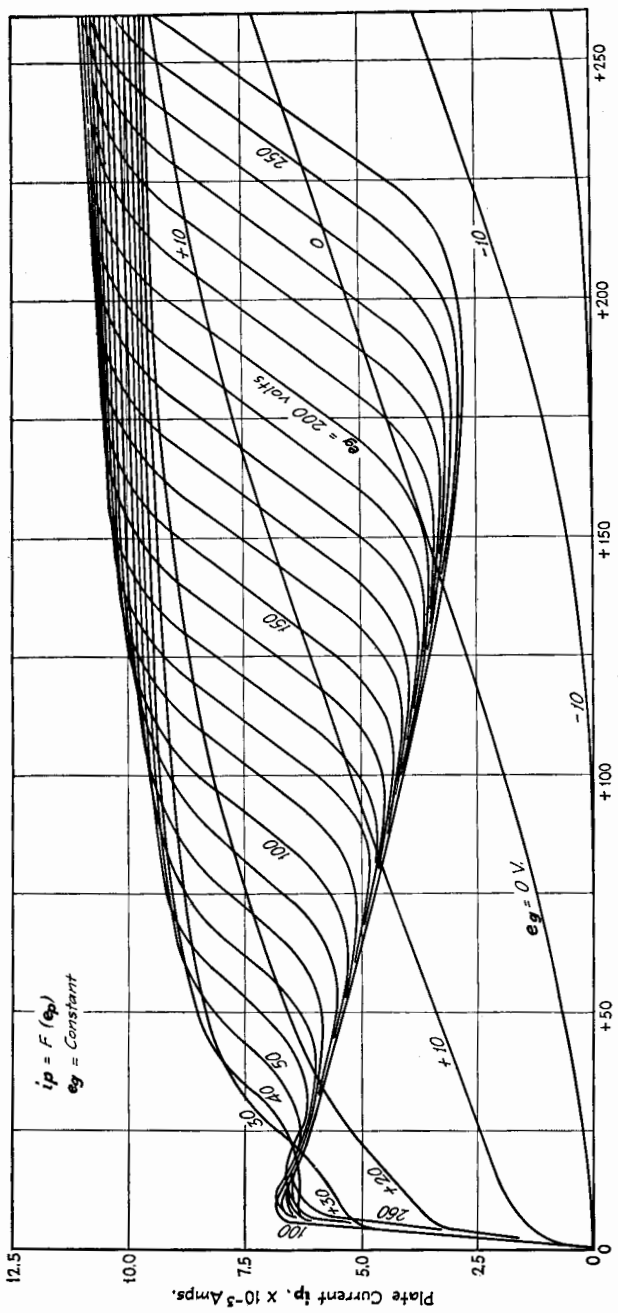


FIG. 57.—Static characteristic curves of the plate current of a triode. Curves of constant grid voltage. (*van der Pol.*)

van der Pol.[5] The curves of Fig. 56 are the $i_p - e_g$ curves obtained by section of the plate-current surface made perpendicular to the grid-voltage axis. Figure 57 shows the $i_p - e_p$ curves for constant values of e_g, and Fig. 58 gives the contours for constant plate current. The portion of the curves of Fig. 58 for negative grid voltages is coincident with the curves of constant

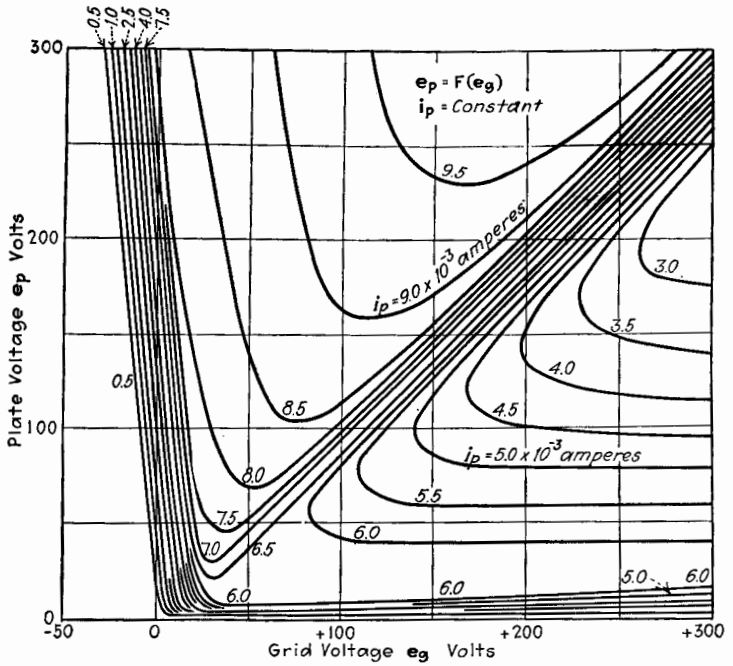


FIG. 58.—Static characteristic curves of the plate current of a triode. Curves of constant plate current. (*van der Pol.*)

total space current, since for negative grid voltages the grid current is zero. This portion of the curves of Fig. 58 shows that D is constant over this region at least, but tells nothing as to the constancy of D for other voltages.

The corresponding families of curves for the grid-current surface are given in Figs. 59, 60, and 61. The curves of Fig. 61 for negative plate voltages are curves of constant total space current and are practically extensions of the portions of the curves of Fig. 58 for negative grid voltages.

[5] VAN DER POL, BALTH, *Zeits. f. Hochfrequenztechnik*, **25**, 125, (1925).

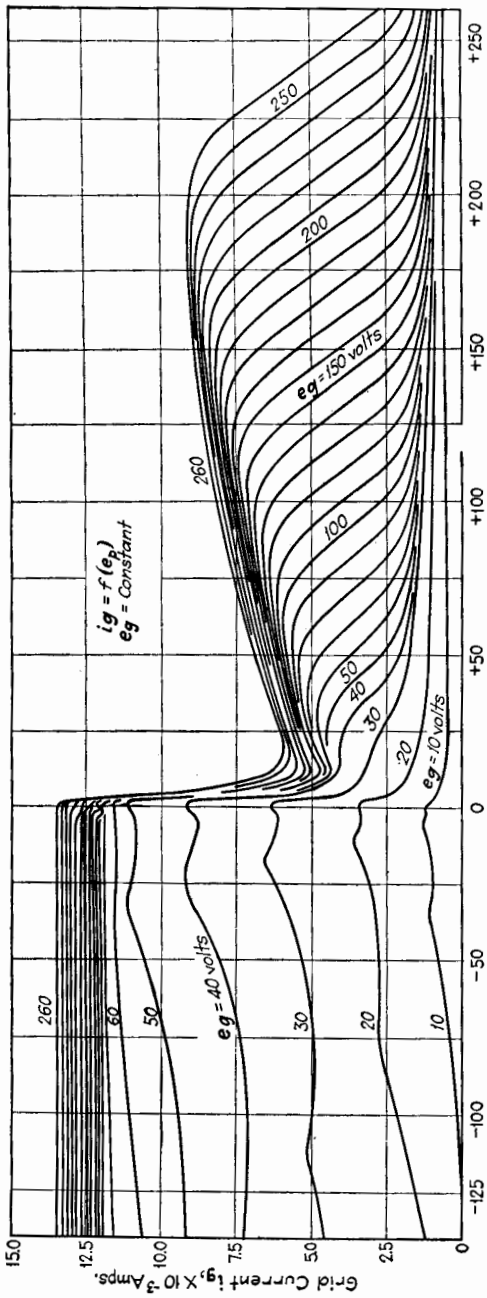


FIG. 59.—Static characteristic curves of the grid current of a triode. Curves of constant grid voltage. (*van der Pol.*)

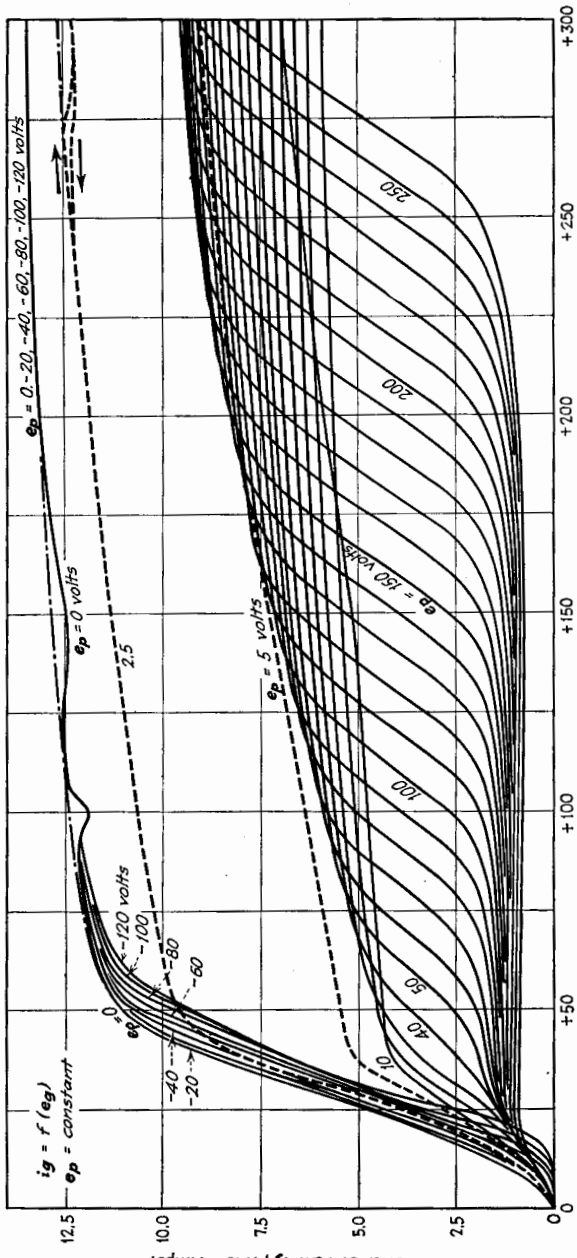


FIG. 60.—Static characteristic curves of the grid current of a triode. Curves of constant plate voltage. (*van der Pol.*)

Referring to the plate-current surface shown in Plate III, we note two important regions. The first region is that of the steep flat portion almost perpendicular to the axis of grid voltage. The second region includes the depression of the upper part of the surface.

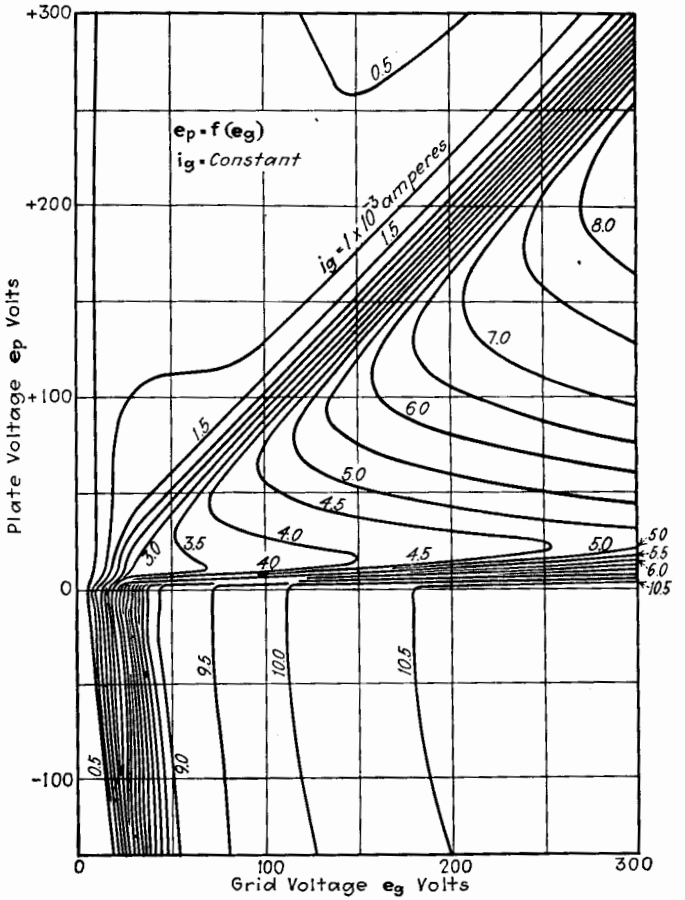


FIG. 61.—Static characteristic curves of the grid current of a triode. Curves of constant grid current. (*van der Pol.*)

The *first* region is that part of the surface used in the normal operation of the triode as an amplifier, detector, and oscillator. In Fig. 57, the curves for grid voltages from -20 to $+20$ volts indicate sections within this region. It is seen that over a considerable part of this region the curves of Fig. 57 are straight,

parallel, and equidistant for equal increments of grid voltage, which shows that this region is practically a plane. The same plane portion is indicated by the closely spaced, parallel, straight lines which cut across the axis of zero grid volts in Figs. 56 and 58.

The depression in the *second* region of the plate-current surface is due partly to secondary emission of electrons and partly to the change in the fraction of the total space current taken by the plate. On one sloping side of the depression, secondary emission occurs at the plate; on the other sloping side of the depression, secondary emission may occur at the grid. For example, in Fig. 57, the slope of the depression indicated by the portions of the curves having a negative slope is due to secondary emission from the plate. On this part of the surface the grid voltage is always greater than the plate voltage, and as the plate voltage is increased, the plate current decreases. This results because the increase in plate voltage causes a larger increase in secondary electrons which are then drawn to the grid than the increase of primary electrons striking the plate. This same part of the surface is indicated in Fig. 56 by the lower, nearly horizontal portions of the curves for constant plate voltage. In Fig. 58, it is indicated by the portions of the contour lines which are approximately parallel to the axis of grid voltage and are marked from 6.0 to 3.0 milliamperes in the direction of increasing plate voltage.

The other sloping side of the depression of the plate-current surface occurs where the plate and grid voltages are approximately equal. As the plate voltage increases, the grid voltage being held constant, the plate current increases rapidly. This is due partly to the normal increase in flow of electrons to a terminal whose voltage is increased, but mostly to the shift of the total space current from grid to plate as the plate voltage becomes equal to and greater than the grid voltage. Secondary emission at the grid may play some part in shaping the surface in those portions of the region where the plate voltage is greater than the grid voltage and the grid voltage is positive and greater than 50 or 100 volts. This slope of the depression is indicated in Fig. 57 by the steeply rising curves for constant grid voltages of from 30 to 250 volts, and in Fig. 56 by the portions of the curves having a negative slope. Since the increase in plate current in this region with increasing plate voltage is due largely to the shift of current from grid to plate, the decrease of grid current is evident as shown in Fig. 59 by the portions of the curves with a negative slope.

Certain special types of vacuum tubes known as *dynatrons* operate on one of the slopes of this depression in the plate-current surface.

The steep plane portion of the characteristic surface of a triode, designated above as the first region, is the region of greatest importance when studying the operation of triodes as amplifiers and detectors. It is the usual practice to plot this region of the characteristic surface to much enlarged scales of

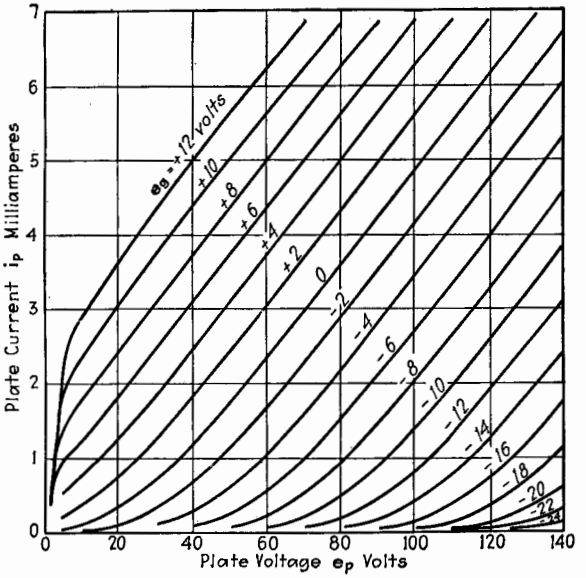


FIG. 62.—Static characteristic curves of a triode. $i_p - e_p$ curves for constant grid voltage (UV 201).

voltage and current and hence much more restricted ranges of voltage and current than for the curves of Figs. 56 to 61. For example, Fig. 62 shows the $i_p - e_p$ curves for various constant values of e_g ranging from -24 volts to $+12$ volts, in 2-volt steps. A chart of this sort will frequently be used in the following discussions. Figure 63 is the chart of the corresponding $i_p - e_g$ curves for various constant values of e_p. Figures 62 and 63 demonstrate that there exists a practically *flat portion* of the plate-current surface. Figure 64 shows the $i_g - e_g$ curves for various constant values of e_p, but differs from Fig. 60 in being for a much smaller range of voltages. Figure 64 includes the operating range of most low-power vacuum tubes. It is par-

ticularly noteworthy that, except for very low values of e_p, the plate voltage has very little effect upon the grid current for the range shown in Fig. 64. This is shown better by the $i_g - e_p$ curves of Fig. 65.

The characteristic curves just described may be considered as typical of the characteristic curves of all high-vacuum triodes.

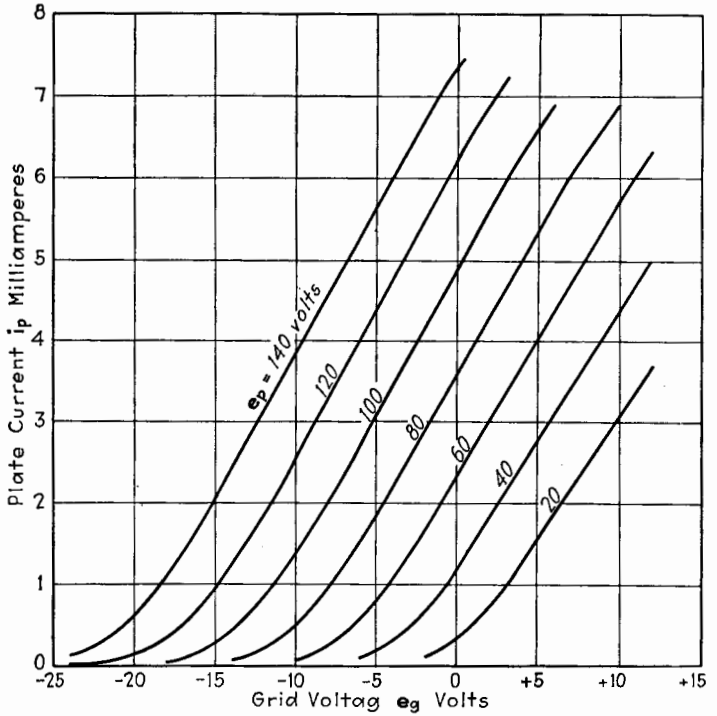


FIG. 63.—Static characteristic curves of a triode. $i_p - e_g$ curves for constant plate voltage (UV 201).

Of course, certain features may be exaggerated in certain tubes, and the scales of current and voltage vary according to the distances between electrodes, size of electrodes, and closeness of grid mesh. The saturation of the plate and grid currents shown in Figs. 56, 57, and 59 for the tube with a tungsten filament is much less pronounced for tubes having oxide-coated or thoriated cathodes. The presence of gas greatly distorts the characteristics, as will be explained in Chap. X.

78. Time Lag in Vacuum Tubes.—In a triode, the only effects which can cause a time lag between the plate current or the grid

current and the applied voltages are the time of flight of an electron from filament to grid or plate, and the time of readjustment of the space charge, the latter being directly dependent upon the velocity of the electrons. The time of flight of an electron is extremely short. For a potential difference of 1 volt and a spacing of ½ cm., this time is about 0.5×10^{-8} sec.

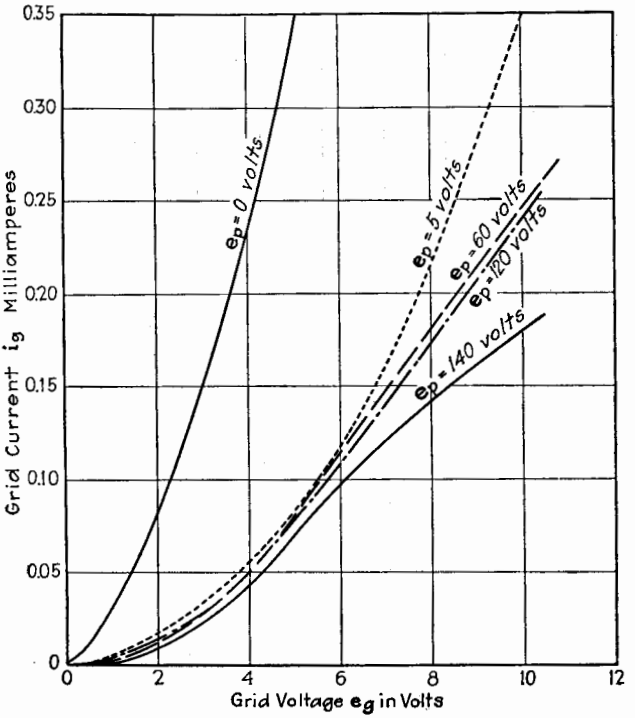


FIG. 64.—Static characteristic curves of a triode. $i_g - e_g$ curves for constant plate voltage (UV 201).

Hence, the time lag due to this cause need not be considered except when the tubes are used for very short waves of the order of 10 meters or less. The time required for the space charge to readjust itself may be greater than this time, because the potential differences which determine the equilibrium of the space charge are small. If gas is present and ionization takes place, the positive ions, having a much greater mass than the electron, move much more slowly, so that considerable lag effects are observed in tubes containing gas. For the present,

only gas-free tubes are considered and the lag due to the electrons is neglected. The static characteristic curves of Figs. 53 to 65 may be considered valid for rapid changes of the variables.

79. Variational Characteristics of Triodes.—The characteristics of triodes have been described thus far in terms of characteristic curves. We shall now develop a general mathematical method for dealing quantitatively with these characteristics.

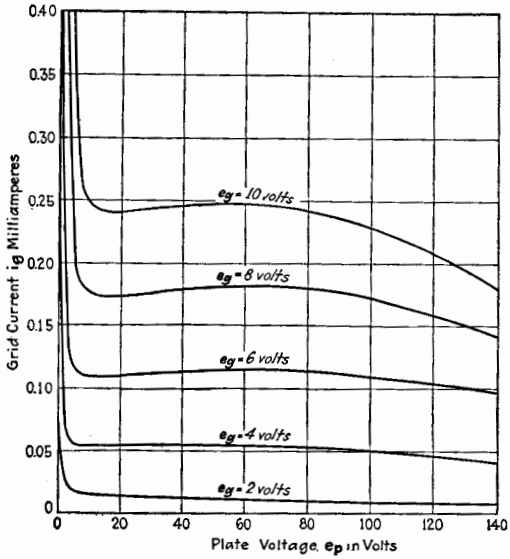


FIG. 65.—Static characteristic curves of a triode. $i_g - e_p$ curves for constant grid voltage (UV 201).

We may start with Eqs. (120) and (121) and assume the temperature of the cathode to be constant. The total differential of Eq. (120) is

$$di_p = \frac{\partial i_p}{\partial e_p} de_p + \frac{\partial i_p}{\partial e_g} de_g \tag{122}$$

Similarly, for Eq. (121),

$$di_g = \frac{\partial i_g}{\partial e_p} de_p + \frac{\partial i_g}{\partial e_g} de_g \tag{123}$$

Expressions (122) and (123) are general relations which depend upon no assumptions or restrictions as to the functional relations (120) and (121) except that these functions be continuous.

The four partial derivatives occurring in Eqs. (122) and (123) are important coefficients and are given specific names. The

coefficient $(\partial i_p/\partial e_p)_{e_g=\text{const.}}$ has the dimensions of conductance. It is called the *cathode-to-plate variational conductance*, or simply the *plate variational conductance*, and is denoted by k_p. Its reciprocal is the *plate variational resistance* and is denoted by r_p. The coefficient $(\partial i_g/\partial e_g)_{e_p=\text{const.}}$ is the *cathode-to-grid variational conductance*, or simply the *grid variational conductance*, and is denoted by k_g. Its reciprocal is the *grid variational resistance* and is denoted by r_g.

The other two coefficients in Eqs. (122) and (123) involve the effect of the grid circuit on the plate circuit or the effect of the plate circuit on the grid circuit and have been called *mutual coefficients*. Exception may well be taken to the use of the word *mutual*, because the effects defined by these coefficients are only one-way effects. For example, the coefficient $(\partial i_p/\partial e_g)_{e_p=\text{const.}}$ has long been known as the *grid-to-plate mutual conductance*, but it does not give the reciprocal effect of the plate circuit upon the grid circuit. A better designation for this coefficient is the *grid-to-plate variational transconductance*, or simply the *variational transconductance*, and is denoted by s_p. Similarly, the remaining coefficient $(\partial i_g/\partial e_p)_{e_g=\text{const.}}$ is the *plate-to-grid variational transconductance*, or simply the *variational reflex transconductance*, and is denoted by s_g. Hence, Eqs. (122) and (123) may be written in the following forms:

$$di_p = k_p de_p + s_p de_g \tag{124}$$

$$di_g = s_g de_p + k_g de_g \tag{125}$$

If the plate and grid voltages are so varied that the plate current remains constant, another useful partial differential coefficient is obtained, namely, $(\partial e_p/\partial e_g)_{i_p=\text{const.}}$ The *negative* of this coefficient is known as the *plate variational amplification factor*, or better the *plate variational voltage ratio*, and is denoted by u_p.

A relation among the three coefficients belonging to the plate circuit can be obtained by putting di_p of Eq. (124) equal to zero. Then.

$$\left(\frac{de_p}{de_g}\right)_{di_p=0} = \left(\frac{\partial e_p}{\partial e_g}\right)_{i_p=\text{const.}} = -u_p = -\frac{s_p}{k_p} \tag{126}$$

As a result, we obtain the important relation

$$s_p = u_p k_p \tag{127}$$

A similar analysis may be carried through for the grid-current equation, Eq. (125). Putting di_g equal to zero,

$$\left(\frac{de_g}{de_p}\right)_{di_g=0} = \left(\frac{\partial e_g}{\partial e_p}\right)_{i_g=\text{const.}} = -\frac{s_g}{k_g} \tag{128}$$

The negative of the partial derivative expressed in Eq. (128) is known as the *variational reflex factor*, or the *variational reflex voltage ratio*, and is denoted by u_g. A relation similar to Eq. (127) obtains for the grid coefficients, namely,

$$s_g = u_g k_g \tag{129}$$

For easy reference the six coefficients are listed below:

$$\left.\begin{aligned} k_p &= \left(\frac{\partial i_p}{\partial e_p}\right)_{e_g} = \frac{1}{r_p} \\ s_p &= \left(\frac{\partial i_p}{\partial e_g}\right)_{e_p} \\ u_p &= -\left(\frac{\partial e_p}{\partial e_g}\right)_{i_p} \end{aligned}\right\} s_p = u_p k_p$$

$$\left.\begin{aligned} k_g &= \left(\frac{\partial i_g}{\partial e_g}\right)_{e_p} = \frac{1}{r_g} \\ s_g &= \left(\frac{\partial i_g}{\partial e_p}\right)_{e_g} \\ u_g &= -\left(\frac{\partial e_g}{\partial e_p}\right)_{i_g} \end{aligned}\right\} s_g = u_g k_g \tag{130}$$

Referring to Eq. (117), page 152, the voltage ratio u is given by

$$u = -\left(\frac{\partial e_p}{\partial e_g}\right)_{i_p+i_g=\text{const.}} \tag{131}$$

If e_g is negative and hence i_g is zero, Eq. (131) reduces to Eq. (126) and u becomes u_p. Further, if e_p is negative, i_p is zero and u becomes $1/u_g$. If we add Eqs. (124) and (125),

$$d(i_p + i_g) = (k_p + s_g)de_p + (s_p + k_g)de_g \tag{132}$$

Following the method used in arriving at Eq. (126), we have

$$u = \frac{s_p + k_g}{s_g + k_p} \tag{133}$$

The six variational coefficients of a triode, as defined in Eqs. (130), are the fundamental factors in terms of which the

theory of operation of the vacuum tube is developed. It is therefore essential to become familiar with the definitions of the coefficients and to have a visual picture of the manner in which the coefficients vary over the characteristic surfaces of a triode. Although triodes may vary considerably in design, the effects of varying the size, shape, and spacings of the electrodes are principally to change the scales of the characteristic surfaces, without materially altering their general shape. As a necessary

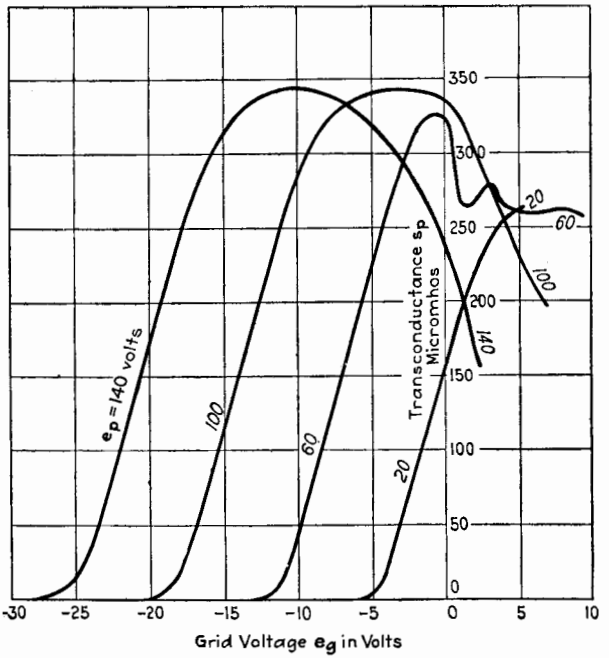


FIG. 66.—Transconductance of a triode (UV 201).

consequence, the variations of the six coefficients over the characteristic surfaces are similar for all triodes, even though not the same in scale and magnitude.

The variational characteristics of a vacuum tube can be plotted in two ways. The first and usual method consists in plotting the values of the coefficients against one of the variables, such as e_g or e_p. Such graphs for the variational coefficients, corresponding to the particular characteristic curves of Figs. 62 to 65, are plotted against e_g in Figs. 66 to 71.

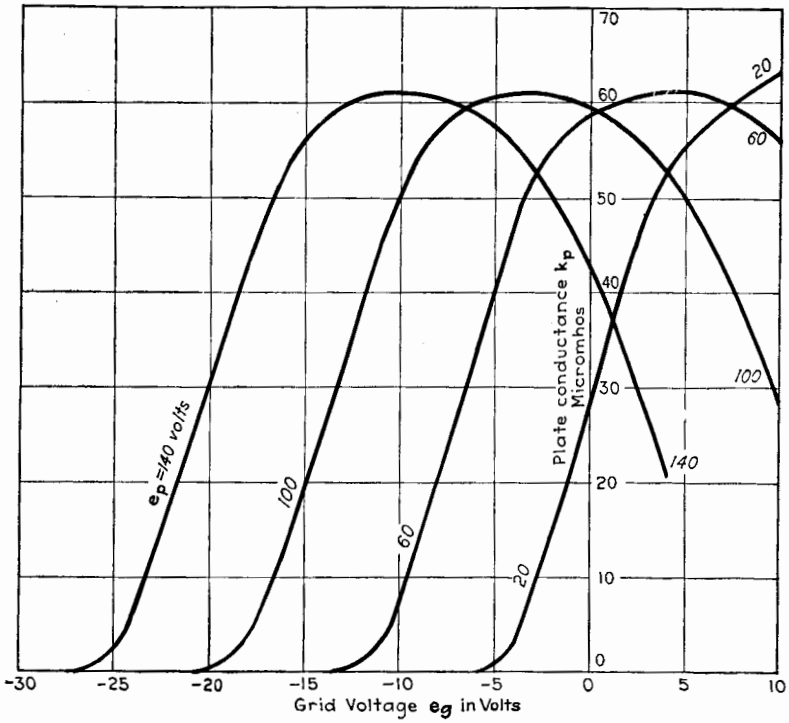


FIG. 67.—Plate variational conductance of a triode (UV 201).

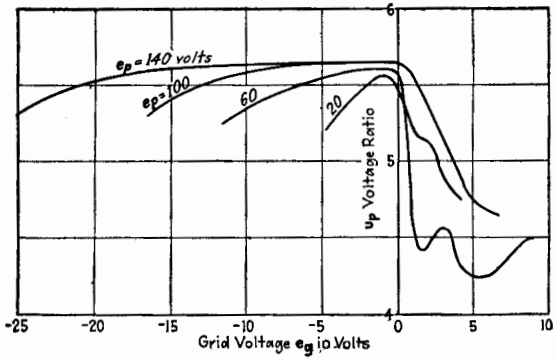


FIG. 68.—Voltage ratio of a triode (UV 201).

The variational properties of a triode may also be shown to advantage by drawing contours for constant values of the coefficients on the static characteristic curves as in Figs. 72 and 73.

The triode coefficients can be approximately determined directly from the static characteristic curves in the region where the curves are approximately straight. For example, let the curves of Fig. 74 be two of the $i_p - e_p$ curves of a tube for two grid voltages differing by a finite amount Δe_g. The ratio $(\overline{ab}/\overline{ca})$ is approximately k_p and is the slope of the $i_p - e_p$ curves. Also,

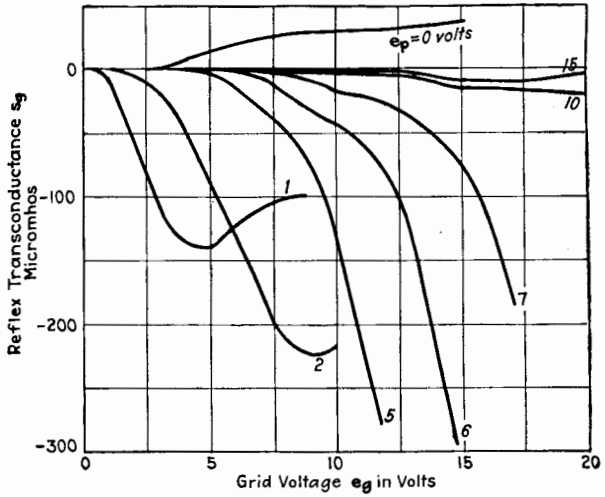


FIG. 69.—Reflex transconductance of a triode (UV 201).

s_p is approximately $(\Delta i_p/\Delta e_g)_{e_p=\text{const.}}$ or $\overline{ab}/\Delta e_g$; and u_p is approximately $(-\Delta e_p/\Delta e_g)_{i_p=\text{const.}}$ or $-(\overline{ca}/\Delta e_g)$.

The values of the coefficients obtained by using finite increments of the variables are inaccurate if the increments are large or if the curvature is appreciable. However, approximate values are conveniently determined by this method. Any one of the three families of curves for the plate current, two of which are shown in Figs. 62 and 63, gives all three of the plate coefficients. Similarly, all three of the grid coefficients can be obtained from any one of the three sets of grid-current curves, two of which are shown in Figs. 64 and 65. The coefficients can be more accurately determined from the static characteristic curves by measuring the slopes of the curves. By this method, the slopes of the curves of Fig. 62 give k_p, the slopes from Fig. 63 give s_p, etc.

The graphical methods as outlined for determining the six coefficients are inaccurate and cumbersome. Electrical methods for directly measuring the coefficients are in common use and will be described in a later chapter.

An examination of Fig. 68 shows that u_p is approximately constant for negative values of e_g for the particular triode used. The

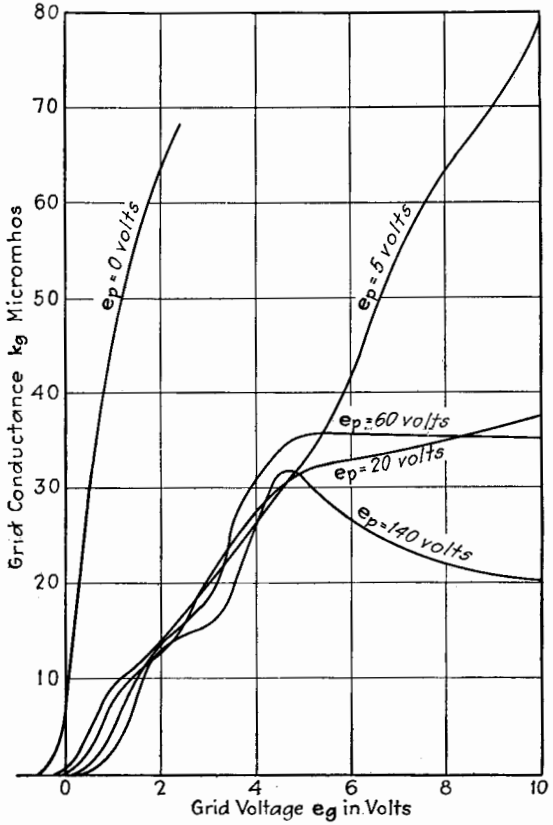


FIG. 70.—Grid variational conductance of a triode (UV 201).

plate conductance k_p, shown in Fig. 67, and the transconductance, given in Fig. 66, are not constant over any considerable range. Wherever the graph of k_p or s_p is straight, the corresponding static characteristic curve is parabolic in shape. If the graph of k_p or s_p is a straight horizontal line, the corresponding static characteristic graph is straight. Figures 69 and 71 show that, within the range of voltages of the graphs, both s_g and u_g are negative

quantities and have very small values except when e_p is small. Figure 61 shows, however, that u_g may be positive as well as

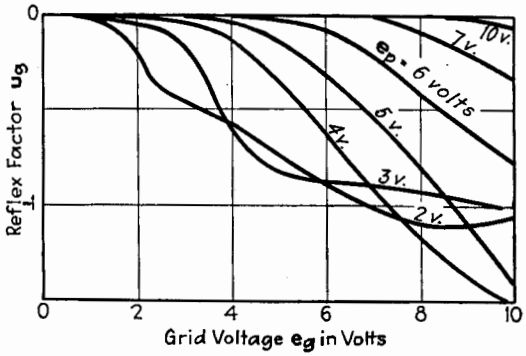


FIG. 71.—Reflex factor of a triode (UV 201).

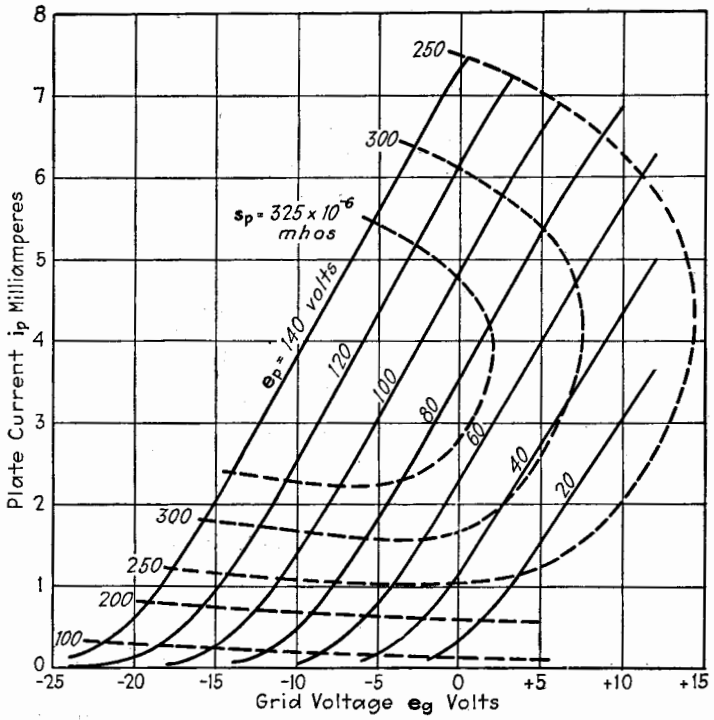


FIG. 72.—Contours of transconductance of a triode (UV 201).

negative if the value of e_g is large. The grid conductance k_g of a high-vacuum tube is practically always positive for positive

values of e_g. It is zero for negative values of e_g because there is then no grid current. Figure 59 shows that s_g is negative for small positive values of e_p. It becomes positive for certain

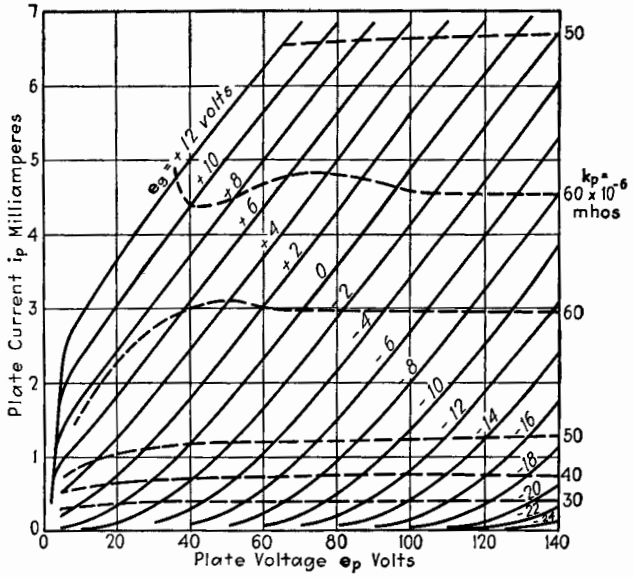


FIG. 73.—Contours of plate variational conductance of a triode (UV 201).

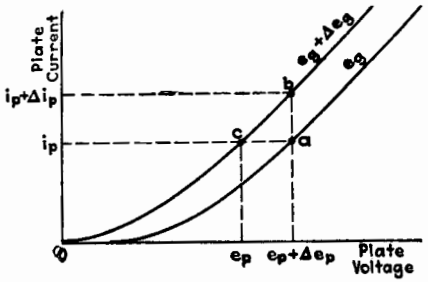


FIG. 74.

values of e_p for which u_g is positive. For negative values of e_p, s_g is in general positive but very small.

Since vacuum tubes are more often operated with negatively polarized grids, and s_g and k_g are then zero, the grid coefficients are of little interest compared to the much more important plate coefficients u_p, k_p, and s_p.

80. Further Study of the Action of the Grid.—The various characteristics of a triode as measured externally have been set forth in the preceding sections. We shall now examine in more detail the action of the grid in controlling the flow of electrons inside the tube. Since the triode is an electrostatic device in that its action is governed by the electrostatic forces acting upon the electrons, a fair insight into the action of the grid can be had by a study of the electrostatic fields about the elements of the triode for various combinations of the plate and grid voltages.

First, let it be pointed out that the electrons travel in obedience to the forces exerted on them not only by the charged electrodes but also by the space charge. The electrostatic flux lines, which pass from the plate to the cathode when the cathode is cold and no emission takes place, are intercepted by the space charge when the cathode is heated. The electrons move toward the plate in sufficient numbers to intercept all of the lines of flux, and hence no flux lines reach the cathode when the space current is limited to saturation value by space charge. The space charge is ordinarily most dense near the cathode. But when the plate voltage is low, the space charge may be great near the plate, as pointed out in Sec. 75. The position of the space charge may considerably alter the shape of the electrostatic field about the elements and thus alter the relative effects of the grid and plate in controlling the current. When the space charge is most dense near the cathode, the distortion of the field due to the space charge is least. Under these circumstances, a fair idea of the action of the grid can be deduced from the shape of the electrostatic field for a cold cathode.

The shape of the electrostatic field about a wire grid placed between two electrodes has been calculated as a problem in pure electrostatics, first by Clerk Maxwell[6] and later by several others.[7,8,9,10] Only simple cases, which geometrically are symmetrical, can thus be treated. For example, a wire grid between two parallel plane surfaces, all three being infinite in extent to eliminate edge effects, can be solved. A cylindrical grid composed of wires parallel to the axis of the cylinder, placed con-

[6] MAXWELL, "Electricity and Magnetism," 3d ed., Vol. I, Par. 203, Oxford, 1904.

[7] ABRAHAM, *Arch. Elektrotechnik*, **8**, 42 (1919).

[8] SCHOTTKY, *Arch. Elektrotechnik*, **8**, 21 (1919).

[9] VON LAUE, *Ann. d. Physik*, **59**, 465 (1919); *Jahrb. Drahtlos. Tel. Tel.*, **14**, 243 (1919).

[10] KING, *Phys. Rev.*, **15**, 256 (1920).

centrically between an inner and outer cylinder, all of infinite length, can also be solved.

The shape of the electrostatic field, when edge effect is present, can be obtained experimentally by setting up a model of a triode in a shallow tray containing acidulated water, and tracing out equipotential lines by a probe connected to a telephone receiver. The equipotential lines of Figs. 75, 76, and 77 were obtained in this way, and the lines of flow (shown in red) of electrons from cathode to the other electrode were sketched by making them everywhere perpendicular to the equipotential lines. For other drawing of equipotential lines and lines of flow, see Möller.[11]

The left half of Fig. 75 represents the condition when the grid potential is so negative that the field at the cathode urges the electrons back into the cathode. The grid potential for the right half of Fig. 75 was calculated to give zero field at the cathode, and hence the figure represents the condition when the electron flow to the plate is practically cut off. The value of u was calculated by formula (137), to be given later. In the right half of Fig. 75, a few electrons pass to the plate around the grid, as indicated by the flow lines. In Fig. 76, the grid potential is zero for the left half of the figure and is positive for the right half. Figure 77 gives the corresponding four cases for a cylindrical arrangement of the electrodes.

If the distance from cathode to grid is not large compared with the distance between grid wires, and the potential of the grid is somewhat negative, there may be regions of the cathode where the field urges electrons back into the cathode while, in the regions between, the field is in the reverse direction and urges electrons away from the cathode. This condition is known as *island emission.*

The electron stream, especially for negative grid potentials, is constricted as it passes between grid wires so that the electrons strike the plate in localized areas, as shown in Figs. 76 and 77. This is often made visible by sharp lines of fluorescence on the plate in some tubes having oxide-coated cathodes. This fluorescence is caused by electronic bombardment of minute quantities of the oxide material which has been deposited on the plate by evaporation from the cathode.

[11] MÖLLER, "Die Elektronenröhren," p. 18, Friedr. Vieweg & Sohn, Akt.-Ges. Braunschweig, 1922.

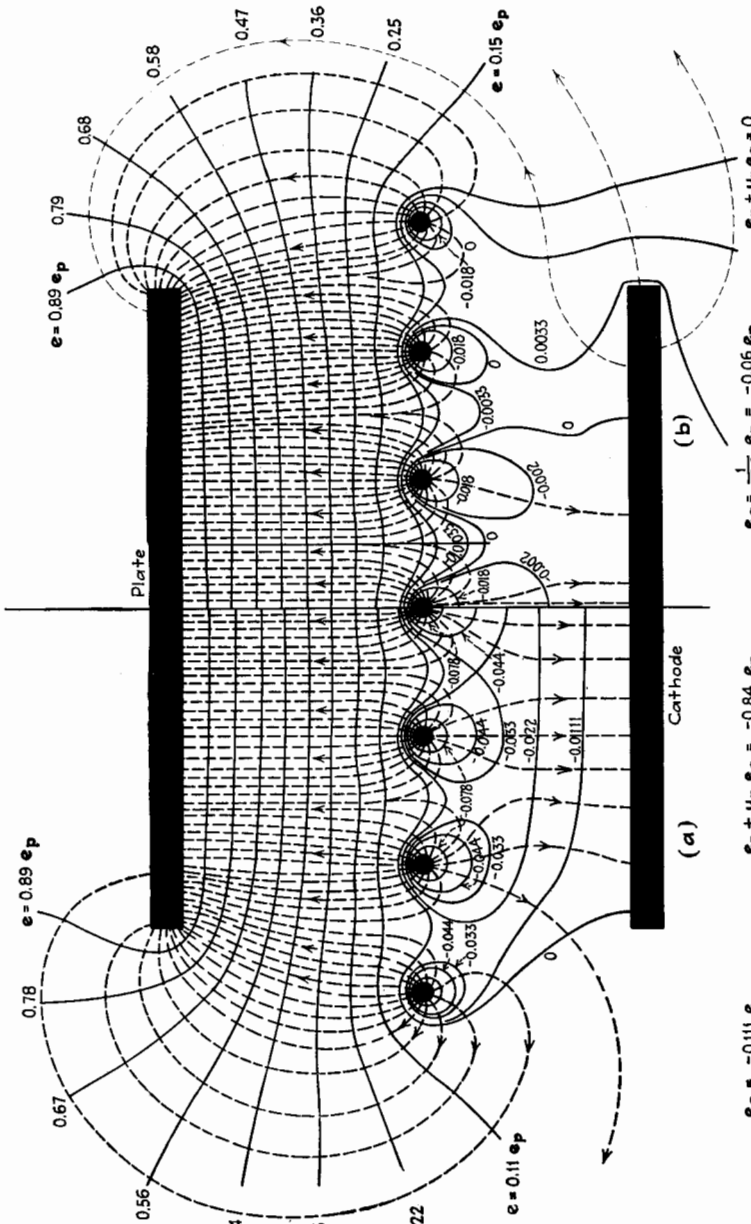


FIG. 75.—Equipotential lines and lines of force in a triode having plane electrodes. Paths of electrons shown in red.

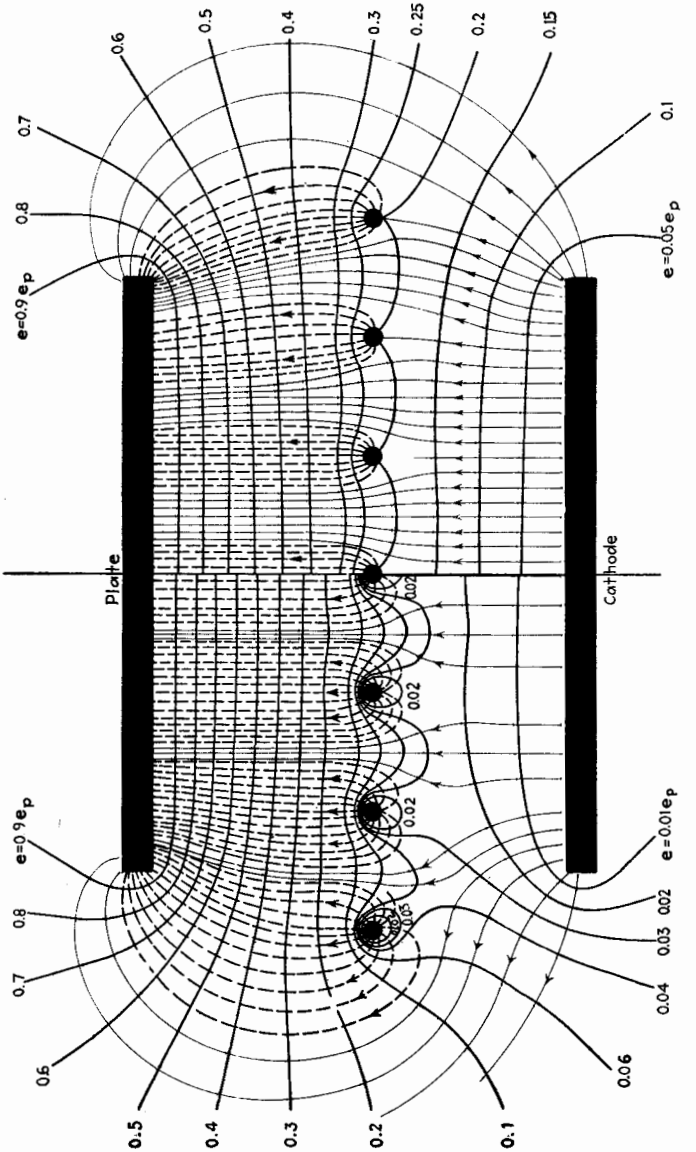


FIG. 76.—Equipotential lines and lines of force in a triode having plane electrodes. Paths of electrons shown in red.

The Durchgriff D or voltage ratio u can easily be derived from the solution of the electrostatic problem. The method in general terms consists in calculating the changes of grid and plate voltages which produce the same change in the electrostatic field at the cathode. Then u is the ratio of these increments in voltage. The reason that the space charge near the cathode has but little effect in making the measured value of u different from the cal-

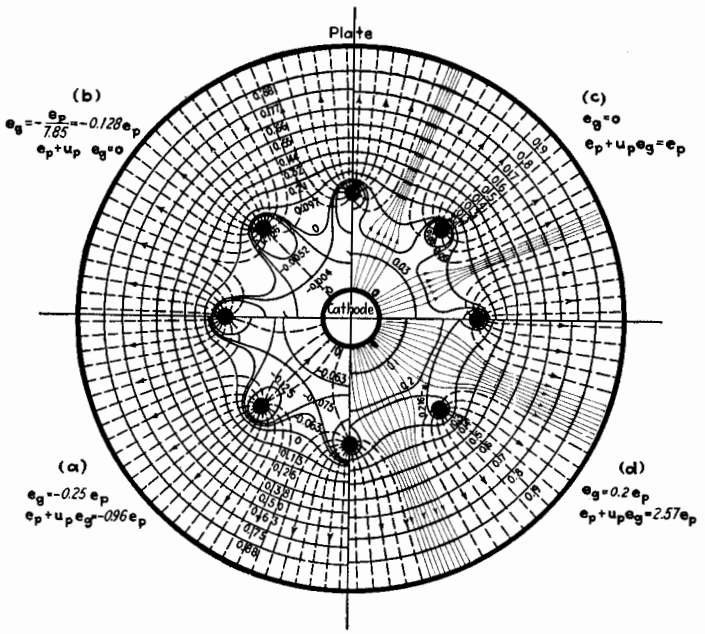


FIG. 77.—Equipotential lines and lines of force in a triode having cylindrical electrodes. Paths of electrons shown in red.

culated value is as follows. The space charge, when very near the cathode, affects the field at the cathode due to charges on the grid and plate by about the same fractional amount. Therefore, the ratio of the increments of voltage on plate and grid to produce the same change in plate current is but slightly affected by the space charge.

Several forms of expression for u have been derived and differ according to the approximations made.[10,12,13]

[12] MILLER, *Proc. I.R.E.*, **8**, 64 (1920).

[13] VOGDES and ELDER, *Phys. Rev.*, **24**, 683 (1924).

Miller in March, 1919, in a report at the Bureau of Standards, which was later published,[12] gives the following formula for the voltage ratio of a triode with plane electrodes. The cathode is assumed to be an infinite emitting plane.

$$u = -\frac{2\pi nb}{\log_\epsilon (2 \sin \pi\rho n)} \text{ (plane electrodes—Miller)} \quad (134)$$

where b is the distance from the grid to the plate, n is the number of grid wires per centimeter length of grid, and ρ is the radius of the grid wires.

King[10] calculated the following two expressions for u, the first for plane electrodes and the second for cylindrical electrodes, the grid wires being assumed parallel to the axis of the cylinders.

$$u = \frac{2\pi nb}{\log_\epsilon \frac{1}{2\pi\rho n}} \text{ (plane electrodes—King)} \quad (135)$$

$$u = \frac{2\pi nb \frac{\rho_g}{\rho_p}}{\log_\epsilon \frac{1}{2\pi\rho n}} \text{ (cylindrical electrodes—King)} \quad (136)$$

In Eq. (136), ρ_g and ρ_p are the radii of the grid and plate. Equations (134) and (135) are practically the same if $\pi\rho n$ is a small angle.

In deriving these expressions for u, the radius of the grid wires was assumed small in comparison with the distance between wires. Vogdes and Elder[13] have given the following expressions which are not thus limited. These are the most accurate expressions yet derived for this important parameter.

$$u = \frac{2\pi nb - \log_\epsilon \cosh 2\pi n\rho}{\log_\epsilon \coth 2\pi n\rho} \text{ (plane electrodes—Vogdes and Elder)} \quad (137)$$

$$u = \frac{2\pi n\rho_g \log_\epsilon \frac{\rho_p}{\rho_g} - \log_\epsilon \cosh 2\pi n\rho}{\log_\epsilon \coth 2\pi n\rho} \text{ (cylindrical electrodes—Vogdes and Elder)} \quad (138)$$

In Eq. (138), if $\frac{\rho_p}{\rho_g}$ is less than 2, $\log_\epsilon \frac{\rho_p}{\rho_g}$ may be expanded in a series, and Eq. (138) becomes

$$u = \frac{2\pi nb\left[1 - \frac{1}{2}\frac{b}{\rho_g} + \frac{1}{3}\left(\frac{b}{\rho_g}\right)^2 - \cdots\right] - \log_\epsilon \cosh 2\pi n\rho}{\log_\epsilon \coth 2\pi n\rho} \quad (139)$$

If b/ρ_g is small in comparison with unity, Eq. (139) becomes the same as Eq. (137) for the plane electrodes. As the value of b/ρ_g increases, the value of u for a cylindrical arrangement of electrodes becomes less than that for a plane arrangement of electrodes, the values of ρ and n remaining the same. The value of the series in brackets, Eq. (139), is plotted against b/ρ_g in Fig. 78. Since the last term in the numerator of Eqs. (137), (138), and (139) is usually small in comparison with the first term, the plot of Fig. 78 gives a fair idea of the ratio of values of u for a plane and a cylindrical arrangement, the values of ρ, n, and b being the same.

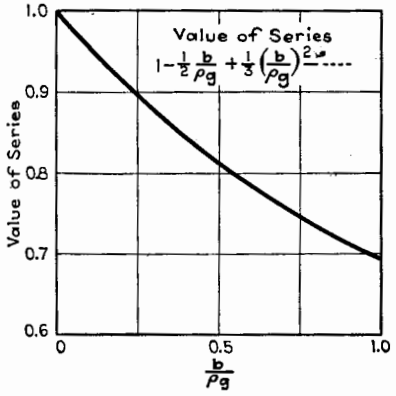


FIG. 78.

The formulas for the cylindrical arrangement of electrodes were derived for grids having wires parallel to the axis of the cylinder, whereas most commercial triodes have helically wound grids. However, formula (138) gives results which are in fair agreement with the measured values of u, as is shown by the following table taken from the paper by Vogdes and Elder.[13]

TABLE XI.—CALCULATED AND OBSERVED VALUES OF u FOR TRIODES WITH HELICAL GRIDS

Tube	(a)	(b)	(c)
Calculated by Eq. (138)	5.97	16.7	34.5
Observed	6.25	18.0	34.0

The variation of u caused by a change in the position of the maximum density of the space charge is well shown in Fig. 79. This is a plot of the measured values of u along constant total-

space-current curves of Fig. 53, page 149. The value of u is practically constant along the constant-current lines until the region of low values of e_p is reached, when u goes through a series of fluctuations, as shown in Fig. 79. Some of the small variations are probably due to ionization of the residual gas in the tube. With some triodes, the value of u for negative plate voltages is greater than for positive plate voltages, as shown in Fig. 79, but this is not always the case.

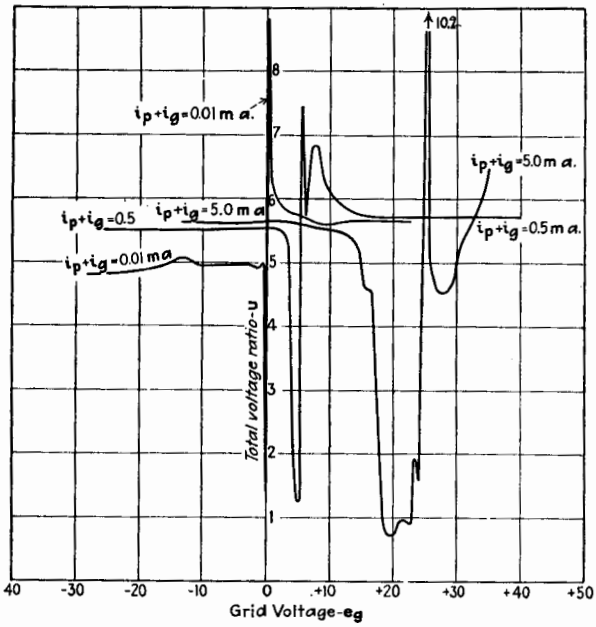


FIG. 79.—Values of u along lines of constant total space current (UV 201).

***81. Dependence of u on Dimensions of Electrodes.**—From a design standpoint, the extent to which u depends on the various dimensions of the electrodes is of interest. Using formulas (137) and (138), the ratio of the fractional variation in u to the fractional variation in any parameter can readily be calculated. This calculation has been made by Prof. R. G. Porter of Northeastern University, and the results are plotted in Figs. 80 to 86 for a particular set of parameters which correspond to a certain commercial triode. Figures 80, 81, and 82 give the results

* See Preface.

for a triode having approximately a plane grid and plate and a V-shaped filamentary cathode.

Figure 80 gives the variation of u and of $du/u \div db/b$ as b, the distance from grid to plate, is varied. The formulas and the

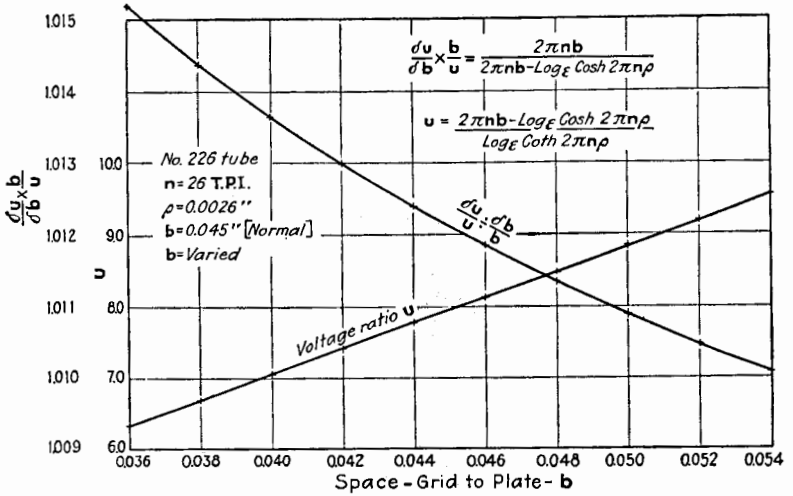


FIG. 80.—Variation of u with distance from grid to plate for a triode having plane electrodes.

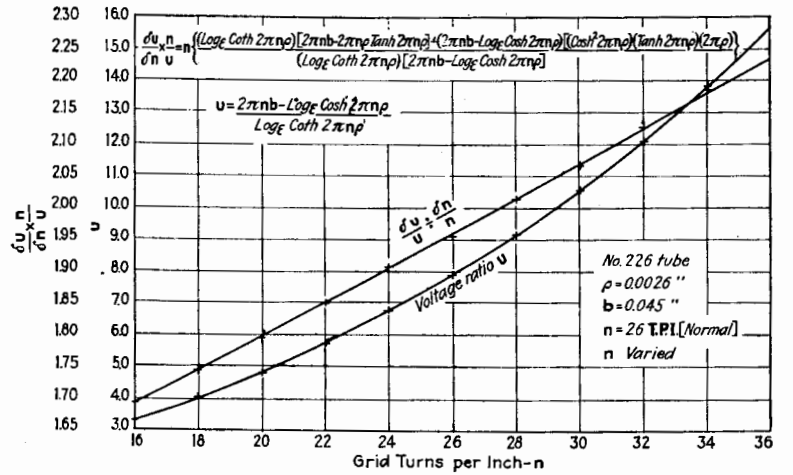


FIG. 81.—Variation of u with grid mesh for a triode having plane electrodes.

values of the parameters are given in the figures. As is evident both from the formula for u and from the graph, u varies nearly linearly with b.

Figure 81 gives the variation in u and in $du/u \div dn/n$ as the number of grid wires per inch, denoted by n, is varied. For the particular conditions of Fig. 81, the variation of u is nearly parabolic. This is shown by the fact that the percentage change in u is approximately twice the percentage change in n.

Figure 82 gives the variation in u and $du/u \div d\rho/\rho$ as the radius of the grid wire is varied.

The most important deduction from Figs. 80, 81, and 82 is that the factor n is the dominating factor in determining the

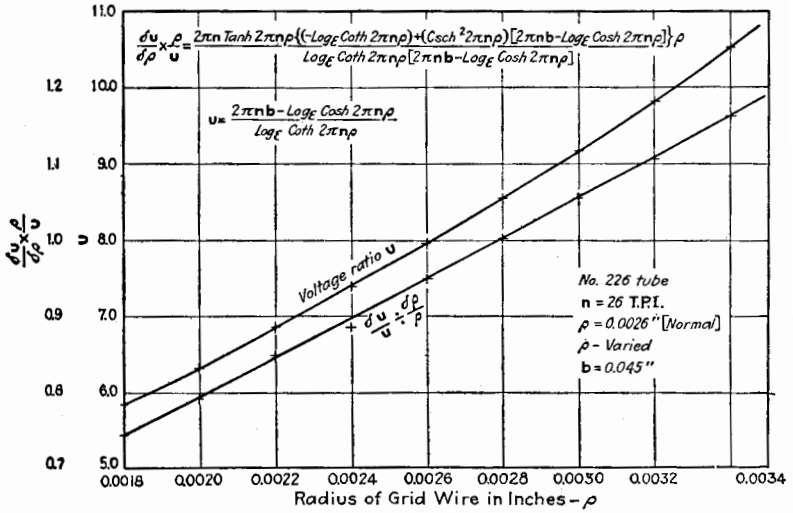


FIG. 82.—Variation of u with radius of grid wire for a triode having plane electrodes.

value of u. Whereas u varies approximately linearly with b and with ρ, it varies about as the square of n. For plane electrodes the distance from the cathode to the grid does not enter into the expression for u.

In the cylindrical arrangement of electrodes of a triode, the distance from cathode to grid and the distance from cathode to plate enter into the determination of the voltage ratio u. There are four factors in the expression for u as given in Eq. (138). The effect upon u of a variation of each of these factors is shown in Figs. 83, 84, 85, and 86. The scheme of plotting is similar to that used for Figs. 80, 81, and 82 and the four figures are self-explanatory.

When a triode has a cylindrical plate and grid, the grid mesh n has about as much effect on u as in the case of the plane arrange-

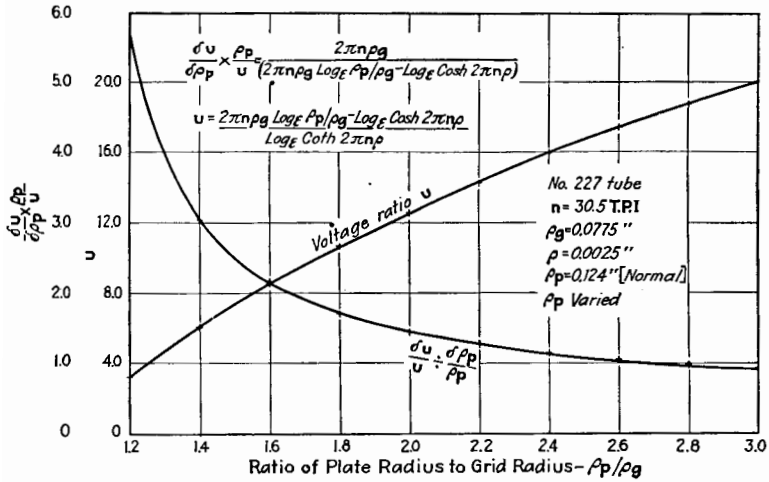


FIG. 83.—Variation of u with ratio of plate radius to grid radius for a triode having cylindrical electrodes. Plate radius varied.

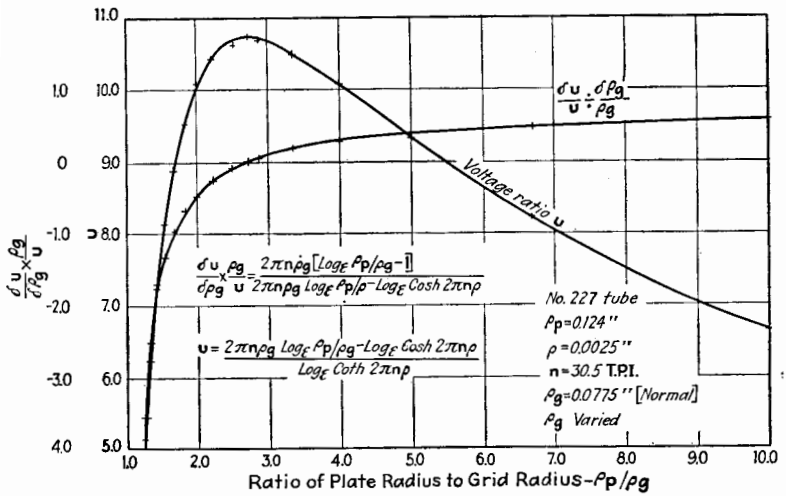


FIG. 84.—Variation of u with ratio of plate radius to grid radius for a triode having cylindrical electrodes. Grid radius varied.

ment of electrodes. The distances from the axis of the cathode to grid and to plate are much more important in their effects on u than in the plane-electrode triode. For example, u passes

through a maximum as ρ_g, the distance from the axis of the cathode to the grid, is varied. And u varies very rapidly as

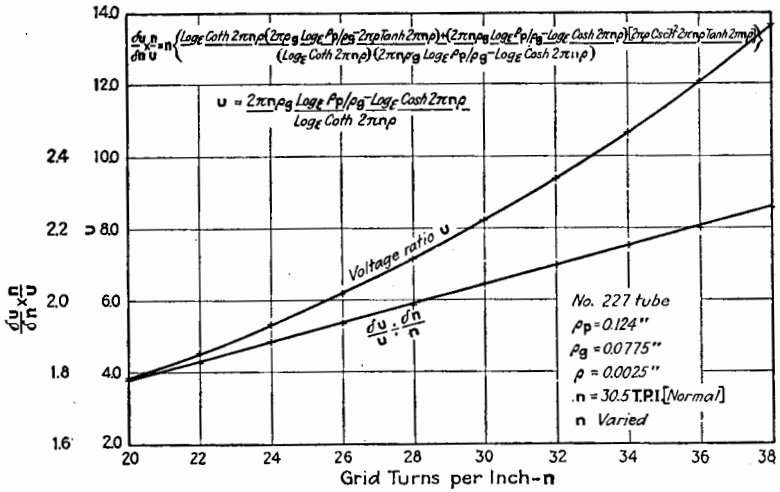


FIG. 85.—Variation of u with grid mesh for a triode having cylindrical electrodes.

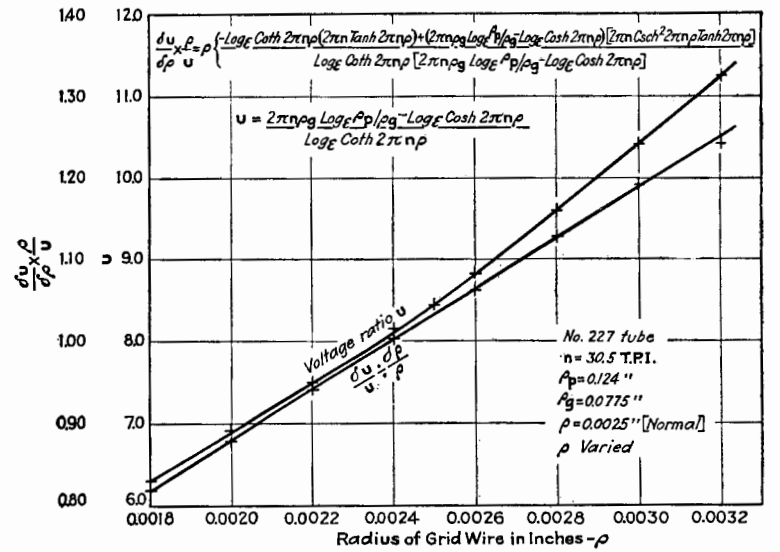


FIG. 86.—Variation of u with radius of grid wire for a triode having cylindrical electrodes.

ρ_g is varied, when ρ_p/ρ_g is much different from the value which gives a maximum value of u. The maximum value of u occurs

when $\log_\epsilon \frac{\rho_p}{\rho_g}$ equals unity, or when $\frac{\rho_p}{\rho_g}$ is equal to 2.72. Figure 83 shows that u also varies rapidly with ρ_p, for all values of the ratio ρ_p/ρ_g. It is evident that the best value of ρ_p/ρ_g is about 2.72, because for this value a small variation of ρ_g has no effect upon u.

The formulas (138) and (139), which are the basis of the discussion concerning the cylindrical arrangement, were deduced on the assumption of a filamentary cathode of very small radius in comparison with the radii of the grid and plate cylinders. In some commercial tubes using separately heated cathodes, the emitting cylindrical surface often has a radius which is a large fraction of the radius of the grid. It might be expected that the value of u for such a tube would deviate considerably from the value given by the formulas. In general such is not the case because, as in the plane-electrode arrangement, changing the distance from cathode to grid has little effect upon the ratio of the electrostatic field produced at the cathode by the charged plate and grid electrodes acting separately.

It is to be noted that formula (138) was deduced for a grid structure having wires parallel to the axis of the cylinder, whereas most commercial tubes have a helically wound grid structure. The value of u is practically the same for the two types of grids, provided the size of wire and the mesh are the same.

82. Effect on u of Non-uniform Dimensions of a Triode.—We have discussed the effects of the various parameters of a triode on the voltage ratio u, each parameter being assumed to have the same value for all portions of the structure. Very frequently, however, some one or more of the important dimensions of the elements of a triode may vary within a single triode. For example, the grid cylinder may not be exactly cylindrical but may be slightly conical, or the spacings between grid wires may vary along the grid so that n is not the same for the whole length of the grid structure. Such a variation in the dimensions causes the value of u to vary with the space current. That a variation in a dimension of a triode causes a variation of u with space current can easily be seen by the following example. Suppose that all dimensions of a particular triode are uniform except the grid mesh n, which is n_1 for half of the grid and n_2 for the other half, n_1 being less than n_2. Such a triode operates like two triodes in parallel, one having a voltage ratio u_1, corre-

sponding to n_1, and the other a voltage ratio u_2, corresponding to n_2. When the grid and plate voltages are such as to reduce to zero the current through the high-u section of the grid, current may flow through the other section, and the equivalent value of u for the tube is u_1. For larger values of the space current, both parts of the grid are operative and the equivalent value of u is greater than u_1.

The above example can be treated analytically as follows. Let the space current through the two portions of the grid be given by

$$i_1 = B_1(e_p + u_1 e_g)^{x_1} \tag{140}$$

$$i_2 = B_2(e_p + u_2 e_g)^{x_2} \tag{141}$$

where x_1 and x_2 may have any values, which usually range, however, from 1 to 2.

The total space current i through the tube is found by adding Eqs. (140) and (141). Since $u = \partial i/\partial e_g \div \partial i/\partial e_p$,

$$u = \frac{B_1 u_1 x_1 (e_p + u_1 e_g)^{x_1 - 1} + B_2 u_2 x_2 (e_p + u_2 e_g)^{x_2 - 1}}{B_1 x_1 (e_p + u_1 e_g)^{x_1 - 1} + B_2 x_2 (e_p + u_2 e_g)^{x_2 - 1}} \tag{142}$$

or

$$u = \frac{B_1^{\frac{1}{x_1}} i_1^{\frac{x_1 - 1}{x_1}} x_1 u_1 + B_2^{\frac{1}{x_2}} i_2^{\frac{x_2 - 1}{x_2}} x_2 u_2}{B_1^{\frac{1}{x_1}} i_1^{\frac{x_1 - 1}{x_1}} x_1 + B_2^{\frac{1}{x_2}} i_2^{\frac{x_2 - 1}{x_2}} x_2} \tag{143}$$

Equation (143) shows that the equivalent value of u depends on the current passed by the two sections of the grid.

We shall now make a more general analysis of a triode having variable dimensions or parameters. First, imagine the triode divided into sections by planes perpendicular to the length or axis of the elements and hence coincident with the lines of flow of the electrons. These sections are to be made so small that the various parameters in a section may be considered constant. The sections may differ owing to one or more of several reasons. Any of the dimensions of the electrodes may be different; the electrodes may be out of line or eccentric or deformed; the emission of the cathode may be different in the various sections owing to varying temperature or to the emission constants varying along the cathode. Let the subscripts 1, 2, 3, etc., denote the quantities pertaining to the several sections. Each section passes a certain plate current and a certain grid current and has certain differential parameters. For example, the mth section

passes a plate current i_{pm} and a grid current i_{gm} and has a transconductance s_{pm} and plate-voltage ratio u_{pm}. Then,

$$i_p = \Sigma i_{pm}, \text{ and } i_g = \Sigma i_{gm} \tag{144}$$

The equivalent transconductance is

$$s_p = \Sigma s_{pm} \tag{145}$$

The equivalent reflex transconductance is

$$s_g = \Sigma s_{gm} \tag{146}$$

The equivalent plate and grid variational conductances are

$$k_p = \Sigma k_{pm} \tag{147}$$

$$k_g = \Sigma k_{gm} \tag{148}$$

The equivalent plate-voltage ratio is, by Eq. (127),

$$u_p = \frac{\Sigma u_{pm} k_{pm}}{\Sigma k_{pm}} \tag{149}$$

and the equivalent reflex factor is, by Eq. (129),

$$u_g = \frac{\Sigma u_{gm} k_{gm}}{\Sigma k_{gm}} \tag{150}$$

Formulas (145) to (150) apply not only to triodes which vary within themselves but also to triodes in parallel. For example, two triodes in parallel act as a single triode having a plate-voltage ratio

$$u_p = \frac{u_{p_1} k_{p_1} + u_{p_2} k_{p_2}}{k_{p_1} + k_{p_2}},$$

an equivalent transconductance equal to the sum of the individual transconductances, and an equivalent plate conductance equal to the sum of the individual plate conductances.

The value of the equivalent voltage ratio given by Eq. (149) nearly always increases with increasing grid voltage. This can best be shown by differentiating Eq. (149) with respect to e_g. The result is

$$\frac{du_p}{de_g} = \frac{\left[\Sigma\left(k_{pm}\frac{du_{pm}}{de_g}\right) + \Sigma\left(u_{pm}\frac{dk_{pm}}{de_g}\right)\right] - \frac{\Sigma u_{pm} k_{pm}}{\Sigma k_{pm}} \Sigma \frac{dk_{pm}}{de_g}}{\Sigma k_{pm}} \tag{151}$$

If the values of u_{pm} be considered to be independent of e_g, the first term of Eq. (151) can be neglected, and

$$\frac{du_p}{de_g} = \frac{\sum k_{pm} \sum u_{pm} \frac{dk_{pm}}{de_g} - \sum u_{pm} k_{pm} \sum \frac{dk_{pm}}{de_g}}{(\Sigma k_{pm})^2} \tag{152}$$

To simplify the discussion, let us return to the consideration of two sections of a triode having different values of u and k, or of two triodes in parallel. Equation (152) becomes

$$\frac{du_p}{de_g} = \frac{(u_{p_1} - u_{p_2}) k_{p_2}^2 \frac{d}{de_g}\left(\frac{k_{p_1}}{k_{p_2}}\right)}{(k_{p_1} + k_{p_2})^2} \tag{153}$$

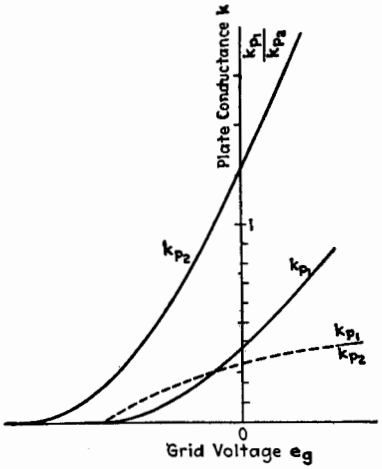


FIG. 87.—Ratio of plate conductances of two triodes.

Evidently from Eq. (153), u_p will not change with e_g if $u_{p_1} = u_{p_2}$, or if the ratio k_{p_1}/k_{p_2} is independent of e_g. Ordinarily the numerator of Eq. (153) is a positive quantity. If, for example, u_{p_1} is greater than u_{p_2}, k_{p_1} is less than k_{p_2} and will increase with increasing e_g, but later than k_{p_2}, as indicated in Fig. 87. It will be noted that k_{p_1}/k_{p_2} has generally a positive slope. However, the courses of the conductance curves may be such as to make u_p decrease with increasing grid voltage, but such a result can be obtained only by combining triodes of very different sizes and by making the polarizing voltages of the two grids different.

Curves A, B, and C of Fig. 88 show the voltage ratio of two triodes of about the same size. The equivalent voltage ratio when the two triodes are operated in parallel for the same and for different polarizing voltages, is shown in curves AC and AB. These curves are plotted against the polarizing voltage of one

of the triodes. Figure 89 gives the variation of u_p in a single triode, the grid of which was carelessly constructed so that the distance between grid wires varied along the grid.

The discussion given for two triodes can be extended to more than two triodes in parallel or can be applied to a single triode with several sections having different values of u_p.

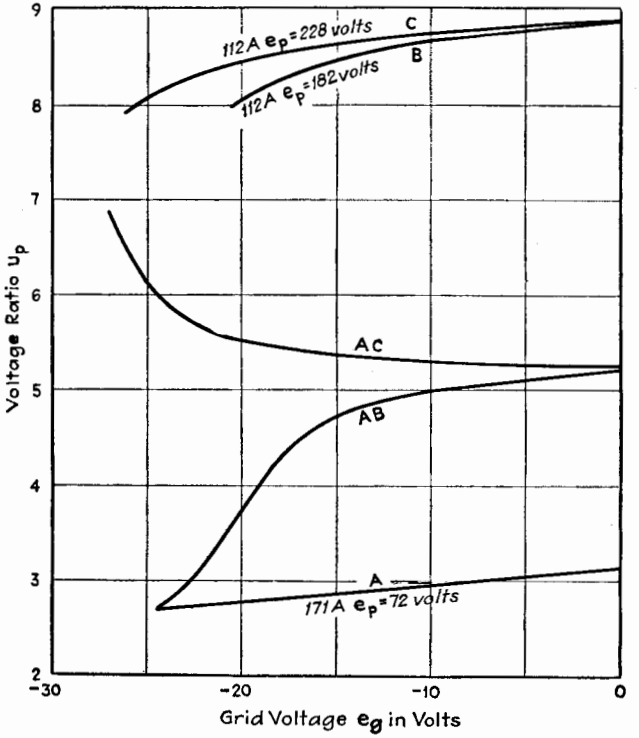


FIG. 88.—Resultant values of u_p for two triodes in parallel.

As pointed out, any irregularity in the structure of a triode may lead to a varying voltage ratio u, but the variation of u, as the grid voltage varies in a positive direction, is always an increase in u. Such a variation of u increases the wave-form distortion and is therefore undesirable in an audio amplifier. When a vacuum tube is used as a radio-frequency amplifier, wave-form distortion is not so important and a variation in u may be an advantage. Ballantine and Snow[14] have shown that

[14] BALLANTINE and SNOW, *Proc. I.R.E.*, **18**, 2102 (1930); *Radio Freq. Lab. Reprint* 22, 1930.

a variable-u tube may materially decrease the "cross talk," as it is called, in radio receivers. This cross talk is a modulation of a selected carrier wave by a modulated carrier wave having a different frequency. Since the variable-u feature is usually

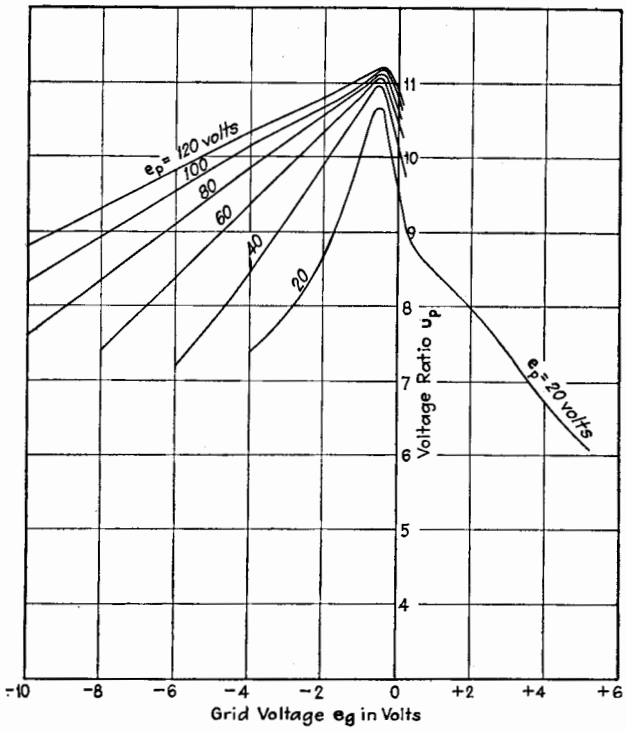


FIG. 89.—Voltage ratio of a triode having cylindrical electrodes and a separately heated cathode. Variable grid mesh.

incorporated in tetrodes and pentodes, its discussion will be deferred to a later chapter.

The reader interested in the design of vacuum tubes is referred to a valuable paper by Yuziro Kusunose,[15] Calculation of Characteristics and the Design of Triodes.

General References

GOETZ: *Phys. Zeits.*, **24**, 35 (1923).
LANGE: *Zeits. f. Hochfrequenztechnik*, **26**, No. 2 (1925).
SIXTUS: *Ann. d. Physik*, **3**, 1017 (1929).

[15] KUSUNOSE, *Proc. I.R.E.*, **17**, 1706 (1929).

For good *résumé* and bibliography of experiments on secondary emission in specially built tubes, see Compton and Langmuir, *Rev. Mod. Phys.*, **2,** 123 (1930).

BITTMANN: Der Einfluss der Sekundäremission auf der Röhren Kennlinien, *Ann. d. Physik*, **8,** 737 (1931).

GREVE: Untersuchungen über den Durchgriff von Empfängerröhren, *Diss. Jena*, 1930; *Zeits. f. Hochfrequenztechnik*, **38,** 234 (1931).

PATRUSCHEW: Die charakteristischen Flächen der Elektronenröhren, *Zeits. f. Hochfrequenztechnik*, **38,** 231 (1931).

MIURA: Graphical Representation of the Three Constants of a Triode, *Proc. I.R.E.*, **19,** 1488 (1931).

CHAPTER VIII

FUNDAMENTAL CONSIDERATIONS PERTAINING TO TRIODES

Before describing the uses of triodes, it is necessary to derive certain fundamental theorems which serve as the basis of almost all the analytical theory of the operation of triodes in their various applications. This chapter presents these theorems and certain other fundamental deductions which are necessary for a thorough understanding of the operation of triodes.

I. EQUIVALENT-CIRCUIT THEOREMS

When the grid potential of a triode is varied, the ratio of the instantaneous plate potential to the instantaneous plate current never remains constant. In other words, the instantaneous resistance $\bar{r}_p$ of the triode is a function of e_g. The plate circuit of a triode consists of a source of electrical power, usually having constant potential, in series with the vacuum tube which acts as a varying resistance, and some circuit in which the desired electrical effects are produced. When this tube resistance is caused to vary by changing the grid potential, the current sent through the circuit by the source of power is thereby caused to vary. Such an action is similar to that of a microphone transmitter, the resistance of which is caused to vary by the changing air pressure on the diaphragm. A circuit containing a *varying* resistance is not easily treated mathematically. However, by means of a fundamental theorem known as the *equivalent-circuit theorem*, the varying resistance may be replaced under certain conditions by a constant resistance and a varying e.m.f. This theorem is applicable to both the plate and grid circuits of a triode.

83. Equivalent-plate-circuit Theorem.[1]—Figure 90 represents a triode with impedances Z_c and Z_b in the grid and plate circuits.

[1] The first reference to this theorem which the author has found is in an article by J. M. Miller, *Proc. I.R.E.*, **6**, 143 (1918). In a footnote, Miller states that the theorem was suggested to him by H. H. Beltz. See also H. W. Nichols, *Phys. Rev.*, **13**, 404 (1919).

These impedances must have a conducting path to allow the steady currents to pass. The sources of electrical power in the grid and plate circuits are indicated by the batteries $\bar{E}_C$ and $\bar{E}_B$, conventionally shown with their negative terminals connected to the cathode. If the cathode is a filament along which there is a steady potential drop, the potential of the negative end

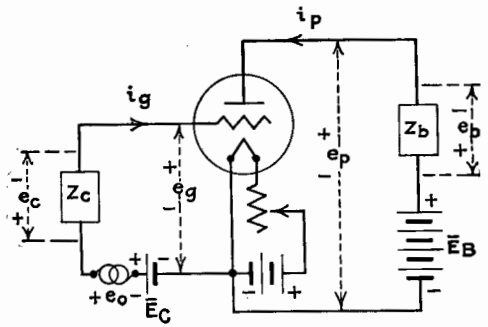


FIG. 90.—Circuits of the triode.

is taken as the potential of the cathode. The total instantaneous currents and potentials are shown in the conventional manner in Fig. 90.

The steady components of the grid and plate potentials are designated by $\bar{E}_g$ and $\bar{E}_p$, and the steady components of the grid and plate currents by $\bar{I}_g$ and $\bar{I}_p$. The grid voltage is caused to change by an e.m.f. indicated by e_0.

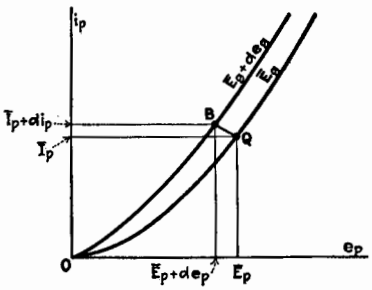


FIG. 91.

Figure 91 is the $i_p - e_p$ diagram for the triode, and if there is no varying potential impressed in the grid circuit, the operating point Q is determined by the steady components of the grid and plate potentials. Point Q is called the *quiescent* or *Q-point.*

Let the grid voltage e_g be given an increment de_g by reason of a small variation of e_0. The plate current and plate voltage change according to Eq. (124) is

$$di_p = k_p de_p + s_p de_g \qquad (124)$$

The operating point Q of Fig. 91 moves to some new position B depending on the relative magnitudes of di_p, de_p, and de_g. Equation (124) alone is not sufficient to determine the values of the increments, but the impedance Z_b and Eq. (124) together do determine the new point B. If e_b is the instantaneous potential across Z_b, it is clear from Fig. 90 that

$$\bar{E}_b - e_b - e_p = 0 \tag{154}$$

Hence,

$$de_p = -de_b \tag{155}$$

The relation between de_b and di_p is completely determined by the impedance Z_b. Combining Eqs. (124) and (155),

$$di_p = k_p(-de_b + u_p de_g) \tag{156}$$

Since $k_p = 1/r_p$,

$$r_p di_p + de_b = u_p de_g \tag{157}$$

Equation (157) shows that *small variations of plate current can be calculated by assuming that a fictitious e.m.f., equal to* $u_p de_g$, *acts in a circuit containing* Z_b *and a constant resistance equal to* r_p. This theorem, known as the *equivalent-plate-circuit theorem*, or the e-p-c. theorem, will also be given in other forms.

Equation (157) is a fundamental relation which cannot be reduced further until we know the relation between de_b and di_p, and the manner of variation of de_g. Usually e_g is a function of time, so we may write Eq. (157) in the equivalent form

$$r_p \frac{di_p}{dt} + \frac{de_b}{dt} = u_p \frac{de_g}{dt} \tag{158}$$

Equation (158) is of value in solving transient effects in vacuum-tube circuits. For most purposes it is not the most useful form.

Equation (157) is expressed in terms of increments of the total values of plate current, plate voltage, and grid voltage. Simplification results if we assume that Eq. (157) holds without appreciable error when these increments are finite though small. We may denote these small finite changes by Δi_p, Δe_b, and Δe_g, which are to be considered as a small current and small potentials measured from Q as a new origin of coordinates. Using this new system of measurement, we rewrite Eq. (157) in the form

$$r_p \Delta i_p + \Delta e_b = u_p \Delta e_g \tag{159}$$

If the impedance Z_b has resistance R_b, inductance L_b, and capacitance C_b, where these constants are the equivalent series values for any network denoted by Z_b, then the well-known expression for Δe_b is

$$\Delta e_b = R_b \Delta i_p + L_b \frac{d(\Delta i_p)}{dt} + \frac{\int (\Delta i_p) dt}{C_b} \tag{160}$$

Substituting Eq. (160) in Eq. (159),

$$L_b \frac{d(\Delta i_p)}{dt} + (R_b + r_p)\Delta i_p + \frac{\int \Delta i_p dt}{C_b} = u_p \Delta e_g \tag{161}$$

Equation (161) shows that *small* variations of plate current, as measured from the steady value of plate current, when caused by any kind of *small* variation in grid voltage, as measured from the steady value of grid potential, can be calculated by assuming that the tube is replaced by a *fictitious* e.m.f. $u_p \Delta e_g$ in series with a constant resistance numerically equal to the value of r_p at point Q. This is another statement of the e-p-c. theorem.

As an illustration of the principle just demonstrated, suppose that the grid voltage of a triode, the plate circuit of which contains inductance and resistance, is suddenly decreased by a small amount. The course of the plate current with respect to time can be calculated in the usual way for a circuit containing inductance L_b and resistance $R_b + r_p$. The current drops exponentially, approaching a new steady value, the time constant for the decrease of current being $\frac{L_b}{R_b + r_p}$. The important restriction must be kept in mind that the change in plate current, calculated by the application of the above simple rule, is true *only* when the changes in current and in potential are so small that r_p and u_p remain essentially constant. If the applied change of grid potential is large, Eq. (158) must be used, in which r_p and u_p are not constant.

As another illustration of the application of Eq. (161), suppose Δe_g is a small sinusoidal e.m.f. of the form $\widetilde{\Delta e}_g = \sqrt{2}\Delta E_g \sin \omega t$, and we desire the steady-state value of the alternating component of the plate current. If ΔE_g is small, the alternating component of plate current can be calculated by assuming that a fictitious voltage $u_p \Delta E_g$ acts in a circuit of impedance $\mathbf{Z}_b + r_p$. Hence,

$$\mathbf{\Delta I}_p = \frac{u_p \mathbf{\Delta E}_g}{r_p + \mathbf{Z}_b} \tag{162}$$

The root-mean-square value of the alternating component of plate current is

$$\Delta I_p = \frac{u_p \Delta E_g}{\sqrt{(R_b + r_p)^2 + X_b^2}} \tag{163}$$

The Δ sign may be omitted if it be remembered that the e-p-c. theorem can be applied *only* when the alternating components of currents and potentials are small or when, with larger variations, there are no appreciable changes in the values of the tube coefficients. The latter restriction implies that the operating point is on a plane region of the plate-current characteristic surface.

84. Equivalent-grid-circuit Theorem.[2]—The derivation of the equivalent-grid-circuit theorem or e-g-c. theorem is exactly similar to the derivation of the e-p-c. theorem. Equations (125) and (129), pages 165 and 166, combine to give

$$di_g = k_g(de_g + u_g de_p) \tag{164}$$

Adding e.m.f.'s in the grid circuit of Fig. 90,

$$e_0 - e_c - e_g + \bar{E}_c = 0 \tag{165}$$

Hence,

$$de_g = de_0 - de_c \tag{166}$$

where e_0 is an impressed e.m.f. from an outside source. Combining Eqs. (164) and (166), using r_g as the reciprocal of k_g,

$$r_g di_g + de_c = de_0 + u_g de_p \tag{167}$$

The differential equation for the grid circuit, which corresponds to Eq. (158), is obtained by dividing Eq. (167) by dt, giving

$$r_g \frac{di_g}{dt} + \frac{de_c}{dt} = \frac{de_0}{dt} + u_g \frac{de_p}{dt} \tag{168}$$

Equation (167) shows that *a small variation in grid current can be calculated by assuming that an equivalent voltage,* $de_0 + u_g de_p$,

[2] The author has been unable to find any derivation or statement of the e-g-c. theorem. Van der Bijl, in his book, "The Thermionic Vacuum Tube and Its Applications," p. 187, McGraw-Hill Book Company, Inc., New York, 1920, gives the equivalent of Eq. (125). Latour in *Electrician*, December, 1916, also derives Eq. (125).

Nichols, *Phys. Rev.*, **13**, 404 (1919), refers to the variation of grid current due to the variation of plate voltage which can be expressed in terms of a voltage introduced into the grid circuit.

Llewellyn takes fully into account the reaction of the plate voltage on the grid circuit, but does not derive or state this theorem (*Bell Tel. Lab. Reprint B-20*).

acts in a circuit containing Z_c and r_g. This is the most general statement of the e-g-c. theorem.

Following the same procedure adopted in developing the e-p-c. theorem, we may extend Eq. (167), without much error, to apply to small finite currents and potentials Δi_g, Δe_0, Δe_c, and Δe_p, measured from the operating point of the grid-current surface. Thus Eq. (167) becomes

$$r_g \Delta i_g + \Delta e_c = \Delta e_0 + u_g \Delta e_p \tag{169}$$

If Δe_c is expressed in terms of its components across the parts of the impedance Z_c,

$$\Delta e_c = R_c \Delta i_g + L_c \frac{d(\Delta i_g)}{dt} + \frac{\int (\Delta i_g) dt}{C_c} \tag{170}$$

which gives, combined with Eq. (169),

$$L_c \frac{d(\Delta i_g)}{dt} + (R_c + r_g)\Delta i_g + \frac{\int (\Delta i_g) dt}{C_c} = \Delta e_0 + u_g \Delta e_p \tag{171}$$

In Eq. (171), Δi_g is a small finite current and Δe_0 and Δe_p are small finite potentials measured from the steady values of the total current and potentials. They are, therefore, small variable current and potentials, measured from a new origin of coordinates, the Δ sign being unnecessary except to indicate that the theorem applies only to such small current and potential changes that r_g and u_g are essentially constant. Equation (171) is another way of stating mathematically the e-g-c. theorem.

If Δe_0 is a small sinusoidal e.m.f. of effective value ΔE_0, solving Eq. (171) gives

$$\mathbf{\Delta I}_g = \frac{\mathbf{\Delta E}_0 + u_g \mathbf{\Delta E}_p}{r_g + \mathbf{Z}_c} \tag{172}$$

When the grid of a triode is polarized negatively, the grid current is zero and r_g is infinite.

85. Equivalent Circuits of the Triode.—The two theorems just derived show that if the variations of potentials and currents in the circuits of a triode are small, so that the triode coefficients are essentially constant over the range of variation, the changes can be calculated by considering them apart from the steady components of currents and potentials. This permits us to draw separate equivalent-circuit diagrams for the separate components of the currents and potentials.

Figure 92 shows the equivalent circuits of a triode for the steady components of the electrical quantities. The circles marked G, P, and F indicate the grid, plate, and filament terminals of the triode. The resistances $\bar{R}_c$ and $\bar{R}_b$ are the resistances for steady current of the external circuits. The steady-current resistances within the triode are $\bar{r}_g$ and $\bar{r}_p$. The currents $\bar{I}_g$

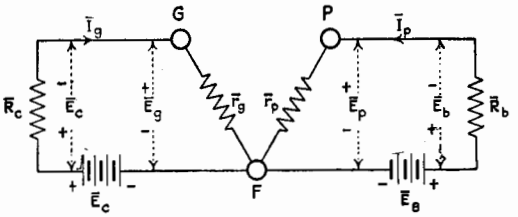


FIG. 92.—Equivalent circuits of the triode for the steady components of current and potential.

and $\bar{I}_p$ are determined by Ohm's law, but, since $\bar{r}_g$ and $\bar{r}_p$ are functions of $\bar{E}_g$ and $\bar{E}_p$, the grid and plate currents are determined graphically.

Figure 93 shows the equivalent circuits for small changes of the electrical quantities, these changes being determined by the way in which the impressed e.m.f. Δe_0 varies with time, and by the constants of the circuits as given on the diagram. The convention of signs as used throughout the discussions in this book is

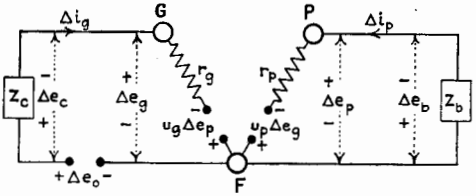


FIG. 93.—Equivalent circuits of the triode for the varying components of current and potential. Electrode capacitances neglected.

indicated. The internal variational resistances of the triode, r_g and r_p, are functions of $\bar{E}_g$ and $\bar{E}_p$ and hence are obtained from the solution of the steady-current case of Fig. 92. The grid and plate circuits are linked by the fictitious e.m.f's. $u_p \Delta e_g$ and $u_g \Delta e_p$ and therefore must be solved simultaneously. Fortunately, however, under the conditions of ordinary use of a triode, the grid resistance is usually infinite and hence the grid current is zero. It is important to note that all effects due to the internal capacitances

of the triode are neglected, so that the circuit diagram of Fig. 93 is applicable only when the rates of change of the electrical quantities are so low that these capacitances have no appreciable effect. In Chap. XI the equivalent circuits will be treated when the capacitance effects are included.

If the impressed e.m.f. Δe_0 is sinusoidal, with r.m.s. value ΔE_0, the r.m.s. currents are calculated by the use of the equivalent circuits of Fig. 94. Here also, the effects of the capacitances of the triode are neglected, which limits the use of the circuits of Fig. 94 to low frequencies. The complete equivalent circuit is

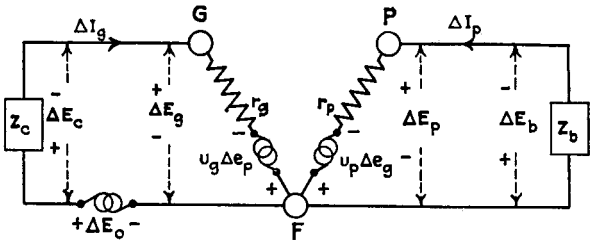


FIG. 94.—Equivalent circuits of the triode for a-c. components of current and potential. Electrode capacitances neglected.

obtained by placing capacitances between each pair of electrode terminals. The complete circuit is discussed in Chap. XI.

86. Alternative Equivalent Circuits of a Triode with Constant-current Generator.—An alternative equivalent circuit, which involves a constant-current generator, can be derived easily from the equivalent-circuit theorems already given. This alternative circuit was first suggested probably by Mayer[3] in 1926 and later strongly recommended by Bligh.[4] It leads to a considerable simplification of some problems, as will presently appear.

Dividing Eq. (157) by r_p,

$$di_p = \frac{u_p de_g}{r_p} - \frac{de_b}{r_p} \tag{173}$$

or

$$s_p de_g = di_p + \frac{de_b}{r_p} \tag{174}$$

The equivalent plate circuit of Fig. 95 is consistent with Eq. (174). The equivalent internal series circuit of Fig. 93 is replaced

[3] MAYER, *Tel. Fernspr. Tech.*, November, 1926.
[4] BLIGH, *E.W. and W.E.*, **7**, 480 (1930).

by a resistanceless generator of voltage de_b which delivers a current $s_p de_g$ to the constant resistance r_p in parallel with the plate load Z_b.

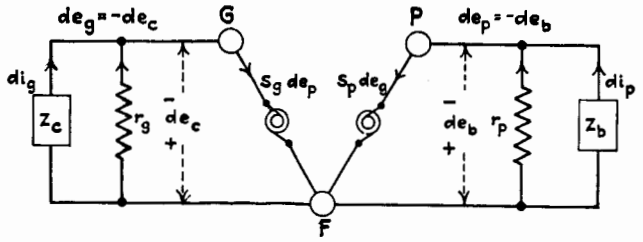


FIG. 95.—Equivalent circuits of the triode with resistanceless generators. Electrode capacitances neglected. Grid side valid only when $de_0 = 0$.

Dividing Eq. (167) by r_g, and assuming de_0 to be zero, we derive the grid-circuit equation. Thus

$$s_g de_p = di_g + \frac{de_c}{r_g} \tag{175}$$

The equivalent grid circuit corresponding to Eq. (175) is shown in Fig. 95. It comprises a resistanceless generator of voltage de_c which delivers a current $s_g de_p$ to the constant resistance r_g in parallel with the grid load Z_c. If de_0 is not zero, the simplifica-

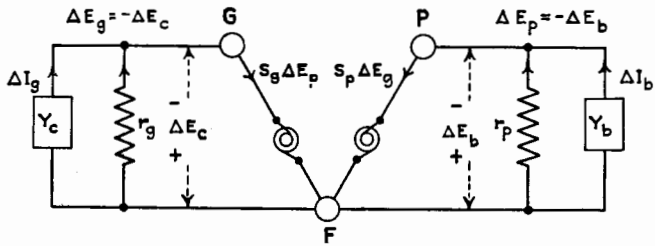


FIG. 96.—Equivalent circuits for alternating currents of the triode with resistanceless generators. Electrode capacitances neglected. Grid side valid only when $\Delta E_0 = 0$.

tion made possible by the use of a resistanceless generator, as illustrated in Fig. 95, does not hold in the grid circuit; the equivalent circuits of Fig. 93 should then be used.

It is evident from the derivation of the equivalent circuits of Fig. 95 that they are valid for transient currents as well as for sinusoidal alternating currents. The equivalent circuits for alternating currents are shown in Fig. 96, which corresponds to the equivalent series circuits of Fig. 94.

The advantage of the equivalent circuits of Figs. 95 and 96 arises from the fact that all of the circuits are in parallel. Since admittances rather than impedances are simpler in dealing with parallel circuits, the circuit elements in Fig. 96 are indicated as admittances. Equations (174) and (175), when expressed in admittances and applied to a-c. problems, become

$$s_p \Delta \mathbf{E}_g = (k_p + \mathbf{Y}_b)\Delta \mathbf{E}_b \tag{176}$$

and

$$s_g \Delta \mathbf{E}_p = (k_g + \mathbf{Y}_c)\Delta \mathbf{E}_c \quad (\Delta \mathbf{E}_0 = 0) \tag{177}$$

An extension of the principles given for triode circuits, in which the currents through the interelectrode capacitances are included, is given in Chap. XI.

II. PATH OF OPERATING POINT

In gaining a thorough understanding of the operation of a device, graphical methods are often of greater assistance than a purely mathematical analysis; as, for instance, a shunt-wound generator has a characteristic curve which cannot easily be expressed in mathematical form. The performance of the shunt generator with any specified load is expressed graphically. Similarly, a triode is a device whose performance is often much more easily explained by diagrams than by equations.

Consider in somewhat greater detail a few of the diagrams showing the path on the characteristic plate and grid surfaces over which the operating points move under various conditions of operation.

87. Steady-current Problem.—For the present, attention will be confined to the plate circuit. The treatment of the steady-current problem of Fig. 92 for the determination of the position of the Q-point is as follows: The curves of Fig. 97 represent the $i_p - e_p$ characteristics of a triode. From a point on the voltage axis, determined by the voltage $\bar{E}_B$ of the plate-circuit battery, a straight line ad is drawn making an angle with the vertical whose tangent, taking account of scales, is equal to the resistance $\bar{R}_b$. The intersection of the line ad with the particular characteristic curve determined by the steady grid potential $\bar{E}_g$ gives the quiescent point Q. Point Q determines the steady plate current $\bar{I}_p$ and the steady plate voltage $\bar{E}_p$. A line drawn from Q to the origin O makes an angle with the vertical axis

whose tangent equals $\bar{r}_p$. It is obvious that $\bar{I}_p$, thus determined, fulfills the circuit requirements of Fig. 92.

In this method of graphical solution for a series circuit, we have superimposed on the characteristic curves of one device, the tube,

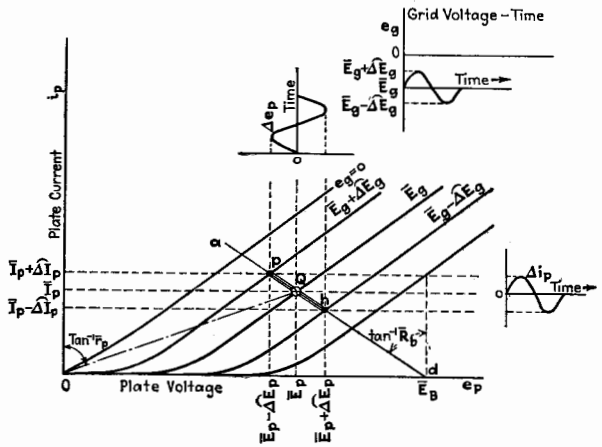


FIG. 97.—Path of operation on the $i_p - e_p$ chart. Pure resistance as plate load. Path on the plane region of the characteristic surface. $\Delta e_g = \widehat{\Delta E_g} \sin \omega t$.

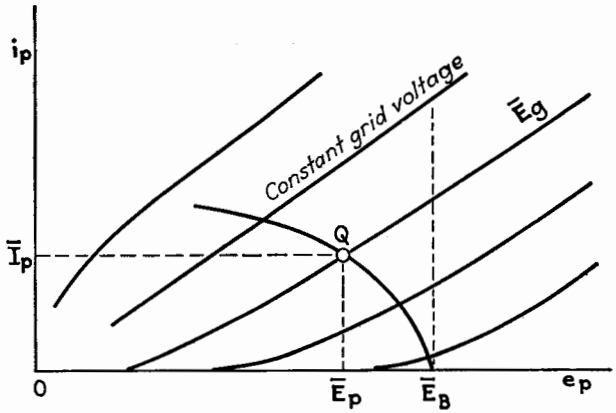


FIG. 98.—Plate-circuit diagram for a triode with a tungsten lamp as plate load.

the characteristic curve of the second device, the external resistance, this curve being reversed and displaced along the voltage axis by the amount of the steady applied voltage $\bar{E}_B$. Line *ad* is the characteristic curve of the resistance $\bar{R}_b$, plotted backward from point $\bar{E}_B$ on the voltage axis as origin, so that

voltages across the resistance are measured to the left from point $\bar{E}_B$. This method of plotting is necessary to make distances on the voltage axis add according to the relation $\bar{E}_B = \bar{E}_p + \bar{E}_b$.

If the external plate circuit contains a device having any steady-current characteristic other than a straight line, the Q-point, and hence the steady current, can be obtained by the same method. For example, suppose that a tungsten lamp is included in the plate circuit. Because of the rise of temperature of the filament with current, the characteristic graph of the lamp is a curve. The characteristic of the lamp is plotted in a reverse direction from point $\bar{E}_B$, as shown in Fig. 98. The Q-point is the intersection of the characteristic curve for the lamp and the triode curve corresponding to constant grid voltage $\bar{E}_g$.

88. Varying-current Problems. $\bar{R}_b = \bar{R}_b$. Q-point on the Plane Region of the Characteristic Surface.—Having determined the steady position of the operating point, we now find the path over which this point moves when the grid voltage is varied. The first and simplest case to consider is a triode having a pure resistance R_b connected in the plate circuit, the resistance being the same for steady current and alternating currents of all frequencies. The triode has plate- and grid-polarizing voltages such that the Q-point is near the center of the plane portion of the characteristic surface. If the grid voltage is varied, the operating point must move along the resistance line, as shown in Fig. 97, which illustrates the case now being considered. If the grid-voltage variation is so small that the operating point does not leave the plane portion of the characteristic surface, the *actual path of operation on the characteristic surface is a straight line*, and all variables are linearly related. The path of operation as projected on each of the three mutually perpendicular planes through the axes is, in this case, also a straight line. The projection of the path on the $i_p - e_p$ plane, shown in Fig. 97, is ph and is determined in length by the grid-voltage variation. The projection of the actual path on the $i_p - e_g$ plane is shown by the straight line ph in Fig. 99, and by a similarly marked line on the $e_p - e_g$ plane in Fig. 100.

If, as assumed in Figs. 97, 99, and 100, the grid-voltage variation is sinusoidal, as shown in the auxiliary diagram of Fig. 97, the plate-current and plate-voltage variations are also sinusoidal, as is indicated in the figures. The *average* value of the grid potential is the same as the quiescent value, so that in this wholly

linear case the average values of the plate current and plate potential are also identical with the quiescent values. Since in more complicated cases it will be necessary to distinguish between average and quiescent values, we shall indicate the quiescent point, or Q-point, by a circle and the average point, or A-point, by a cross.

If the grid-voltage variations contain a steady component or contain even harmonics instead of being sinusoidal as was assumed in Figs. 97, 99, and 100, the average value of the grid

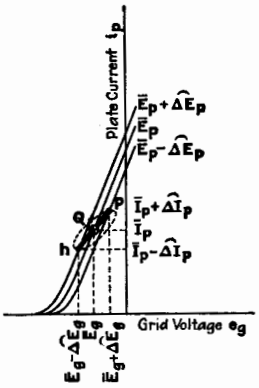


FIG. 99.—Path of operation on the $i_p - e_g$ chart. Path on the plane region of the characteristic surface. $\Delta e_g = \Delta\hat{E}_g \sin \omega t$.

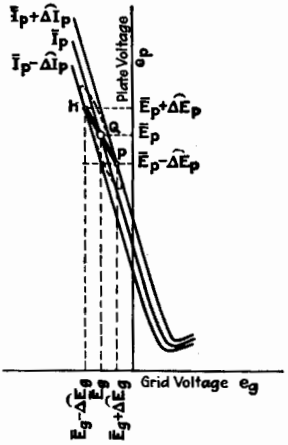


FIG. 100.—Path of operation on the $e_p - e_g$ chart. Path on the plane region of the characteristic surface. $\Delta e_g = \Delta\hat{E}_g \sin \omega t$.

voltage is different from the quiescent value and the average point does not coincide with the quiescent point. Since, however, the actual path on the characteristic surface is straight and consequently all variables are linearly related, the average point lies on the characteristic surface and hence on the straight-line paths on all three diagrams.

89. Varying-current Problem. $\bar{R}_b \neq \tilde{R}_b$. Q-point on the Plane Region of the Characteristic Surface.—Frequently, the external plate circuit offers a very different resistance to alternating currents than to steady currents. Let $(R_b)_\omega$ be the resistance offered to alternating currents of frequency $\omega/2\pi$, and let $\bar{R}_b$ as before represent the resistance offered to steady currents. Suppose that the Q-point is situated on the plane portion of the

characteristic surface, as assumed in the cases just discussed, and that the grid-voltage variations are sinusoidal. The Q-point is determined as before, but the alternating variations follow a

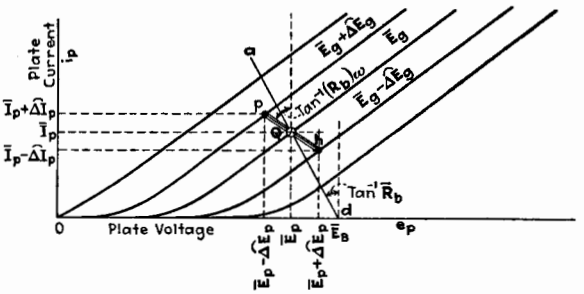


FIG. 101.—Path of operation on the plane region of the characteristic surface. Resistance load. $\bar{R}_b \neq \bar{R}_b$, $\Delta e_g = \hat{\Delta E}_g \sin \omega t$.

path ph, which passes through Q but makes an angle with the vertical whose tangent equals $(R_b)_\omega$, as shown in Fig. 101 (see Appendix A). The average and quiescent points coincide in this case.

If, on the other hand, the small grid-voltage variations are nonsinusoidal, as shown in the auxiliary diagram of Fig. 102, the

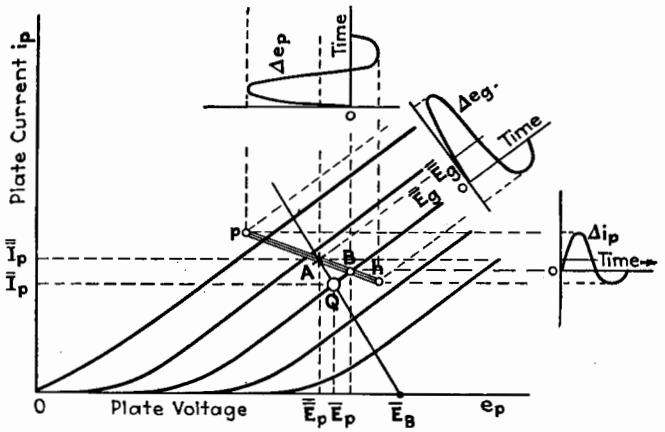


FIG. 102.—Path of operation on the plane region of the characteristic surface. Resistance load. $\bar{R}_b \neq \bar{R}_b$. Grid voltage nonsinusoidal.

grid-voltage variations can be resolved into an average value $\overline{\Delta E}_g$ and certain sinusoidal components. The steady component changes the steady grid-polarizing potential and hence moves the steady-current resistance line to the average point A. If the

external resistance $(R_b)_\omega$ is the same for all a-c. components of the grid-voltage variations, the a-c. path is through A, as indicated by ph of Fig. 102. This average point lies on the characteristic surface of the triode because all variables are linearly related, and hence the curve for average grid voltage $\bar{\bar{E}}_g$ passes through A. In this case the Q-point does not lie on the path of operation. The instantaneous values of plate current and plate voltage, corresponding to the instantaneous value of grid voltage numerically equal to $\bar{E}_g$ (when $\Delta e_g = 0$), are determined by a new point B on the path of operation. Points Q, A, and B are all on the characteristic surface of the triode.

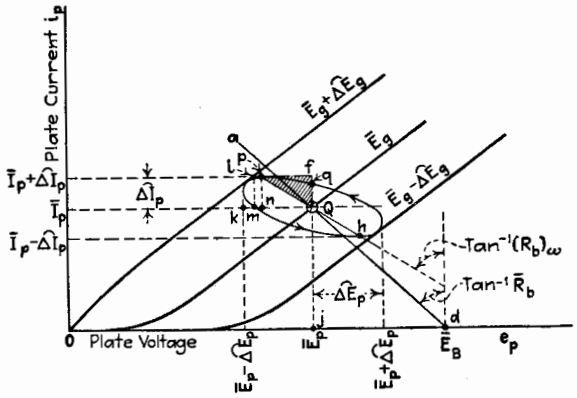


FIG. 103.—Path of operation on the plane region of the characteristic surface. $\bar{R}_b \neq \tilde{R}_b$ and $X_b \neq 0$. $\Delta e_g = \widehat{\Delta E}_g \sin \omega t$. Arrows give direction for capacitive reactance.

90. Varying-current Problem When Z_b Has a Resistance, $\bar{R}_b \neq \tilde{R}_b$, and a Reactance. Q-point on the Plane Region of the Characteristic Surface.—Consider the path followed by the operating point when the path is still located entirely on the plane portion of the plate-current characteristic surface, but when the impedance Z_b has a resistance to steady current of $\bar{R}_b$, an equivalent resistance $(R_b)_\omega$ to alternating current of frequency $\omega/2\pi$, and a reactance X_b for the same frequency. If the alternating component of the plate current is $\Delta i_p = \widehat{\Delta I}_p \sin \omega t$,

$$\begin{aligned}\Delta e_p = -\Delta e_b &= -\widehat{\Delta I}_p \sqrt{(R_b)_\omega^2 + X_b^2} \sin\left(\omega t + \tan^{-1}\frac{X_b}{(R_b)_\omega}\right) \\ &= -\widehat{\Delta I}_p (R_b)_\omega \sin \omega t - \widehat{\Delta I}_p X_b \cos \omega t \\ &= -(R_b)_\omega \Delta i_p - X_b \sqrt{\widehat{\Delta I}_p^2 - \Delta i_p^2}\end{aligned}$$

Transposing and squaring,

$$\Delta e_p^2 + 2(R_b)_\omega \Delta e_p \Delta i_p + Z_b^2 \Delta i_p^2 = X_b^2 \widehat{\Delta I}_p^2 \qquad (178)$$

Equation (178) is the equation of the path of the operating point expressed in terms of coordinates measured from Q as origin. The path is an ellipse, as shown in Fig. 103. The operating point travels in a clockwise direction if X_b is positive or inductive, and in a counterclockwise direction if X_b is negative or capacitive. The quiescent and average points are coincident.

If Δi_p in Eq. (178) is given its maximum value $\sqrt{2}\Delta I_p$, Δe_p is equal to $(R_b)_\omega \widehat{\Delta I}_p$. Therefore, in Fig. 103, $\overline{Qn} = (R_b)_\omega \widehat{\Delta I}_p$. Hence, line ph, determined as before by the resistance $(R_b)_\omega$, intersects the elliptical path at its points of tangency with the horizontal lines $\bar{I}_p + \widehat{\Delta I}_p$ and $\bar{I}_p - \widehat{\Delta I}_p$.

Again, if, in Eq. (178), Δi_p is zero, Δe_p is equal to $X_b \widehat{\Delta I}_p$, showing that the horizontal intercept $\overline{Qm}$ is equal to $X_b \widehat{\Delta I}_p$. Furthermore, the distance $\overline{Qk}$ is equal to $\widehat{\Delta E}_p = (Z_v)_\omega \widehat{\Delta I}_p$. It is clear that $\overline{Qm}/\overline{Qn} = X_b/(R_b)_\omega = \tan\theta$, and $\overline{Qn}/\overline{Qk} = (R_b)_\omega/(Z_b)_\omega = \cos\theta$, where θ is the phase angle of $(Z_b)_\omega$. The vertical intercept $\overline{Qq}$ is equal to $X_b/(Z_b)_\omega \cdot \widehat{\Delta I}_p$.

The actual path on the characteristic surface, the projection of which on the $i_p - e_p$ plane is shown in Fig. 103, is an ellipse and has elliptic projections upon the two other planes, as shown by the dotted ellipses of Figs. 99 and 100. In all three ellipses the direction indicated by the arrow is the direction of motion of the operating point for capacitive reactance. The direction for capacitive reactance is opposite to that for inductive reactance.

91. Varying-current Problem. $\bar{R}_b = \tilde{R}_b$. Q-point on a Curved Region of the Characteristic Surface.—We shall now study a more difficult case, in which the path of the operating point is on a curved region of the plate-current characteristic surface. The grid voltage is assumed to vary sinusoidally about a steady value $\bar{E}_g$ and, in order to accentuate the effects, the grid-voltage variations are not restricted to small amplitudes. The $i_p - e_p$ diagram assumes the form shown in Fig. 104, where $\bar{E}_B$ is the steady potential of the plate battery. Line ad, as before, is laid off making an angle with the vertical whose tangent is equal to $\bar{R}_b$, the resistance of the impedance to a steady current. The quiescent plate current is determined as before by the intersection of the resistance line ad and the $i_p - e_p$ curve corresponding to the quiescent grid voltage $\bar{E}_g$. Two positions of Q

are shown in Fig. 104, corresponding to two assumed values of $\bar{E}_g$. One position is in the lower curved region of the characteristic surface and the other is in the curved region due to filament saturation. If the plate impedance is a pure resistance, having the *same* value for alternating current as for steady current, the projection of the path of operation upon the $i_p - e_p$ diagram is a *straight line* through Q and coincident in direction with line ad. The extent of travel ph on the $i_p - e_p$ diagram is determined by the amplitude of variation of the grid potential. Although the grid potential varies sinusoidally, the plate current is nonsinusoidal because of the *nonlinear* relation between the variables caused by the curvature of the characteristic surface. Plots of the plate

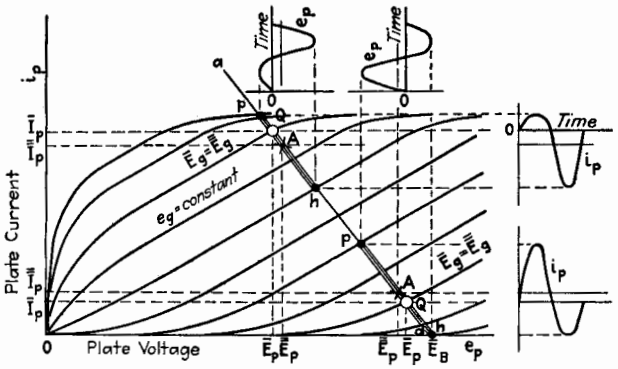


FIG. 104.—Paths of operation on curved regions of the characteristic surface. Pure resistance as plate load. $R_b = R_b$. $e_g = \bar{E}_g \sin \omega t$.

current and plate potential with respect to time are shown also in Fig. 104.

The quiescent operating point is represented in Fig. 104 by the circles on the line ad. When the operating point moves over the path ph, the quiescent values of plate current and plate voltage are not the average values. Since i_p and e_p are linearly related (see Appendix A), the point determined by the average values of e_p and i_p lies on the path ph. The relations between i_p and e_g and between e_p and e_g are not linear, however, so that the average point does not lie on the path when it is projected upon the other two planes. Therefore, the *point determined by average values, i.e., by d-c. instrument readings, does not lie upon the characteristic surface when the portion of the surface traversed by the operating point is curved.* In Fig. 104, $\bar{E}_g$ is the average

value of the grid voltage, but the $i_p - e_p$ curve for $\bar{\bar{E}}_g$ does not pass through the point A which is determined by $\bar{\bar{I}}_p$ and $\bar{\bar{E}}_p$.

92. Varying-current Problem. $\bar{R}_b \neq \tilde{R}_b$. Q-point on a Curved Region of the Characteristic Surface.—As a final case in the study of the path of operation, assume the external plate circuit to offer the same pure resistance to all alternating components of the plate current, but the a-c. resistance $\tilde{R}_b$ to be different from the steady-current resistance $\bar{R}_b$. The Q-point, assumed to be on a curved portion of the characteristic surface, is determined as before. Because of the nonlinearity of the actual path, there are average components of the changes in

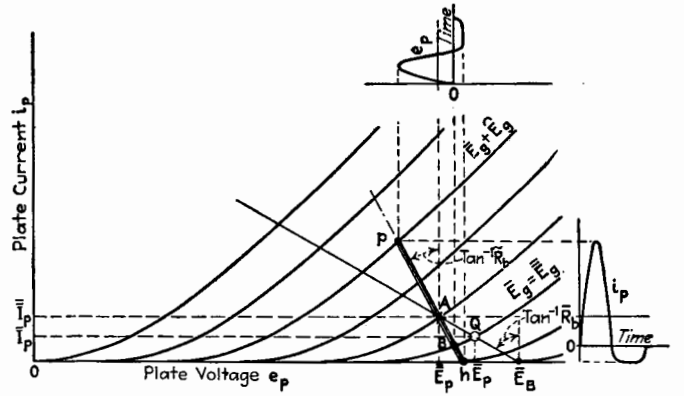


FIG. 105.—Path of operation on a curved region of the characteristic surface $\bar{R}_b \neq \tilde{R}_b$. $e_g = \bar{E}_g \sin \omega t$.

plate current and in plate potential. These average-value increments determine the average point A on the steady-current resistance line. The path of operation is a straight line ph which passes through A projected on the $i_p - e_p$ diagram shown in Fig. 105. Since the variables are not linearly related, point A does *not* lie on the characteristic surface of the triode. Hence, in the $i_p - e_g$ and $e_p - e_g$ diagrams for this case, point A does *not* lie on the projected path of operation which is curved instead of straight. If the grid-voltage variations are sinusoidal, $\bar{E}_g$ is the average value of grid voltage, but that A is not on the characteristic surface is shown by the fact that the curve for $\bar{E}_g$ does not pass through A. The instantaneous values of plate current and plate potential when Δe_g is zero are determined by point B, the intersection of the curve for $\bar{E}_g$ with the path of operation. Points Q and B are on the characteristic surface.

The reader may be interested in proving that the Q-point lies on the path of operation on the $i_p - e_p$ diagram only when X_b is zero for all components of the plate current, and when either $\bar{R}_b = \tilde{R}_b$ or $\bar{I}_p = \tilde{I}_p$.

93. Varying-current Problem. Effect of Grid-circuit Resistance. $\bar{R}_c = \tilde{R}_c$, $\bar{R}_b = \tilde{R}_b$.—The operating paths of the preceding discussion have been derived with no reference to the grid circuit, since it was assumed that the grid-potential variation was known. If, however, the grid circuit has impedance connected in series with the sinusoidal e.m.f. e_0, the grid-potential variation is no longer sinusoidal, and its value at any instant depends on the drop of potential through the grid-circuit impedance. In order to derive the plate-current wave form, the grid circuit as well as the plate circuit must be considered. The derivation of the exact wave form for the plate and grid currents is in most cases a difficult problem. The solution for one special simple case follows as an illustration of the method.

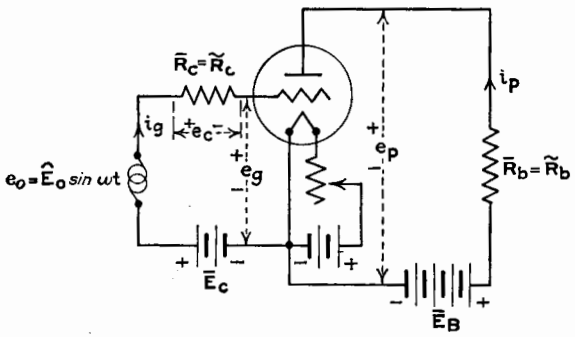


FIG. 106.—A triode having a pure resistance in grid and plate circuits.

Figure 106 gives the circuit constants and connections for the case considered. Pure resistances, whose values are independent of frequency, form the grid- and plate-circuit external impedances. Figures 107 and 108 are the $i_p - e_p$ and $i_g - e_g$ diagrams of the tube. First, the resistance line ad is laid off in Fig. 107 from the point d on the voltage axis and at a distance from O equal to the steady-plate-battery voltage $\bar{E}_B$. The quiescent point cannot yet be determined since it depends upon the quiescent grid potential $\bar{E}_g$ which is as yet unknown. The operating point must travel over some portion of the line ad, also unknown. Each point of its path corresponds to a certain

value of e_p and a certain value of e_g. For each point of line *ad* of Fig. 107, there is a corresponding point on the $i_g - e_g$ diagram of Fig. 108, determined by the corresponding values of e_p and e_g for each point. The path line *ad* of Fig. 107 has a corresponding path line on the $i_g - e_g$ diagram, shown as *uv* in Fig. 108. Line *uv* is the projection upon the $i_g - e_g$ plane of a path on the grid-current characteristic surface and corresponds to the path on the plate-current surface determined by the impedance Z_b.

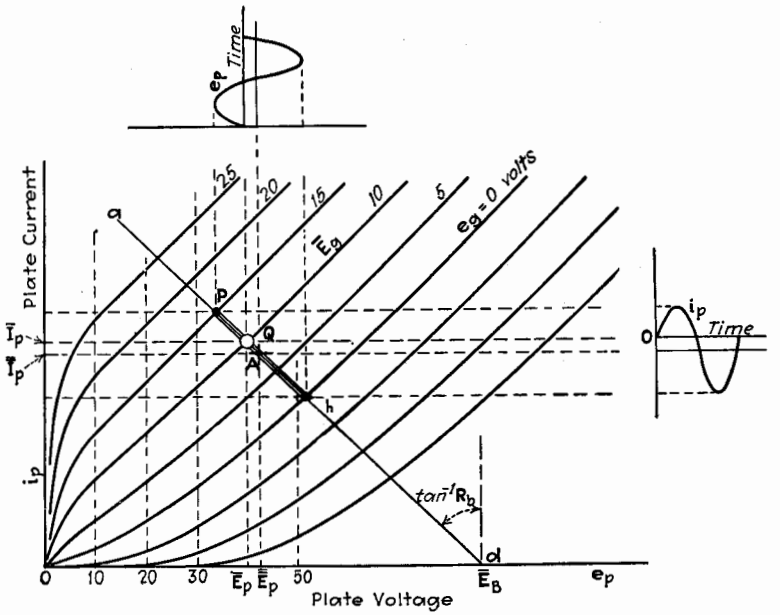


FIG. 107.—Path of operation on the plate-current surface when $\bar{R}_b = R_b$ and $\bar{R}_c = R_c$. Grid current not zero.

The next step is to determine the grid-current and grid-voltage time plots from Fig. 108. At a distance $\bar{E}_c$ from the i_g-axis a circle is drawn, the radius of which is equal to $\hat{E}_0$. A reference vector $\hat{E}_0$, rotating with angular velocity ω, gives, by projection upon the voltage axis, the instantaneous values of e_0. The time plot of e_0 is shown in Fig. 108. The distance from the i_g-axis to this plot gives the value of $\bar{E}_c + \tilde{e}_0$. But, by referring to Fig. 106, it is seen that

$$\bar{E}_c + \tilde{e}_0 = e_g + e_c \tag{179}$$

The value of e_g at any instant t can be found from the intersection of the path line uv with a line lw laid off from the point on the voltage axis representing $\bar{E}_c + \tilde{e}_0$, and making an angle with the vertical whose tangent is equal to $\bar{R}_c = \tilde{R}_c$. By per-

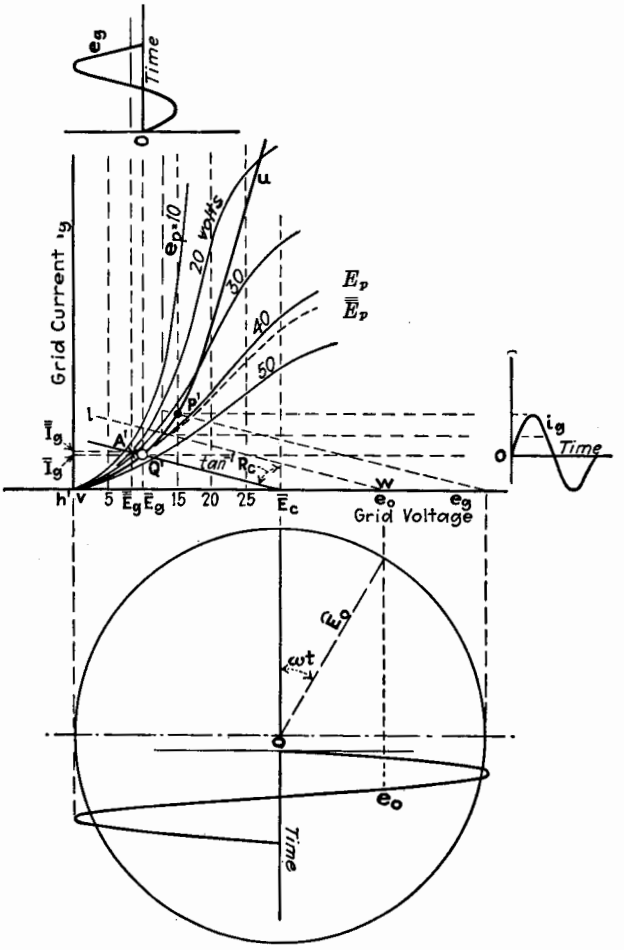


FIG. 108.—Path of operation on the grid-current surface when $\tilde{R}_b = \bar{R}_b$ and $\tilde{R}_c = \bar{R}_c$.

forming this construction for various times during a cycle, the wave forms of e_g and i_g can be determined. The quiescent point, indicated by a circle, lies on the path line uv, but the average point, shown by a cross, does *not* lie on the path, because

of the curvature of the path line. Furthermore, since the average point is not on the grid-current characteristic surface, the $i_g - e_g$ curve passing through the average point does not correspond to the average plate voltage. The actual path is $p'h'$.

The wave form of e_g plotted against time has been determined. To find the plate current and plate voltage at any time t, it is necessary merely to project lines horizontally and vertically from the intersection of line *ad*, (Fig. 107), with the particular $i_p - e_p$ curve which corresponds to e_g at time t. The repetition of the process of projection yields the wave forms of i_p and e_p, as shown in Fig. 107. The quiescent point in Fig. 107 can now be found from the quiescent point in Fig. 108, but the average point, shown by a cross, can be found only after determining the average value of i_p or e_p. The average point lies on the line *ad* but does *not* lie on the $i_p - e_p$ curve corresponding to the average grid voltage $\overline{E}_g$.

III. POWER RELATIONS IN TRIODE CIRCUITS

In studying the power relations in triode circuits, *i.e.*, the power sources and the power absorbed by the several parts of the circuits, two cases must be recognized. The *first case* includes those types of operation of the triode in which there is *no* transfer of power in either direction between the grid and plate circuits. The *second case* includes the more complex types of operation when power may be transferred either from the grid circuit to the plate circuit, or from the plate circuit to the grid circuit, or in both directions.

94. Conditions of Power Interchange between the Grid and Plate Circuits.—Preliminary to the study of power relations, examine the conditions under which power may be transferred between the two circuits of a triode. It is obvious that no power can pass from or to the grid circuit when there is no grid current. There is no grid current when the grid is polarized negatively and when the grid-potential variations are so small as never to reach positive values. Further, there is no transfer of power when the frequency is so low that no appreciable current passes through the capacitance between grid and plate. There is no sharp line of negligibility of capacitance current, because some capacitance current exists at all frequencies. A very small capacitance current may cause power exchange from or to the

grid circuit, and, although this current may be very small, the power exchange may be appreciable compared to the total power available. Ordinarily, at audio frequencies the power interchange through the grid-to-plate capacitance of a triode is negligibly small.

If, however, there is a flow of grid current, it does not necessarily follow that power is transferred between the grid and plate

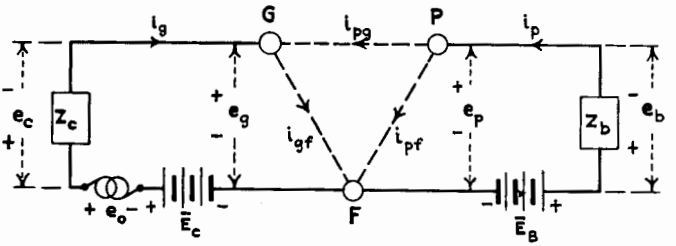


FIG. 109.—Schematic diagram showing currents between the electrodes of a triode.

circuits. If, for example, electrons pass from cathode to grid as well as from cathode to plate, there is no *actual* e.m.f. introduced into either circuit by the other circuit, and the currents in the two circuits are independent. Hence, no power can be transferred from one circuit to the other. More complicated conditions may arise, however, due to ionization, secondary emission, or capacitance between electrodes, by reason of which current may pass directly between grid and plate or *vice versa* as conventionally illustrated by the current i_{pg} in Fig. 109. In this case, as we shall see by examining analytically the power relations, power may be transferred from one circuit to the other.

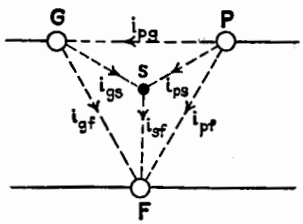


FIG. 110.—Diagram illustrating currents in a triode when ionization occurs.

It may be objected rightfully that Fig. 109 does not represent every condition that may arise in a triode. Figure 109 does represent conditions when current i_{pg} is the current through the capacitance between electrodes or when secondary emission from the grid or plate gives rise to a current to the other electrode. Secondary emission may take place, however, from some other point, such as S in Fig. 110, which is not at the potential of either the grid or the plate. Again, ionization may take place in the space between electrodes, by reason of which currents

may exist, as illustrated in Fig. 110 by the currents i_{gs}, i_{ps}, and i_{sf}, where S is the point of ionization. When ionization plays an important part in the conduction, there are many points indicated by S.

We shall now analyze the simpler and more usual case illustrated in Fig. 109. The extension of the treatment to the more complicated case of Fig. 110 is left for the reader.

The sum of the voltages in the plate circuit, multiplied by the current i_p and averaged over a time T, which is the longest period of the electrical variations, gives the power equation

$$\frac{1}{T}\int_0^T \bar{E}_B i_p dt - \frac{1}{T}\int_0^T e_b i_p dt - \frac{1}{T}\int_0^T e_p i_p dt = 0 \qquad (180)$$

Since $i_p = i_{pf} + i_{pg}$, Eq. (180) becomes

$$P_B - P_b - \frac{1}{T}\int_0^T e_p i_{pf} dt - \frac{1}{T}\int_0^T e_p i_{pg} dt = 0 \qquad (181)$$

The first integral in Eq. (180) is the power delivered by the plate battery and is denoted by P_B. The second integral in Eq. (180) is the total power delivered to the load Z_b and is denoted by P_b. The third term in Eq. (181) represents power lost in heat within the tube and is denoted by H_{pf}. The last term in Eq. (181) will be dealt with presently.

The power equation for the grid circuit is obtained in a similar manner and is

$$\frac{1}{T}\int_0^T \bar{E}_C i_g dt + \frac{1}{T}\int_0^T e_0 i_g dt - \frac{1}{T}\int_0^T e_c i_g dt - \frac{1}{T}\int_0^T e_g i_g dt = 0 \qquad (182)$$

Since $i_g = i_{gf} - i_{pg}$, Eq. (182) becomes

$$P_C + P_0 - P_c - \frac{1}{T}\int_0^T e_g i_{gf} dt + \frac{1}{T}\int_0^T e_g i_{pg} dt = 0 \qquad (183)$$

The terms in Eq. (183) are power delivered by the C-battery, power delivered by the e.m.f. e_0, power delivered to the load Z_c, heat lost within the tube and denoted by H_{gf}, and a term to be considered.

The sum of Eqs. (181) and (183) is

$$P_B + P_C + P_0 = P_b + P_c + H_{pf} + H_{gf} + \frac{1}{T}\int_0^T (e_p - e_g) i_{pg} dt \qquad (184)$$

The left-hand side of Eq. (184) is the total power input to the system, and the right-hand side is the equal amount of power,

part of which is delivered to the external impedances and part lost in heat inside the tube. Evidently the last integral represents heat loss in the tube due to the current i_{pg} and is denoted by H_{pg}. Hence,

$$H_{pg} = \frac{1}{T}\int_0^T e_p i_{pg} dt - \frac{1}{T}\int_0^T e_g i_{pg} dt \tag{185}$$

In general, the term H_{pg} is a positive quantity, although we must not ignore the possibility that electrons may go between grid and plate or *vice versa* against the electric field between the electrodes. This phenomenon may take place because of initial velocity of emission of the secondary electrons, or because of momentum imparted to a carrier due to impact of another carrier. In such cases H_{pg} may be negative.

If $\frac{1}{T}\int_0^T e_p i_{pg} dt$ is *negative*, power flows into the plate circuit from the grid circuit, or else, as a remote possibility, the negativity of the integral is due partly to energy imparted to the carriers in current i_{pg} by the carriers in i_{pf}. Then H_{pf} is increased by the amount of energy transferred directly from the current stream i_{pf} to current i_{pg}.

Similarly, if $\frac{1}{T}\int_0^T e_g i_{pg} dt$ is *positive*, power flows into the grid circuit from the plate circuit, or else a similar phenomenon, as described above for the plate circuit, takes place so that a part of the integral $\frac{1}{T}\int_0^T e_g i_{pg} dt$ comes from H_{gf}. We shall study this case of power interchange between the two circuits in more detail following the analytical study of power relations for the simpler cases given in the next section.

95. Power Relations with No Power Interchange between the Grid and Plate Circuits.—In deducing the general power relations in the plate circuit of a triode, the circuit arrangement is as indicated in Fig. 111. The plate load Z_b is any form of impedance. The plate-circuit ammeter is a d-c. instrument which reads the average plate current $\bar{\bar{I}}_p$. A d-c. voltmeter is to be imagined connected from plate to filament to give the average plate voltage $\bar{\bar{E}}_p$. An instrument is not shown connected across the tube, because of the disturbing effect of a current path in that position. The plate current i_p, which may be of any periodic form, can be resolved into an average component $\bar{\bar{I}}_p$ and a component $\bar{i}_p$

which has zero average value and is measured from the average value $\bar{\bar{I}}_p$. Thus

$$i_p = \bar{\bar{I}}_p + \bar{i}_p \tag{186}$$

Similarly, the plate voltage e_p may be resolved, giving

$$e_p = \bar{\bar{E}}_p + \bar{e}_p \tag{187}$$

and the load voltage may be resolved, giving

$$e_b = \bar{\bar{E}}_b + \bar{e}_b \tag{188}$$

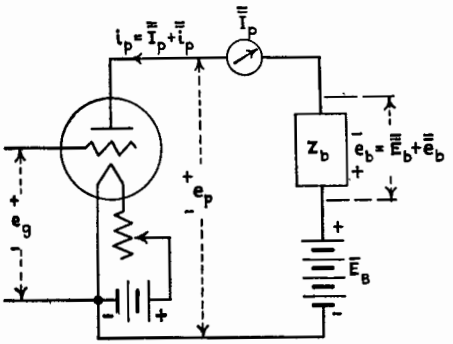


FIG. 111.—A triode and its plate load.

The power supplied by the plate battery is

$$\left.\begin{aligned} P_B &= \frac{1}{T}\int_0^T \bar{E}_B i_p dt \\ &= \frac{1}{T}\int_0^T \bar{E}_B \bar{\bar{I}}_p dt + \frac{1}{T}\int_0^T \bar{E}_B \bar{i}_p dt \\ &= \bar{E}_B \bar{\bar{I}}_p \end{aligned}\right\} \tag{189}$$

where T is the time of one complete cycle. Since the plate battery is the only source of power in the plate circuit, Eq. (189) gives the input power. The battery power is correctly given by the battery voltage times the average current.

The power delivered to the tube is

$$\left.\begin{aligned} P_p &= \frac{1}{T}\int_0^T e_p i_p dt \\ &= \frac{1}{T}\int_0^T (\bar{\bar{E}}_p + \bar{e}_p)(\bar{\bar{I}}_p + \bar{i}_p) dt \\ &= \bar{\bar{E}}_p \bar{\bar{I}}_p + \frac{1}{T}\int^T \bar{e}_p \bar{i}_p dt \end{aligned}\right\} \tag{190}$$

Equation (190) shows that the power delivered to the tube is expressible in two terms, one the d-c. power calculated from average values or d-c. instrument readings, and the other the a-c. power, the a-c. values being measured from the average value as axis.

The power delivered to the load or impedance Z_b is

$$\left.\begin{aligned} P_b &= \frac{1}{T}\int_0^T e_b i_p dt \\ &= \frac{1}{T}\int^T (\bar{\bar{E}}_B + \bar{\bar{e}}_b)(\bar{\bar{I}}_p + \bar{\bar{i}}_p)dt \\ &= \bar{\bar{E}}_b\bar{\bar{I}}_p + \frac{1}{T}\int_0^T \bar{\bar{e}}_b\bar{\bar{i}}_p dt \end{aligned}\right\} \tag{191}$$

The load power consists of a d-c. component and an a-c. component, just as with the tube power.

The input power must equal the sum of the tube and load powers. Hence,

$$\bar{E}_B\bar{I}_p = (\bar{\bar{E}}_p + \bar{\bar{E}}_b)\bar{\bar{I}}_p + \frac{1}{T}\int_0^T (\bar{e}_p + \bar{\bar{e}}_b)\bar{\bar{i}}_p dt \tag{192}$$

Adding the voltages in the plate circuit,

$$\bar{E}_B - \bar{\bar{E}}_b - \bar{\bar{E}}_p = \bar{\bar{e}}_b + \bar{\bar{e}}_p \tag{193}$$

Equation (193) shows that $\bar{\bar{e}}_b + \bar{\bar{e}}_p$ is constant; but since both $\bar{\bar{e}}_b$ and $\bar{\bar{e}}_p$ have zero average values, the two parts of Eq. (193) are equal to zero. Thus,

$$\bar{E}_B - \bar{\bar{E}}_b - \bar{\bar{E}}_p = 0 \tag{194}$$

$$\bar{\bar{e}}_b = -\bar{\bar{e}}_p \tag{195}$$

Examining Eq. (192) in the light of Eqs. (194) and (195), it is seen that the power equation can be divided into two equations, one for d-c. power, denoted by dashes over the letter P, and the other for a-c. power, denoted by $\tilde{P}$. Hence

$$\bar{P}_B = \bar{P}_p + \bar{P}_b \tag{196}$$

$$\tilde{P}_p = -\tilde{P}_b \tag{197}$$

where $\tilde{P}_b$ is in general a positive quantity representing dissipation of energy in the load. The total power P_b in the load is

$$P_b = \bar{P}_b + \tilde{P}_b \tag{198}$$

and the total power P_p in the tube is

$$P_p = \bar{P}_p + \tilde{P}_p = \bar{P}_p - \tilde{P}_b \tag{199}$$

The results of the analysis may be stated as follows:

1. When there is no secondary emission or ionization, the plate battery is the only source of power in the plate circuit. The power supplied by the plate battery is given by the average current output multiplied by the battery voltage.

2. The power supplied to the load consists of two parts: the d-c. power, calculated from the average current through the load and the average voltage across the load, and the a-c. power.

3. The power dissipated as heat at the plate of the tube is equal to the d-c. power, calculated from average values of plate current and plate e.m.f., minus the a-c. power supplied to the load.

If the special case of sinusoidal variations of plate current and plate potential is considered, the results can be somewhat simplified. Average values are now the same as the quiescent values. Figure 112, which is the $i_p - e_p$ diagram for sinusoidal variations with a pure resistance as plate load, is similar to Fig. 101. The power supplied by the battery is $\bar{E}_B \bar{I}_p$, which is given by the area of rectangle *bsdO*. The d-c. power $\bar{P}_b$ supplied to the load is given by the area of rectangle *Qsdj*, and the d-c. power $\bar{P}_p$ supplied to the tube is given by the area of rectangle *bQjO*. The a-c. power $\tilde{P}_b$ supplied to the load is given by

$$\tilde{P}_b = \frac{(\Delta \widehat{I}_p)^2 \tilde{R}_b}{2} = \frac{\overline{Qf} \times \overline{pf}}{2} \tag{200}$$

Therefore $\tilde{P}_b$, which is the useful a-c. power output, is given by the area of triangle *pfQ*. Since $\tilde{P}_b = -\tilde{P}_p$, the power lost as heat at the plate is given by the area $(bQjO - pfQ)$, and the total power in the load by area $(Qsdj + pfQ)$.

Even if the $i_p - e_p$ curves were parallel straight lines over the entire $i_p - e_p$ plane, and if the amplitude of variation of plate current were as great as possible while maintaining sinusoidal variations, the area of triangle *pfQ* could never be greater than one-half the area of rectangle *bQjO*. Even if $\bar{R}_b$ were zero, the theoretical maximum limit of efficiency is 50 per cent for sinusoidal operation. Because of the curvature of the $i_p - e_p$ lines, the practical maximum efficiency for sinusoidal operation is much less than 50 per cent.

If the plate impedance has reactance, and the variations of all electrical quantities are sinusoidal, the a-c. power terms can be represented by certain areas. Refer to Fig. 103 for the discussion of this case. The d-c. components of power are exactly as described with reference to Fig. 112. The a-c. power $\tilde{P}_b$ is given by Eq. (200), so that $\tilde{P}_b$ is given by the area of triangle pfQ, as in Fig. 112.

The quadrature power is equal to $\widehat{\Delta E}_p \times \widehat{\Delta I}_p/2 \times X/Z$ or $(\widehat{\Delta I}_p)^2/2 \times X$. It has been shown that the intercept Qm of Fig. 103 is equal to $\widehat{\Delta I}_p \times X$. Since the line lm is equal to $\widehat{\Delta I}_p$, the area

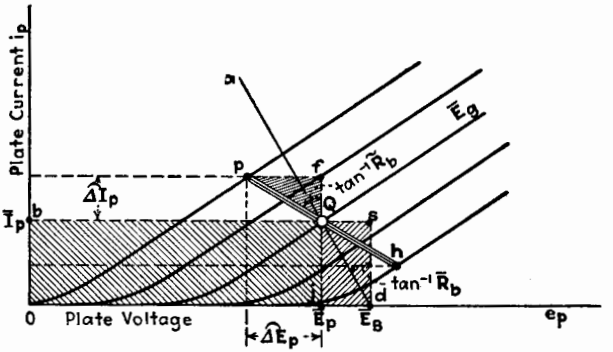


FIG. 112.—Areas representing power in the plate circuit of the triode.

of either triangle lmQ or triangle lfQ represents the quadrature power.

96. Non-supply of Power by Fictitious Voltage.—The equivalent plate-circuit theorem has shown that for sinusoidal operation of a triode, when the path on the plate-current characteristic surface is so small that u_p and k_p are essentially constant, the a-c. component of plate current can be calculated by assuming a *fictitious* e.m.f. The value of this e.m.f. is given by u_p times the a-c. component of grid voltage. This fictitious e.m.f. acts through the external plate-circuit impedance in series with a constant resistance equal to r_p of the tube. This theorem is usually limited to very small values for the a-c. components of current and potential. It has been emphasized that the voltage $u_p \Delta E_g$ is a *fictitious* voltage, and hence *it cannot supply power.* To illustrate the fallacious reasoning in which one is liable to indulge, assume a triode with a pure resistance, independent of frequency, as plate impedance. Then construct as usual the

path ph on the $i_p - e_p$ diagram in Fig. 113. According to the e-p-c. theorem,

$$\widehat{\Delta I}_p = \frac{u_p \widehat{\Delta E}_g}{r_p + \tilde{R}_b} \tag{201}$$

If this equation is cleared of fractions and multiplied throughout by $\widehat{\Delta I}_p/2$, we have the apparent power equation

$$\frac{(\widehat{\Delta I}_p)^2}{2} \cdot r_p + \frac{(\widehat{\Delta I}_p)^2}{2} \cdot \tilde{R}_b = u_p \frac{\widehat{\Delta E}_g \widehat{\Delta I}_p}{2} \tag{202}$$

Equation (202) must be a true equation, but two of the terms *do not represent actual power*. This will perhaps be clearer if we proceed as follows: r_p is equal to $(\partial e_p/\partial i_p)_{e_g = \text{constant}}$ and this

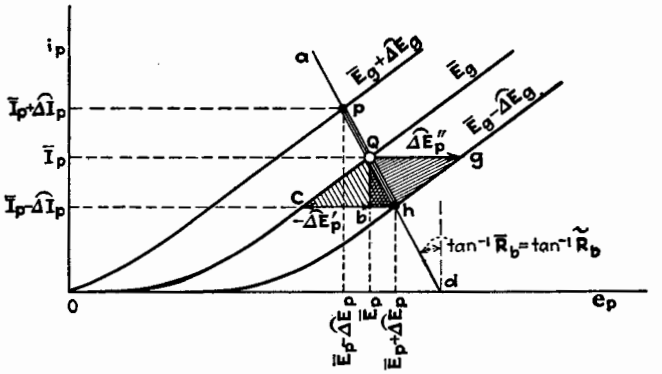


FIG. 113.—The $i_p - e_p$ diagram of a triode illustrating certain fictitious power terms.

becomes $(\widehat{\Delta E}_p'/\widehat{\Delta I}_p)_{e_g = \bar{E}_g}$ for the particular case illustrated in Fig. 113. Therefore, the first term of Eq. (202) becomes $\widehat{\Delta I}_p \times \widehat{\Delta E}_p'/2$ which is given by the area of triangle Qbc. Since $\widehat{\Delta E}_p'$ is merely a construction line on the diagram and is not a voltage in the actual operation, this area cannot represent real power. Therefore, $(\widehat{\Delta I}_p)^2 r_p$ does *not* represent power lost in the tube. Similarly, $u_p = -(\partial e_p/\partial e_g)_{i_p = \text{constant}}$ and this is numerically equal to $(\widehat{\Delta E}_p''/\widehat{\Delta E}_g)_{i_p = \bar{I}_p}$ for the case of Fig. 113. The right-hand side of Eq. (202) is consequently equal to $\widehat{\Delta I}_p \times \widehat{\Delta E}_p''/2$, which is given by the area of triangle Qhg. This, too, does *not* represent actual power, because $\Delta E_p''$ does not actually exist in the operation over path ph. Finally, $\tilde{R}_b = de_b/di_p$, which is numerically equal to $\widehat{\Delta E}_p/\widehat{\Delta I}_p$ for the case of Fig. 113.

The middle term of Eq. (202) becomes $\widehat{\Delta E_p} \times \widehat{\Delta I_p}/2$, which is represented by the area of triangle Qbh. Since both $\widehat{\Delta E_p}$ and ΔI_p actually exist, the middle term of Eq. (202) and the area of triangle Qbh represent the actual a-c. power dissipated in $\bar{R}_b$. Examination of Fig. 113 shows that Eq. (202) is satisfied, because area Qbc + area Qbh is equal to area Qhg.

Similarly, the fictitious potential $u_g \Delta E_p$ does not represent an actual e.m.f. Consequently power is not supplied to the grid circuit from the plate circuit because of this voltage. It will be shown in a later chapter that the effect of the plate circuit on the grid circuit is to alter the equivalent impedance of the grid circuit and hence the power loss in it.

Attention is called to a very important limitation to the validity of the power-distribution theorems just given. In each case it was assumed there was no source of power in the plate circuit other than the B-battery or generator, and in the grid circuit that the only sources of power were the C-battery or generator and the applied e.m.f. e_0. Consequently it follows that no power can be transferred from the grid circuit to the plate circuit or *vice versa*. This assumption is true *only* when the carriers of electric charge, electrons and positive ions, pass directly from the filament to the plate, from the filament to the grid, or to both. In other words, *if there is any ionization, or if any current passes between grid and plate or vice versa because of reflection of electrons or of secondary emission, the assumption is not valid and the theorems given for the distribution of power are not valid.*

***97. Power Relations When Power Interchange Exists between the Grid and Plate Circuits.**—Consider further the analysis begun in Sec. 94 of the power relations in the grid and plate circuits of a triode, when power may be interchanged between the two circuits.

Express the currents and potentials of Fig. 109 in terms of average values and of periodic values measured from the average values, as in Eqs. (186) and (187). Equation (180) can be broken up by the method of Sec. 95 into the two following equations, the first for the steady power and the second for the periodic power.

$$\left.\begin{aligned} \bar{\bar{E}}_B \bar{\bar{I}}_p - \bar{\bar{E}}_b \bar{\bar{I}}_p - \bar{\bar{E}}_p \bar{\bar{I}}_p &= 0 \\ \bar{P}_B - \bar{P}_b - \bar{P}_p &= 0 \end{aligned}\right\} \quad \text{or} \quad (203)$$

and

$$\left.\begin{aligned} \frac{1}{T}\int^{T} \bar{e}_b \bar{i}_p dt &= -\frac{1}{T}\int_0^T \bar{e}_p \bar{i}_p dt \\ \text{or} \qquad \bar{P}_b &= -\bar{P}_p \end{aligned}\right\} \tag{204}$$

Equation (204) corresponds exactly to Eq. (197). Since $\bar{P}_b$ is generally a positive quantity, $\bar{P}_p$ must generally be negative.

A similar treatment of the grid-circuit equation, Eq. (182), gives

$$\bar{P}_C + \bar{P}_0 - \bar{P}_c - \bar{P}_g = 0 \tag{205}$$

and

$$\bar{P}_0 - \bar{P}_c - \bar{P}_g = 0 \tag{206}$$

In Eq. (206), $\bar{P}_0$ and $\bar{P}_c$ are nominally positive quantities. $\bar{P}_g$ may, therefore, be either positive or negative.

We may now compare this case with the simple case explained in Sec. 95. In the simple case, $\bar{P}_p$, when $\bar{e}_p$ is zero, represents power dissipated *at the plate*, whereas, in the case in hand, the total tube loss when $\bar{e}_p$ and $\bar{e}_g$ are zero is $\bar{P}_p + \bar{P}_g$. It is impossible to tell without more information what fraction is lost at each electrode. If $\bar{\bar{I}}_{pf}$ and $\bar{\bar{I}}_{pg}$ are pure electronic currents passing to the plate, the power lost as heat at the plate is

$$\bar{H}_p = \bar{\bar{E}}_p \bar{\bar{I}}_{pf} + (\bar{\bar{E}}_p - \bar{\bar{E}}_g)\bar{\bar{I}}_{pg} = \bar{P}_p - \bar{\bar{E}}_g \bar{\bar{I}}_{pg} \tag{207}$$

and if $\bar{\bar{I}}_{gf}$ is a pure electronic current, the power lost as heat at the grid is

$$\bar{H}_g = \bar{\bar{E}}_g \bar{\bar{I}}_{gf} = \bar{P}_g + \bar{\bar{E}}_g \bar{\bar{I}}_{pg} \tag{208}$$

The currents and potentials are indicated as average values though if $\bar{e}_p$ and $\bar{e}_g$ are zero, they are the quiescent values.

If, however, $\bar{\bar{I}}_{pg}$ is an electronic current passing to the grid and is due to secondary emission at the plate, the power lost as heat at the plate is

$$\bar{H}_p = \bar{\bar{E}}_p \bar{\bar{I}}_{pf} = \bar{P}_p - \bar{\bar{E}}_p \bar{\bar{I}}_{pg} \tag{209}$$

and the power lost as heat at the grid is

$$\bar{H}_g = \bar{\bar{E}}_g \bar{\bar{I}}_{gf} + (\bar{\bar{E}}_p - \bar{\bar{E}}_g)\bar{\bar{I}}_{pg} = \bar{P}_g + \bar{\bar{E}}_p \bar{\bar{I}}_{pg} \tag{210}$$

In Eqs. (209) and (210), $\bar{\bar{I}}_{pg}$ is intrinsically negative.

Considering now the periodic power, the power term $\tilde{P}_p$ in Eq. (204) is made up of two terms, shown in the following equation:

$$\tilde{P}_p = \frac{1}{T}\int_0^T \bar{\bar{e}}_p \bar{\bar{i}}_{pf} dt + \frac{1}{T}\int_0^T \bar{\bar{e}}_p \bar{\bar{i}}_{pg} dt \tag{211}$$

The power term $\tilde{P}_g$ can be expanded as follows:

$$\tilde{P}_g = \frac{1}{T}\int^{T} \bar{e}_g \bar{i}_{gf} dt - \frac{1}{T}\int^{T} \bar{e}_g \bar{i}_{pg} dt \qquad (212)$$

The first integral of Eq. (211) usually represents a definite decrease in the heat lost at the plate, which, in the simple case of Sec. 95, accounts for the total a-c. power in the load. In the present case, we have an additional term, the second integral of Eq. (211), which may be either positive or negative. The integrals of Eq. (212) may be positive or negative. A clearer picture may be had by assuming that the variations in current and potential are sinusoidal and small enough so that the tube coefficients are constant. Let

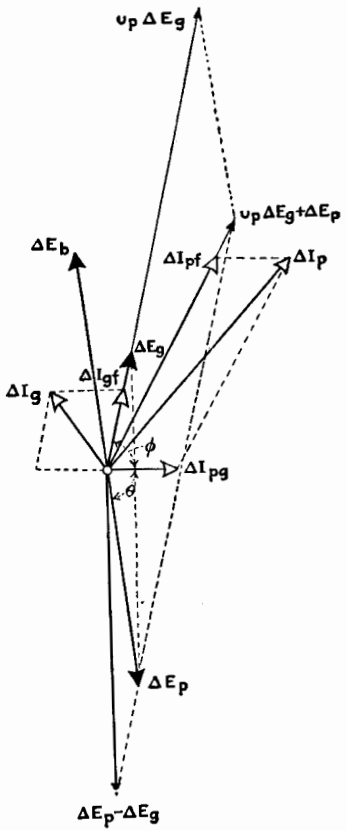


FIG. 114.—Vector diagram of an oscillator with capacitance feedback.

$$\left.\begin{aligned} \bar{i}_{pg} &= \sqrt{2}\Delta I_{pg} \sin \omega t \\ \bar{e}_p &= \sqrt{2}\Delta E_p \sin (\omega t + \theta) \\ \bar{e}_g &= \sqrt{2}\Delta E_g \sin (\omega t + \phi) \\ \bar{i}_{pf} &= \sqrt{2}\Delta I_{pf} \sin (\omega t + \alpha) \\ \bar{i}_{gf} &= \sqrt{2}\Delta I_{gf} \sin (\omega t + \beta) \end{aligned}\right\} (213)$$

Equation (211) becomes

$$\tilde{P}_b = -\tilde{P}_p = -\Delta E_p \Delta I_{pf} \cos (\theta - \alpha) - \Delta E_p \Delta I_{pg} \cos \theta \qquad (214)$$

and Eq. (212) becomes

$$\tilde{P}_0 - \tilde{P}_c = \Delta E_g \Delta I_{gf} \cos (\phi - \beta) - \Delta E_g \Delta I_{pg} \cos \phi \qquad (215)$$

We shall now examine two illustrations of the application of the above discussion. As the first illustration, let Fig. 114 be the vector diagram for a triode having an oscillatory circuit in both grid and plate circuits. $\tilde{P}_0$ is assumed to be zero, and the current i_{pg} is solely through the interelectrode capacitance between grid and plate. In Fig. 114, ΔI_{pg} represents this capacitance current which leads the potential difference $\Delta E_p - \Delta E_g$ by 90 electrical degrees.

The grid and plate voltages ΔE_g and ΔE_p are represented in the diagram. Evidently H_{pg} of Eq. (185) is zero. Referring to Eq. (215), $\tilde{P}_c$ is a positive quantity, since it is power dissipated in the oscillatory circuit. The first term on the right-hand side of Eq. (215) is intrinsically positive, as it represents heat lost at the grid. The last term in Eq. (215) represents power transferred from the plate circuit to the grid circuit to supply the losses just referred to, *i.e.*, in the grid circuit and at the grid.

In Eq. (214), $\tilde{P}_b$ is positive, as it represents power dissipated in the oscillatory circuit connected to the plate. The last term in Eq. (214) is numerically equal to the last term in Eq. (215) and represents the power transferred from the plate circuit to the grid circuit. Both $\tilde{P}_b$ and $\Delta E_p \Delta I_{pg} \cos \theta$ arise from the term

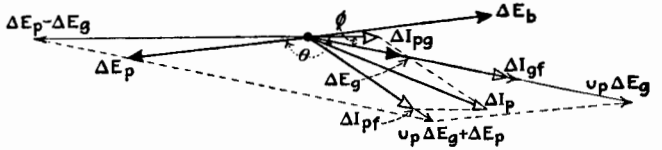


FIG. 115.—Vector diagram showing transfer of energy between plate and grid circuits when secondary emission exists.

$-\Delta E_p \Delta I_{pf} \cos (\theta - \alpha)$, a positive quantity which represents a decrease in the heating of the plate of the triode.

As a second illustration we shall consider the case in which the current i_{pg} is due entirely to secondary emission at the grid. The vector diagram is shown in Fig. 115. The current ΔI_{pg} is shown opposite in phase to $\Delta E_p - \Delta E_g$, because the secondary emission increases as ΔE_g increases. Although the steady portion of H_{pg}, Eq. (185), is positive and increases the heating of the plate, the alternating portion of H_{pg}, represented by $\tilde{H}_{pg} = (\Delta E_p - \Delta E_g)\Delta I_{pg} \cos 180°$, is negative and causes a decrease in heating of the plate.

The term $\tilde{P}_b$ in Eq. (214) is a positive quantity. The quantity $\Delta E_p \Delta I_{pf} \cos (\theta - \alpha)$ in Eq. (214) is negative and also represents a decrease in heating of the plate. The term $\Delta E_p \Delta I_{pg} \cos \theta$ in Eq. (214) is negative and represents, in part, power transferred from the plate circuit to the grid circuit.

In Eq. (215), all of the terms are positive. The term $\Delta E_g \Delta I_{gf} \cos (\phi - \beta)$ represents an increase in the heat developed

at the grid, whereas the term $\Delta E_g \Delta I_{pg} \cos \phi$ is positive and represents a gain in power in the grid circuit.

The interesting features of this case are as follows: In the quiescent condition, the plate is heated as a result of the bombardment of the electrons in both current streams $\bar{I}_{pf}$ and $\bar{I}_{pg}$. When the alternating e.m.f. E_0 is impressed, a-c. power appears in Z_c and Z_b, and an increase in heat at the grid occurs, *all* resulting from the decrease in power lost as heat at the plate.

98. Condition for Maximum Power Output for Constant E_g.— When dealing with power systems one is frequently interested in knowing the condition under which a maximum of output power is obtained. In triodes the condition for maximum power output depends greatly upon the restrictions imposed upon the operation of the system. One of the several practical problems which arise under different restrictive conditions of operation will now be considered. Others will be discussed in Chap. XII.

The problem considered here is that of finding the value of the plate-circuit resistance which gives maximum power in the plate load when the alternating grid potential is held constant. Assume the a-c. component of the applied grid voltage, E_g, to be sinusoidal and limited in amplitude so that the coefficients u_p and k_p do not vary appreciably over the path of the operating point on the plate-current characteristic surface. This implies that if the operating point is on a curved portion of the characteristic surface, the amplitudes of the a-c. components of plate current and plate potential must be very small. But if the operating point is situated on a plane portion of the plate-current surface, the amplitudes of the a-c. components may be as large as will retain a linear relation between i_p, e_p, and e_g.

If the plate impedance has a d-c. resistance $\bar{R}_b$ and an a-c. impedance $\tilde{R}_b + jX_b$, the a-c. power in the impedance, which is considered as the power output, is

$$\tilde{P}_b = \text{power output} = I_p^2 \tilde{R}_b = \frac{u_p^2 E_g^2 \tilde{R}_b}{(r_p + \tilde{R}_b)^2 + X_b^2} \tag{216}$$

To find the value of $\tilde{R}_b$ to give a maximum $\tilde{P}_b$, equate to zero the derivative of Eq. (216) with respect to $\tilde{R}_b$, giving

$$(r_p + \tilde{R}_b)^2 + X_b^2 - 2\tilde{R}_b(r_p + \tilde{R}_b) = 0 \tag{217}$$

Equation (217) reduces to

$$\tilde{R}_b = \sqrt{r_p^2 + X_b^2} \tag{218}$$

From Eq. (216), it is clear that the presence of X_b reduces the power output, but if X_b is present, the best value of $\tilde{R}_b$ is given by Eq. (218). If X_b is zero, for maximum power output

$$\tilde{R}_b = r_p \text{ (for max. } \tilde{P}_b \text{ if } X_b = 0) \tag{219}$$

Referring to Fig. 112, Eq. (219) means that for maximum power output the slope of the path ph should equal numerically the slope of the $i_p - e_p$ curve which passes through the quiescent point.

General References

BENHAM: The Internal Action of Thermionic Systems at Moderately High Frequencies, *Phil. Mag.*, **5**, 630 (1928); **11**, 457 (1931).

JACKSON: The Transient Response of the Triode Valve Equivalent Network, *Phil. Mag.*, **13**, 143 (1932).

CHAPTER IX

DYNAMIC MEASUREMENT OF TRIODE COEFFICIENTS

Attention has been called to the importance of the triode coefficients in expressing the performance of the triode, and in the development of its theory. The determination of the coefficients is usually made by direct methods known as *dynamic methods*. These methods make use of networks in one branch of which is a telephone receiver or its equivalent. An alternating potential, usually of a frequency of 1,000 cycles per second, is impressed across two points of the network, and the constants of the circuits are adjusted until there is no sound in the telephone receiver. Obviously, the coefficients are determined for the impressed frequency of 1,000 cycles per second. Since, however, in high-vacuum tubes, there are no appreciable lag effects except at very high frequencies, the coefficients are all real quantities, *i.e.*, not complex. Therefore they are practically independent of frequency, excluding frequencies greater than about 10^7 cycles per second. The coefficients of hard (high-vacuum) tubes at very high frequencies and of soft (containing gas) tubes at all frequencies are complex quantities.

99. General Considerations in the Measurement of Triode Coefficients.—Before describing the specific circuits suitable for the measurement of the coefficients of a triode, a few general considerations applicable to all methods may be pointed out. The impressed alternating e.m.f., though small, is finite and causes the potentials and currents of the tube to vary by appreciable amounts or, in other words, causes the operating point on the characteristic surface to move over a path of appreciable length. If the path over which the operating point moves is practically straight and if the alternating grid and plate potentials are not greater than a few tenths of a volt, no great error is introduced in the coefficient being determined. If the path has curvature, the coefficient being determined is more nearly the true coefficient at a given point, the smaller the path traversed. This means that the applied alternating potential must be as small as possible

without unduly reducing the sensitivity of the apparatus. The applied alternating e.m.f. should be such that further reduction in its magnitude makes no appreciable change in the observed value of the coefficient being determined. Usually a one-stage or a two-stage amplifier is necessary to obtain sufficient sensitivity and accuracy, but in the following diagrams the detecting device, for simplicity, is indicated as a telephone receiver.

The plate and grid circuits, when carrying steady currents, should have, where possible, low-resistance paths for these steady components. Otherwise, the actual steady plate or grid voltages are different from the battery voltages, and correction must be made by an amount equal to the external resistance drop. When the amplifier or telephone receivers are in a circuit carrying a direct current, this steady current can be shunted to advantage around the high-resistance telephone or amplifier transformer by a low-resistance choke coil having a high reactance to the alternating current. The primary winding of a filament transformer or of a bell-ringing transformer is suitable for this purpose.

Another important consideration is the effect of stray capacitances, particularly the capacitances between the electrodes of the tube, between the parts of the socket and wires, and from the batteries to ground. These capacitances must be considered as part of the bridge network and, as far as possible, must be balanced out in order to obtain the most accurate results. This can be done by the addition of other variable capacitances, or by a small mutual inductance, properly connected, as will appear when considering the diagrams of connections. The circuits should not be grounded except when grounding causes no error in the balance. The Wagner ground connection should be used where possible.

The a-c. impedance of the telephone receivers or the input to the amplifier system should in each case be adjusted, by transformer if necessary, to give greatest sensitivity for the particular circuit. Highest sensitivity is obtained when the input impedance of the detecting device is approximately equal to the impedance of the circuit to which it is connected.

Usually, the grid and plate voltages are made continuously variable by means of voltage dividers, as shown for the grid voltage in Fig. 116*b*. When this arrangement is used, the alternating component of the current should be shunted around the potential divider by a large condenser of the order of several

microfarads. The other diagrams are simplified by indicating batteries only, but it is to be understood that these potentials may be made variable by the method shown in Fig. 116*b*, or by a variable tap on the battery.

100. Measurement of u and u_p when $i_g = 0$.—Figure 116*a* shows the circuit arrangement for the dynamic method of measuring the voltage ratio u of a triode, where

$$u = -\left(\frac{\partial e_p}{\partial e_g}\right)_{i_p+i_g=\text{const.}}$$

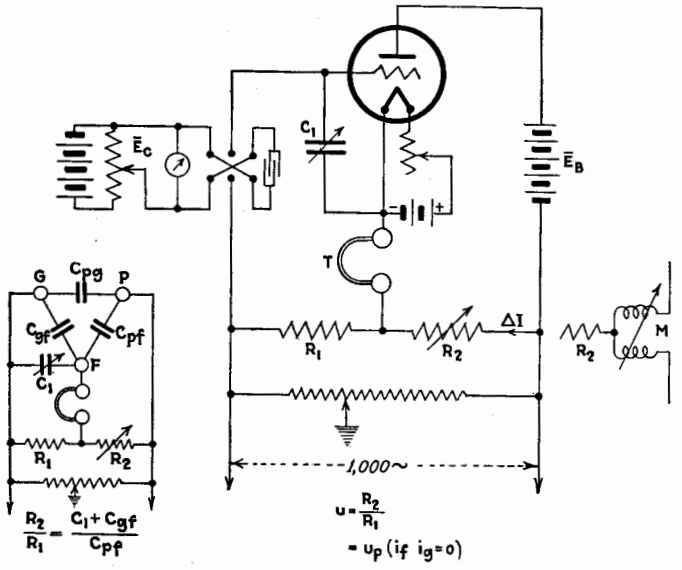


FIG. 116.—First method for the measurement of u, and u_p (if $i_g = 0$).

The resistance R_1 is generally held constant, 10 ohms being a convenient value in most cases. R_2 is variable in tenths, units, and tens of ohms for ordinary triodes.

When R_2 is adjusted to give balance, the alternating grid voltage is $(\Delta I)R_1$. The plate-voltage change in the opposite direction to that of the grid voltage is $(\Delta I)R_2$. If one of these voltages is adjusted so that the alternating current through the telephone receivers is zero, the operating point moves over a portion of a curve of Fig. 53, page 149, determined by the values of the steady polarizing potentials. Then

$$u = \frac{R_2}{R_1} \tag{220}$$

If i_g is zero, Eq. (220) also gives the value of u_p.

A Wagner ground connection is shown in Fig. 116*a* across the source of alternating potential. This variable ground connection should be adjusted so that touching either terminal of the telephone receivers with the wetted finger causes no change in sound.

If the source of alternating potential is not of low resistance, it should be shunted by a low-resistance choke coil in order to provide a low-resistance path for the steady plate current. Then most of the plate current passes through the fixed low resistance R_1 instead of through R_2. The telephone receivers must also be shunted by a low-resistance choke coil.

When the grid voltage is so negative that the plate current is very small, the internal resistance r_p of the tube is of the same order of magnitude as the reactances of the tube capacitances. Under this condition the effects of these capacitances in altering the balance are most marked, and the added refinements to balance out the effects of these capacitances must be used if high accuracy is desired. The diagram of Fig. 116*a* shows the three terminals of the tube, G, P, and F, and the capacitances which exist between these terminals. When R_2 is adjusted to give the approximate balance for u, R_2 is u times R_1. The added capacitance C_1 can then be adjusted to balance the capacitances by making C_1 satisfy the relation

$$\frac{R_2}{R_1} = \frac{C_1 + C_{gf}}{C_{pf}} \tag{221}$$

An alternative method sometimes used for balancing out the quadrature components of potential across the telephone receivers is shown at the right of Fig. 116*b*. A *small* mutual inductance is inserted as shown. The capacitance methods of balance are preferred to this method, because the conditions of balance using mutual inductance are complicated and the value of u obtained using mutual inductance is liable to be in error.

The methods just described for balancing out the capacitance effects require adjustments for every change in the ratio of R_2 to R_1. Figure 117*a* is a diagram of a method[1] of balancing out the capacitance effects which has the advantage that, after the capacitances C_1 and C_2 have once been adjusted, the capacitances are nearly balanced for all values of R_1 and R_2, and little if any readjustment of the condensers is required. A potential divider

[1] Suggested by F. V. Hunt of the Cruft Laboratory, Harvard University.

is connected across the telephone receivers, and the two variable capacitances C_1 and C_2 are connected as shown. The two sections of this potential divider, indicated as R' and R'', together with capacitances C_{pf} and C_2, form a capacitance bridge, the condition of balance for the bridge being

$$\frac{R'}{R''} = \frac{C_2}{C_{pf}} \tag{222}$$

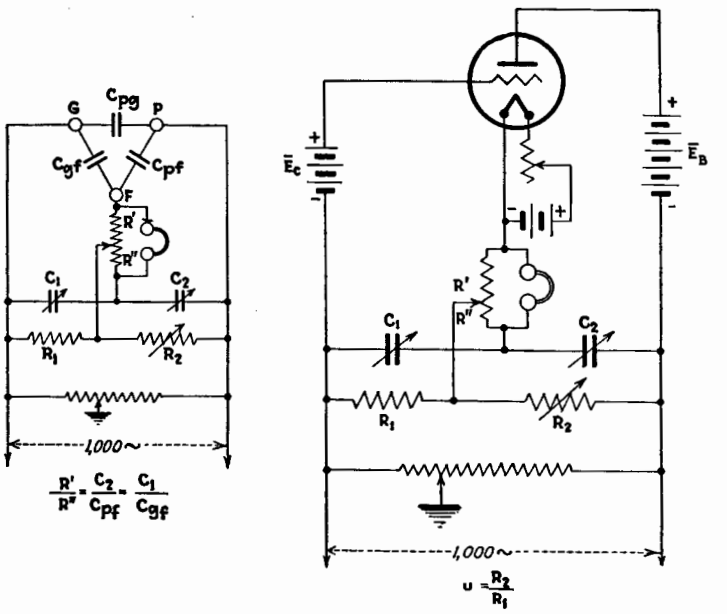


FIG. 117.—Second method for the measurement of u, and u_p (if $i_g = 0$).

This condition is independent of the value of R_2, the voltage across R_2 being the voltage impressed upon this auxiliary bridge. The capacitance current through C_{pf} is thus balanced out from the telephone receivers. Again, the two sections R' and R'' of the potential divider, together with capacitances C_1 and C_{gf}, form another capacitance bridge, the condition of balance being

$$\frac{R'}{R''} = \frac{C_1}{C_{gf}} \tag{223}$$

Thus the capacitance current through C_{gf} is balanced out, although this is of less importance because of the low value of

R_1. The current through C_{pg} has no effect, because C_{pg} is simply a shunt path across the source of potential.

To balance the auxiliary bridges, the cathode should be cold, R_1 should be reduced to zero, and R_2 made as large as possible. C_2 is then adjusted to give balance. Then, with R_2 equal to zero and R_1 large, C_1 is adjusted to balance. A slight readjustment of C_2 may be necessary when the cathode is hot because of a slight change in the tube capacitance due to the presence of the space charge. In many cases sufficient accuracy is obtained by omitting C_1.

It is important that the reactance of C_2 be large in comparison with R_2 in order that the potential across R_2 shall not be appreciably altered by the shunt path through C_2. If R_2 is of the order of 100 ohms, C_2 may be of the order of 1,000 $\mu\mu$f. If C_{pf} is of the order of 10 $\mu\mu$f, R'' is about one one-hundredth of R'. The path for the steady plate current is then through a low-resistance choke coil connected across the telephone receivers and through R''. Therefore, R'' should not be much greater than 10 ohms. A thousand-ohm resistance is suitable, therefore, for the potential divider across the telephone receivers. This somewhat reduces the sensitivity of the detector, but an amplifier gives ample sensitivity. The actual connections are shown in Fig. 117*b*.

Negative values of u from 0 to 1 can be measured by reversing certain connections to R_2, as shown in Fig. 118*b*. For values of u *algebraically less* than -1, certain connections to R_1 are reversed, as shown in Fig. 118*c*. Negative values of u are seldom met with in practice and hence are of little practical interest.

101. Measurement of u_p When $i_g \neq 0$.—The method just given for measuring u is recommended for measuring u_p if the grid current is zero. If there is gaseous ionization or if the grid voltage is positive, the method of Sec. 100 does not give the value of u_p, and the following methods must be used. The methods to be presented are valid for the determination of u_p under all conditions, but the circuits are not so symmetrical and the method of grounding is less satisfactory than in the methods just described. Figure 118*d* gives diagrammatically the arrangement of resistances and balancing capacitances for measuring positive values of u_p. Capacitances C_1 and C_2 balance the triode capacitances C_{pg} and C_{pf}, respectively. The capacitance C_{gf} is not balanced and hence its effect is made small by making

R_1 small. The complete diagram for the measurement of u_p is obtained by adding the batteries to the diagram of Fig. 118*d*.

Negative values of u_p can be measured by reversing certain connections to R_2 or to R_1, as shown in the diagrams of Fig. 118, *e* and *f*.

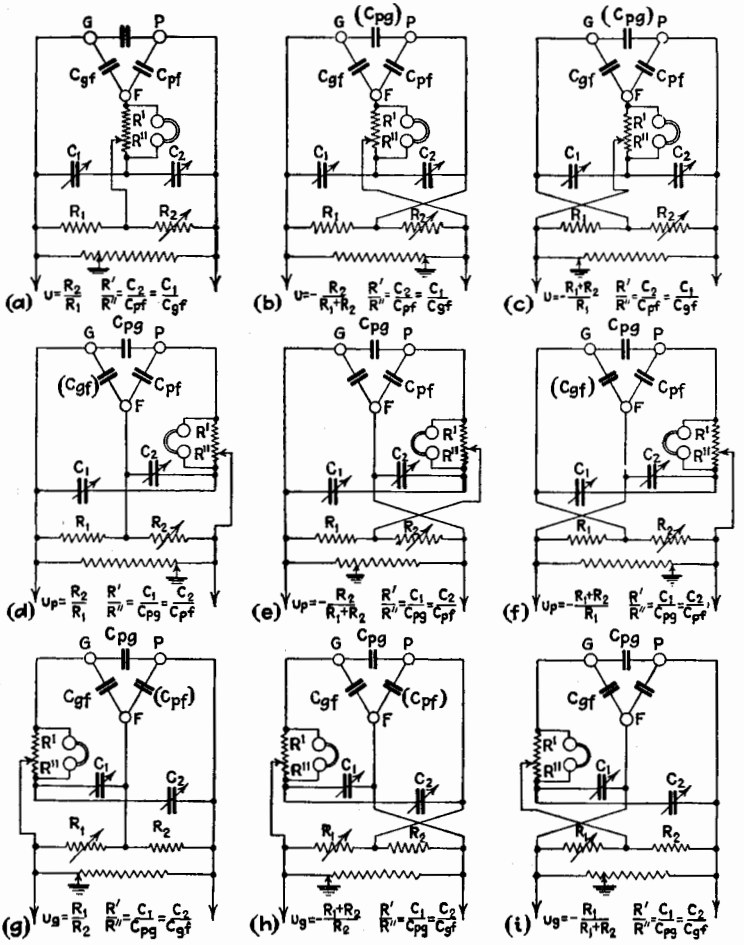


FIG. 118.—Circuits for the measurement of all voltage ratios. Formulae are accurate only when $i_g = 0$.

102. Measurement of u_g.—The quantity u_g is different from zero only when there is a conduction current to the grid, which occurs in high-vacuum tubes only for positive grid voltages. When there is gaseous ionization, there is always a grid current

and u_g is different from zero. Therefore, u_g is of interest when dealing with high-vacuum tubes only when the tube is used as a detector under certain conditions, or as an inverted vacuum-tube voltmeter, or in any other use in which a grid current flows. For gaseous tubes, u_g is of considerable importance. u_g is usually a small positive quantity, but under certain conditions it may assume large negative values.

Positive values of u_g can be measured by the circuit shown in Fig. 118*g*; negative values of u_g can be measured by the circuits of Fig. 118, *h* and *i*.

The nine schemes for measuring voltage ratios are collected for easy comparison in Fig. 118. They can be obtained from a single set-up by the use of reversing switches for R_1 and R_2, and jacks for plugging the telephone receivers, together with their shunt potential divider, into any one of the three positions shown in Fig. 118. The capacitances enclosed in parentheses are not balanced out, but their effect can be made small by giving a small value to resistance R_1 or R_2, whichever shunts the capacitance. The triode capacitances not lettered have no effect on the bridge balance. In all cases R' should be large compared to R'', and a low-resistance steady-current path must be provided by shunting the telephone receivers and the source of alternating potential by low-resistance choke coils.

103. Measurement of k_p.—The coefficient k_p can be measured best by the use of the ordinary four-branch bridge, as shown diagrammatically in Fig. 119. At balance, k_p is given by the relation

$$k_p = \frac{R_2}{R_1 R} \tag{224}$$

It is convenient to give R a fixed value of 10,000 ohms, and R_1 a fixed value of 10 or 100 ohms. The resistance R_2 should be variable in tenths, units, tens, and hundreds of ohms.

The Wagner ground connections should be used as shown rather than using a direct ground connection to the bridge.

The steady plate current passes through a low-resistance choke coil across the telephone receivers and then through the low resistance R_1.

The capacitance $(C_{pf} + C_{pg})$ can be balanced out by a variable condenser C_1 across R_1. The magnitude of C_1 demanded for balance is given by the relation

$$C_1 = (C_{pf} + C_{pg})\frac{R}{R_2} \tag{225}$$

When k_p is very small, R_2 is also small, and, with the connections shown in Fig. 119, an inconveniently large capacitance is required. In such a case, C_1 can be connected across R instead of across R_1. The magnitude of C_1 required in this case is

$$C_1 = (C_{pf} + C_{pg})\frac{R_1}{R_2} \tag{226}$$

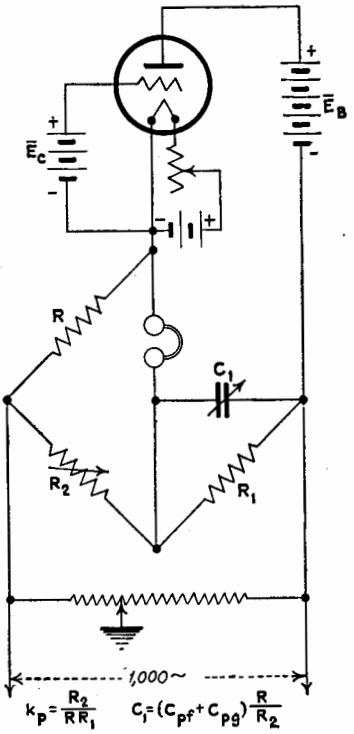


FIG. 119.—First method for the measurement of k_p.

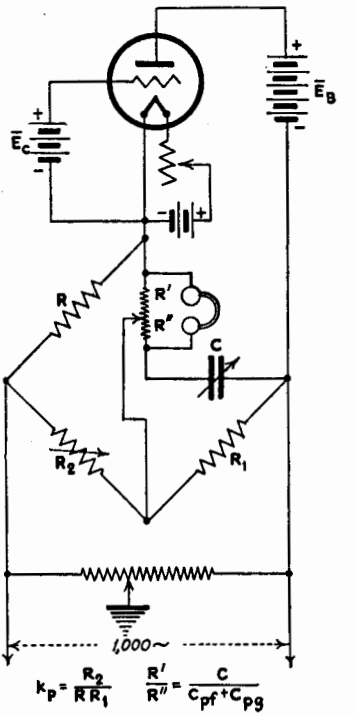


FIG. 120.—Second method for the measurement of k_p.

An alternative and much more convenient method of balancing out the capacitance $(C_{pf} + C_{pg})$ is shown in Fig. 120. The auxiliary bridge, comprising R', R'', $(C_{pf} + C_{pg})$, and C, can be balanced when the filament is cold and R_2 is reduced to zero. R'' should be a small fraction of R' so that C will be of convenient magnitude. With this arrangement, the balance for capacitance is nearly independent of the balance for k_p.

Negative values of k_p can be measured by shunting a known positive conductance across the arm of the bridge which is connected to the tube. This conductance must be such as to make the sum of it and k_p a positive quantity which can then be measured by the bridge. Then the negative value of k_p is

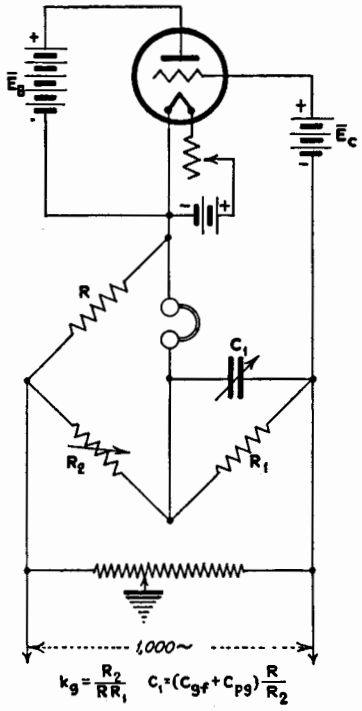


FIG. 121.—First method for the measurement of k_g.

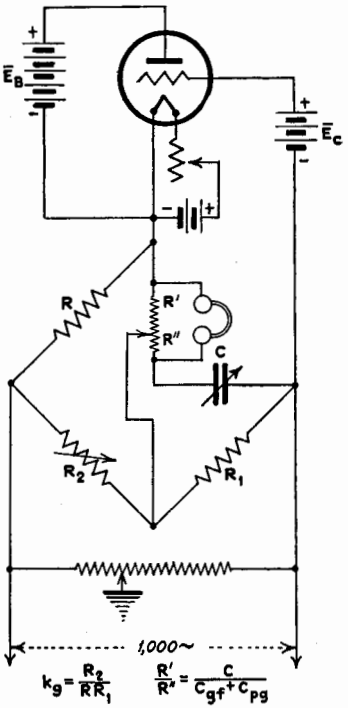


FIG. 122.—Second method for the measurement of k_g.

obtained by subtracting the added conductance from the result obtained from the bridge measurement.

104. Measurement of k_g.—The measurement of k_g is made by a bridge arrangement similar to that described for the measurement of k_p. Figures 121 and 122 correspond to the two arrangements shown in Figs. 119 and 120.

105. Measurement of s_p.—A well-known method of measuring positive values of the transconductance s_p is shown in Fig. 123*a*. If there is no grid current, the grid potential is

$$\Delta E_g = R\Delta I_0 \tag{227}$$

The plate current, according to the e-p-c. theorem, is

$$\Delta I_p = \frac{u_p \Delta E_g}{r_p + R_2} = s_p \Delta E_g \frac{r_p}{r_p + R_2} \tag{228}$$

For no sound in the telephone receivers,

$$(\Delta I_p) R_2 = (\Delta I_0) R_1 \tag{229}$$

Combining Eqs. (227), (228), and (229),

$$s_p = \frac{R_1}{R R_2} \cdot \frac{r_p + R_2}{r_p} \tag{230}$$

Equation (230) is not valid if grid current flows, since account must then be taken of the effect of the reflex voltage ratio u_g and the conductivity of the grid circuit k_g. The complete form of Eq. (230), taking account of these corrections, is

$$s_p = \frac{R_1}{R R_2} \cdot \frac{R_2 + r_p}{r_p} \cdot \frac{R + r_g}{u_g R_1 + r_g} \tag{231}$$

or

$$s_p = \frac{R_1}{R R_2} \cdot \frac{(1 + k_p R_2)(1 + k_g R)}{1 + s_g R_1} \tag{232}$$

For ordinary low-power triodes, it is generally expedient to make R equal to 1,000 ohms, R_2 equal to 100 ohms, and R_1 variable in tenths, units, and tens of ohms. When r_p is very large in comparison with R_2 (see Eq. (231)) and, at the same time, r_g is very large in comparison with R and $u_g R_1$, the arrangement of Fig. 123*a* shows a very convenient direct-reading method of measuring s_p. Unfortunately these conditions do not often obtain, and the calculation of s_p is awkward, requiring a knowledge of r_p and sometimes of r_g and u_g. The factor $\frac{R_2 + r_p}{r_p}$ may differ from unity by several per cent. The other factor $\frac{R + r_g}{u_g R_1 + r_g}$ may differ greatly from unity, but only when grid current flows.

When the circuit of Fig. 123*a* is used, the steady component of the plate current should be forced to pass through the fixed resistance R_2 by the use of high-resistance telephone receivers, or by a large capacitance in series with the telephone receivers. The reason for this suggestion is that if any of the steady plate current passes through R_1, the correction necessary to obtain the steady plate potential is inconvenient, and, of still greater impor-

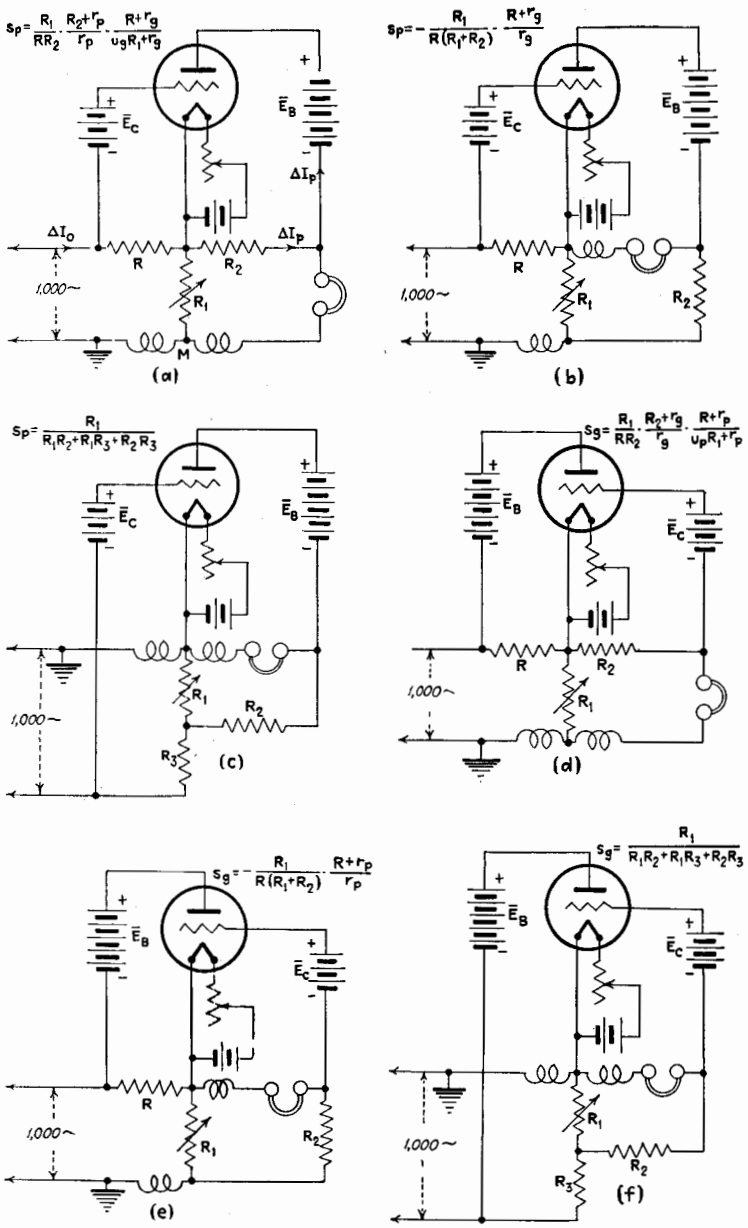


FIG. 123.—Circuits for the measurement of s_p and s_g.

tance, the steady grid potential is dependent upon the voltage drop through R_1 and hence varies with I_p.

The steady grid current, if any, must pass through R_1 and then to the grid by way of a low-resistance choke coil across the source of alternating potential.

In the arrangement of Fig. 123*a*, the quadrature component of current through the telephone receivers due to stray capacitance is small and can be balanced out best by a small mutual inductance connected as shown in the figure. The value of the mutual inductance need not be taken into account in the calculation of s_p. The quadrature component can also be balanced out by an arrangement of capacitance and ratio arm similar to the arrangements used in the measurement of u_p and k_p.

An alternative method of measuring positive values of s_p is shown in Fig. 123*c*. This second method, although not so convenient for calculation as the method of Fig. 123*a* when the corrections are negligible, is accurate under all circumstances, requiring no correction factors and no correction of the steady grid and plate potentials. The value of s_p using the method of Fig. 123*c* is given by

$$s_p = \frac{R_1}{R_1R_2 + R_1R_3 + R_2R_3} \tag{233}$$

Calculation by this formula is facilitated by giving R_2 and R_3 constant values, such as 500 ohms, and making R_1 variable in tenths, units, and tens of ohms. Large values of s_p are conveniently measured by making R_3 zero, in which case s_p is equal to the reciprocal of R_2. This latter modification of the method is particularly convenient in measuring the transconductance of power tubes.

The path for the steady plate current is through a low-resistance choke coil across the telephone receivers, and the path for the steady grid current is through a similar shunt across the source of alternating potential.

Negative values of s_p can be measured by the circuit shown in Fig. 123*b*. In this arrangement s_p is given by

$$s_p = -\frac{R_1}{R(R_1 + R_2)} \cdot \frac{R + r_g}{r_g} \tag{234}$$

106. Measurement of s_g.—The reflex transconductance s_g is of little practical importance, but, because of its occasional use

and to make the description of the methods of measuring tube coefficients complete, the method of determining s_g is given here.

Figure 123*d* gives the circuit arrangement for measuring negative values of s_g, and the circuit of Fig. 123*e*, or the alternative arrangement shown in Fig. 123*f*, enables positive values of s_g to be measured. The formulas for the calculation of s_g are given in the figures.

CHAPTER X

EFFECTS OF GAS IN A TRIODE

The preceding chapters have been devoted to the study of high-vacuum or hard triodes, where the term *high-vacuum* is used to indicate a degree of exhaustion such that the traces of gas or vapor remaining have no appreciable effect upon the conduction through the triode at potentials well above the ionizing potentials of the gas. When the amount of gas is sufficient to have an appreciable effect upon the character of the conduction through the triode, the tube is said to be *soft*. In this chapter the variety of effects obtained in a soft triode under all conditions of gas pressure and potential will not be described. This chapter is devoted to a brief description of the principal deviations from the static and dynamic characteristics of a hard tube caused by a *very small* trace of gas. It includes a description of the operation of the ionization gauge, a practical device of great utility. When the amount of gas is increased or the potentials impressed on the tube are increased beyond the ranges considered in this chapter, the discharge changes, generally suddenly, into the glow discharge. The study of the glow discharge and of practical glow-discharge devices is entirely outside the scope of this chapter.

107. General Effects of Small Traces of Gas.—Before describing the specific changes in the characteristic curves of a triode caused by a trace of gas, we shall consider some of the general effects upon the operation of a triode which result from the presence of gas.

As pointed out in Chap. V, certain gases which combine chemically with the active emitting material of the cathode may cause a considerable decrease in the electron emission from the cathode. Langmuir[1] has studied very fully the effects of various gases in changing the emission of tungsten, and reference is made to the original papers for detailed information. Oxygen is particularly

[1] LANGMUIR, *Phys. Zeits.*, **15**, 520 (1914); *J. Am. Chem. Soc.*, **40**, 1361 (1918); *Phys. Rev.*, **2**, 450 (1913).

active at high temperatures in combining with tungsten, and with the active islands of barium, strontium, or other alkali-earth metals which form the emitting centers of the oxide-coated cathodes. Oxygen is also particularly destructive of the monatomic films of thorium or caesium. The emission of tungsten may be reduced in the ratio of one to several hundred thousand by a trace of oxygen. Nitrogen also combines chemically with tungsten causing reduction in emission. Water vapor, carbon dioxide, and other normally inert compounds may be dissociated by the hot cathode and may then reduce the emission by chemical action. In general the highly electronegative gases, such as chlorine, are deleterious, whereas the highly electropositive or reducing gases, such as hydrogen and carbon monoxide, have no harmful chemical effect at the cathode. Many vapors, such as that of mercury, and the inert gases, such as helium, argon, and neon, apparently have no effect upon the emission of the several types of thermionic cathodes except for the effect now to be described.

Besides the chemical effect at the cathode, the presence of gas may result in a purely physical effect known as *sputtering*. The positive ions produced by ionization may be driven against the cathode with sufficient velocity to knock off particles of molecular dimensions from the cathode, and these particles are deposited on the walls of the tube. The walls become blackened and the cathode reduced in size. This effect is very important at the higher gas pressures and at potentials which are sufficiently high to produce glow discharge. The effect is present even at low pressures and is of particular importance in tubes using cathodes having monatomic films. These delicate films are rapidly knocked off by the slightest positive-ion bombardment when the ions have high velocities. Kingdon and Langmuir[2] have investigated the sputtering of monatomic films resulting from bombardment by various positive ions and find that for each kind of ion there is a critical velocity below which no appreciable sputtering occurs. These critical velocities expressed in terms of the equivalent potential drop E_0, as given by Kingdon and Langmuir for a cold surface, are given in Column 2 of Table XII. A. W. Hull[3] observed somewhat lower critical potentials when the bombarded surface is hot, as for an operating cathode, and when the current

[2] KINGDON and LANGMUIR, *Phys. Rev.*, **22**, 148 (1923).

[3] HULL, *Trans. A.I.E.E.*, **47**, 798 (1928).

TABLE XII.—CRITICAL ENERGY OF BOMBARDING ION FOR SPUTTERING OF THORIUM FILM ON TUNGSTEN

1	2	3
Ion	E_0 (cold surface),[1] volts	E_0 (hot surface),[2] volts
H	>600	
He	About 35	
Ne	45	27
Ar	47	25
Cs	52	
Hg	55	22

[1] Kingdon and Langmuir.
[2] Hull.

density is greater than that used by Kingdon and Langmuir. Hull's results are recorded in part in Column 3 of Table XII. The curves of Fig. 124 show Hull's results on the variation of the emission of a thorium-coated filament in mercury vapor at a pressure of 0.005 mm. At the critical anode potential, the emission rapidly decreases. Figure 124 shows the dependence of sputtering voltage on temperature. Hull[4] has also investigated the effect of bombardment of mercury ions on the emission of oxide-coated cathodes and finds a critical energy corresponding to about 25 volts, below which sputtering of the active emitting material does not take place, and the cathode has a long life. From these results there has developed a line of very valuable high-current, hot-cathode, mercury-vapor tubes.

The changes in the electrical characteristics of a triode, aside from the change in emission, are due almost entirely to the various actions of the positive ions. Hence they occur only for potentials above ionization potential, although small kinks may be produced in the characteristic curves by inelastic collisions at the resonance potentials of the gas. These small kinks are elusive and generally of little importance.

Whenever there are positive ions present, they may be drawn into the region of negative space charge around the cathode, resulting in a partial neutralization of the space charge. One positive ion, because of its low velocity, neutralizes the negative space charge due to hundreds or thousands of electrons. There-

[4] HULL, *Gen. Elec. Rev.*, **32**, 213, 390 (1929).

fore, a small amount of ionization may considerably alter the space charge and hence the space current, if the current is less than the saturation current for the cathode. This effect was explained in Chap. V and will be described more in detail in the next section.

The space current through a tube may be limited not only by the space charge but also by negative charges accumulated on the walls of the tube or on other insulating members, such as

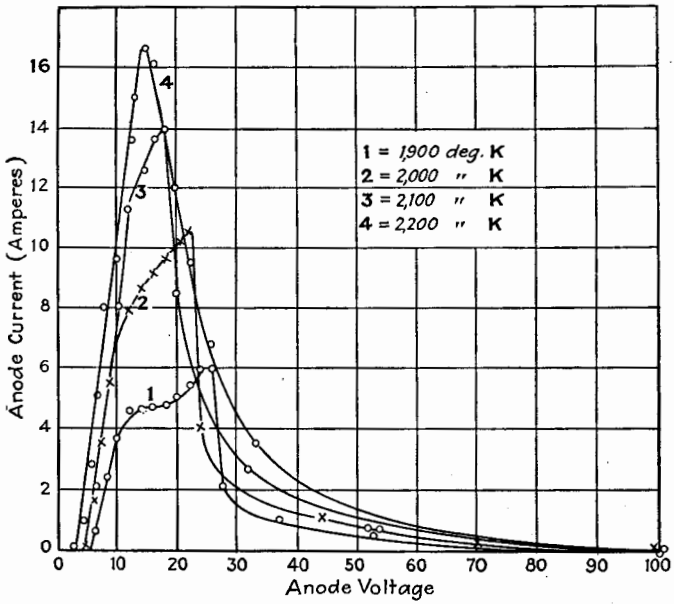


FIG. 124.—Characteristics of a thoriated filament in mercury vapor at 0.005-mm. pressure. (*Hull.*)

insulating supports or insulated conductors. These wall charges are more important the more open the structure of the electrodes. The positive ions produced in a soft tube may also partially or completely neutralize these wall charges and so affect the current.

The positive ions produced in a triode drift in the direction of the negative potential gradient. The positive-ion current, although small in comparison with the electronic space current producing the ionization, is nevertheless appreciable and must be considered in studying the operation of the triode.

Because of the high velocity of the electrons passing from cathode to the other electrodes of a triode, pure electronic conduc-

tion has no lag except for the very high radio frequencies. Gaseous ions travel so much more slowly than electrons that the conduction through a soft tube is generally accompanied by a lag which may be appreciable even at audio frequencies. Hence the dynamic characteristics of a soft tube are in general complex quantities. The equivalent circuits of a soft triode are not as shown in Figs. 93, 94, 95, and 96, but contain, in addition to the resistances r_p and r_g, certain equivalent series reactances x_p and x_g which depend upon the kind of gas and the potentials. The question of power interchange between the grid and plate circuits when ionization is present was considered briefly in Chap. IX.

Ionization, being a discontinuous process, imposes a random fluctuation in the currents through a triode. This is of much the same nature as the "shot effect" of pure electronic conduction, but its magnitude is much greater. At and near the ionization potentials this shot effect, enhanced sometimes by ionic oscillations in the triode, often produces a loud hiss in the telephone receivers placed in the current circuit. This hiss is often an objectionable effect limiting the use of soft tubes which have, however, other very desirable characteristics. For potentials well above the ionization potentials, this hiss is often very slight.

108. Effects of Trace of a Gas on the Static Characteristic Curves of a Triode.—The most important changes in the static characteristic curves of a triode caused by a trace of gas are illustrated by the curves of Figs. 125 to 129, which were obtained for a triode similar to the one used in obtaining the characteristic curves of Figs. 53, 62, 63, 64, and 65, except that after the tube was thoroughly exhausted a small amount of mercury was introduced and the tube sealed from the pump. Any desired pressure of mercury vapor could be obtained by immersing the tube in an oil bath at a suitable temperature. The curves of Figs. 125 to 129 were taken for a temperature of 26°C. which gives a pressure of approximately 0.002 mm. of mercury, or 2 microns. Mercury vapor was used because of the convenience in obtaining and maintaining any desired pressure, and because mercury vapor has no effect upon the emission of the tungsten filament. The static characteristic curves are essentially similar in shape to those obtained for gases other than mercury, although the scale of pressure and potential may be slightly shifted because of the difference in mobility and in mass of various gas ions and the difference in the ionization potentials.

Figure 125 gives the $i_p - e_p$ characteristic curves for the soft tube containing mercury vapor at a pressure of 0.002 mm. Since the ionizing potential of mercury is about 10.4 volts, the portions of the curves for grid and plate potentials of less than 10 volts is practically the same as for a hard tube of the same dimensions. The rapid rise in current as the plate potential exceeds 10 volts is due to the neutralization of the space charge by the positive ions resulting from ionization.

The effect of the positive ions in partially neutralizing the space charge is also shown by the plate-current curves at constant plate voltages, Fig. 126. The curve for a plate voltage of 8 volts

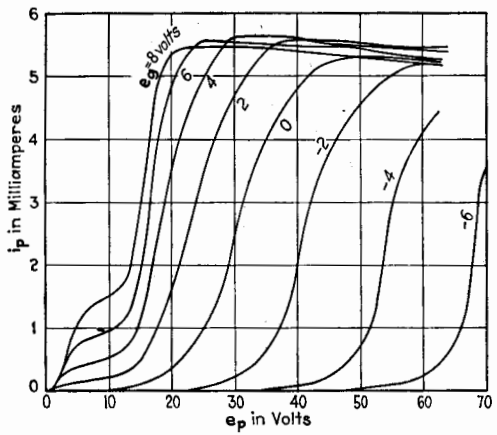


FIG. 125.—Plate-current *vs.* plate voltage for constant grid voltage in mercury vapor at pressure of 0.002 mm.

shows no effect of ionization. As the plate voltage exceeds ionizing voltage, the steepness of the curves increases. The curves of Fig. 126 may be compared with those of Fig. 63 for a hard tube of approximately the same dimensions.

Figure 127 shows the curves of constant total space current which correspond to those of Fig. 53. Ionization exists for all points on the diagram above the horizontal line for a plate voltage of about 10 volts and destroys the parallelism of the curves or, in other words, makes the factor u less constant than for the same tube without ionization. Figure 128 gives the curves of constant plate current.

The effects of ionization are also observable from the shape of the grid-current curves. Figure 129 shows the grid current

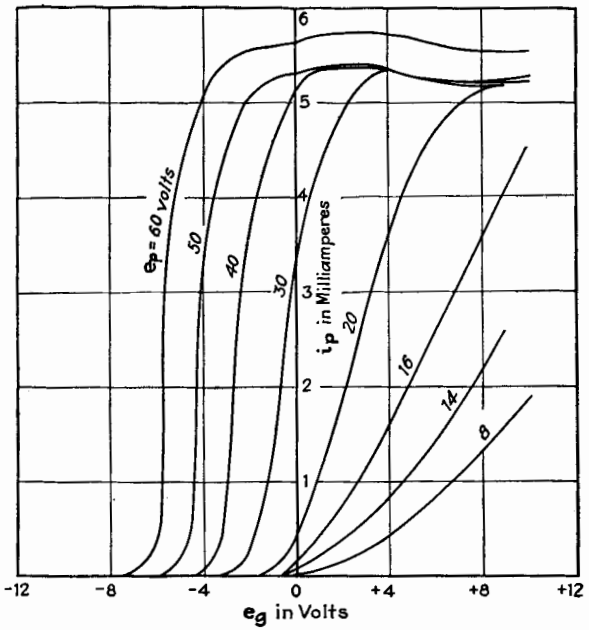


FIG. 126.—Plate current *vs.* grid voltage for constant plate voltage in mercury vapor at pressure of 0.002 mm.

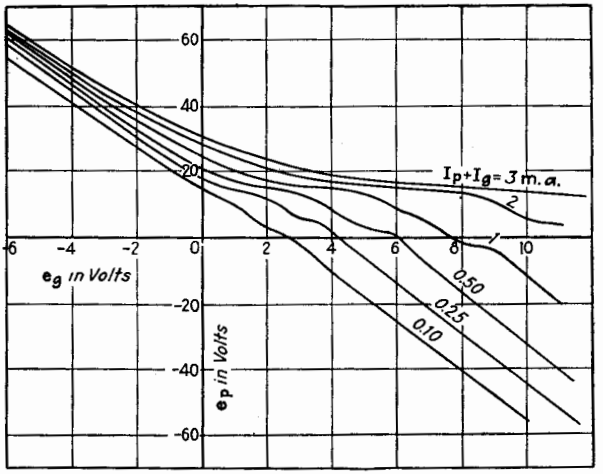


FIG. 127.—Curves of constant total space current in mercury vapor at a pressure of 0.002 mm.

plotted against grid voltage for various constant plate voltages. The curve for a plate voltage of 8 volts, which is below ionization voltage, is a normal curve for a high-vacuum triode and corresponds to the curves of Fig. 64. For this plate voltage, the grid current is a pure electron current. The curve for a plate voltage of 16 volts shows the marked effect of the partial neutralization of the space charge by the positive ions. For a plate voltage of 20 volts a small negative grid current is observed for grid voltages from about -1 to $+1.1$ volts. This negative current is due to positive ions produced in the tube at regions

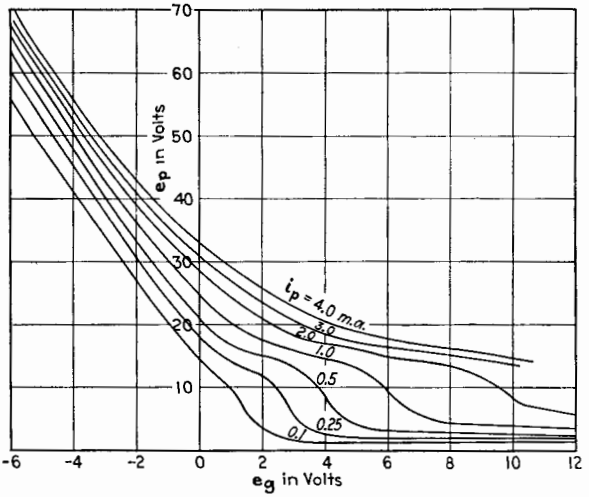


FIG. 128.—Curves of constant plate current in mercury vapor at a pressure of 0.002 mm.

where the potential is more than $+10.4$ volts above the potential of the filament, and these positive ions drift in the direction of negative potential gradient toward the grid and filament. Some go to the grid and some to the filament, the fraction going to each electrode depending upon the distances of the electrodes from the point of ionization and upon their potentials. Besides the positive-ion current, an electron current flows to the grid for positive grid potentials. The actual grid current is the difference between the electron current and the ionization current. These two currents neutralize each other for a grid potential of 1.15 volts and plate potential of 20 volts. The positive-ion current is much more marked for plate potentials

of 40 and 60 volts, as shown by the curves in Fig. 129. The plate-current curve for a plate voltage of 40 volts is added to show that the positive-ion current and plate current increase together. At constant plate voltage, the positive-ion current is approximately proportional to the plate current which produces the ions, and to the pressure of the gas. The grid current for negative grid voltages is entirely positive-ion current, since no electrons go to a negatively charged grid. When the grid voltage is highly negative, practically all of the positive ions go to the grid and many are swept out before they recombine with electrons. The ratio of the positive-ion current to the plate current is constant within experimental error. For example, in Fig. 129, when e_p is 40 volts and when the grid voltage ranges from -4 to -2 volts, this ratio has the value 0.0024. As the grid voltage is increased positively from about -2 volts, a change takes place as shown by both the grid- and plate-current curves. This change, which is also evident on the curve for a plate voltage of 60 volts, may have been due to the beginning of ionic oscillations within the tube, or to a sudden change in the distribution of potentials. For grid voltages on the positive side of -2 volts, the ratio of the ionic grid current to the plate current for a plate voltage of 40 volts increases from the value 0.0024 just given to a maximum of 0.00403 at a grid voltage of about $+1$ volt.

One commonly applied test for appreciable gas in a triode is to measure the positive-ion current to the negatively charged grid when a sufficiently high plate voltage is impressed to cause a fair plate current to flow. A high-vacuum triode will show less than a microampere of positive-ion current.

109. Effects of Small Traces of Gas on the Dynamic Characteristic of a Triode.—The effect of the presence of gas on the dynamic characteristics can best be determined by measuring the dynamic characteristics by the methods described in Chap. IX. However, a rough idea of the trend of the dynamic characteristics can be easily obtained from an examination of the static characteristic curves just described.

One of the most striking effects of gas is the great increase in the transconductance s_p of a soft tube, as shown by the steepness of the curves of Fig. 126. For example, the transconductance derived from the 60-volt curve of Fig. 126 is about 1,500 micromhos as compared to about 370 micromhos for the corre-

sponding high-vacuum triode. In general, however, the inferior stability of a soft tube and the large amount of tube noise make the advantage of such high transconductance of little avail.

Referring now to Fig. 129, it is obvious that k_g can be negative, as well as positive. When the grid conductance is negative, ionic oscillations may arise inside the tube and in the attached circuits, even if the external circuits contain only resistances. If an oscillatory circuit is connected between filament and grid,

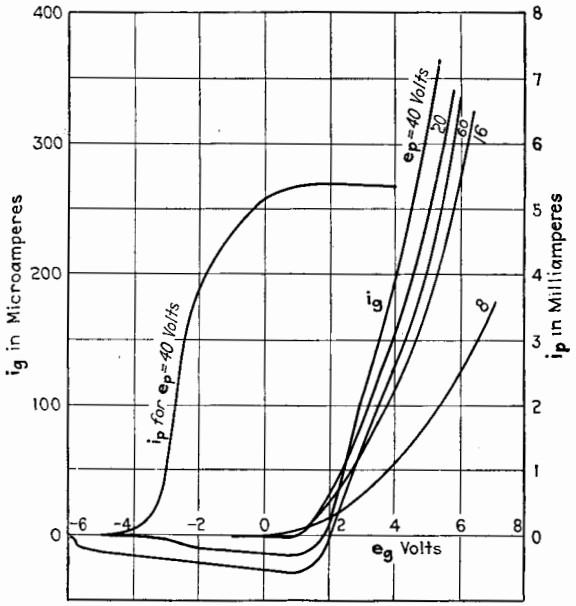


FIG. 129.—Grid current *vs.* grid voltage for constant plate voltage in mercury vapor at a pressure of 0.002 mm.

oscillations may be generated in the circuit if the negative conductance of the filament-to-grid path neutralizes the positive conductance of the oscillatory circuit.

The coefficients u_g and s_g can assume both positive and negative values which are numerically much greater than for the corresponding high-vacuum tube.

The coefficients of a soft tube are actually complex quantities, owing to the lag in the motion of the heavy positive ions. In the brief description of the shape of the dynamic characteristic curves just given, this lag effect was neglected in order that a general picture might be obtained, and the dynamic coefficients

were taken as the slopes of the static characteristic curves. Taking account of the lag effects, the dynamic coefficients are not simple slopes of curves but must be defined in terms of complex quantities at a definite frequency n as given below:

$$(\mathbf{u}_p)_n = -\left(\frac{\Delta \mathbf{E}_p}{\Delta \mathbf{E}_g}\right)_{i_p=\text{const.}} = (u_p)_n' + j(u_p)_n' \tag{235}$$

$$(\mathbf{k}_p)_n = \left(\frac{\Delta \mathbf{I}_p}{\Delta \mathbf{E}_p}\right)_{e_g=\text{const.}} = (g_p)_n + j(b_p)_n \tag{236}$$

$$(\mathbf{r}_p)_n = \left(\frac{\Delta \mathbf{E}_p}{\Delta \mathbf{I}_p}\right)_{e_g=\text{const.}} = (r_p)_n + j(x_p)_n \tag{237}$$

$$(\mathbf{s}_p)_n = \left(\frac{\Delta \mathbf{I}_p}{\Delta \mathbf{E}_g}\right)_{e_p=\text{const.}} = (g_{pg})_n + j(b_{pg})_n \tag{238}$$

The other coefficients $(\mathbf{u}_g)_n$, $(\mathbf{k}_g)_n$, $(\mathbf{r}_g)_n$, and $(\mathbf{s}_g)_n$ are similarly defined.

The circuits given in Chap. IX for the measurement of the simple dynamic coefficients may be extended to the measurement of the complex coefficients also. For example, Fig. 117*b* may be adapted to the measurement of complex values of u, or of u_p for negative grid voltages, by adding a variable capacitance in parallel with R_1. The capacitances of the tube and connections should first be balanced when the cathode is cold or when the potentials are such as to produce no space current. Then it is easily shown that if C_1' is the added capacitance,

$$(u_p)_n' = \frac{R_2}{R_1} \text{ and } (u_p)_n'' = 2\pi n C_1' R_2 \tag{239}$$

Similarly, the circuits of Figs. 119 and 120 are adapted to the measurement of $(\mathbf{r}_p)_n$ by adding a variable inductance in series with R or a capacitance in parallel with R_2. As before, the capacitances of the tube and connections are first balanced with no space current. The capacitance across R_2 is probably the most convenient arrangement, in which case

$$r_p = \frac{R_1 R}{R_2} \text{ and } (x_p)_n = 2\pi n (l_p)_n = 2\pi n R R_1 C_2' \tag{240}$$

The circuits of Chap. IX for the measurement of the other coefficients of a triode can readily be adapted to the measurement of the complex coefficients.

That the coefficients are complex quantities when ionization takes place is demonstrated by the values of $(\mathbf{u}_p)$ and $(\mathbf{r}_p)$ at

a frequency of 1,000 cycles per second, shown in Figs. 130 and 131. The triode used is known as UV200 and contains a trace of some permanent gas, such as argon or nitrogen. In both figures the quantity being measured has no quadrature component until the plate potential reaches the voltage of about 15 volts, at which point ionization begins. For voltages above 15 volts the quadrature component rises rapidly to a maximum

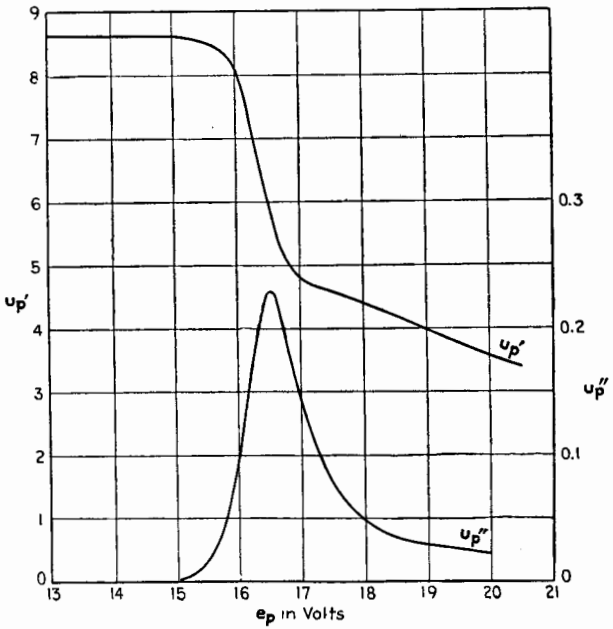


FIG. 130.—Components of $\mathbf{u}_p$ for a triode containing gas. $\mathbf{u}_p = u'_p + ju''_p$ $e_g = 0$.

and the real component decreases. In Fig. 131, the ratio x_p/r_p is also plotted.

Figure 132 shows the variation of r_p and x_p with frequency for certain selected values of the electrode voltages.

110. Static Characteristic Curves of a Soft Triode for Negative Plate Voltages.—No current flows to the plate of a high-vacuum tube when the plate voltage is more than a few tenths of a volt negative, but with a soft tube positive ions are attracted to a negatively charged plate. The static characteristic curves for negative plate voltages are of particular interest as they explain the action of a triode when used as an *ionization gauge.*

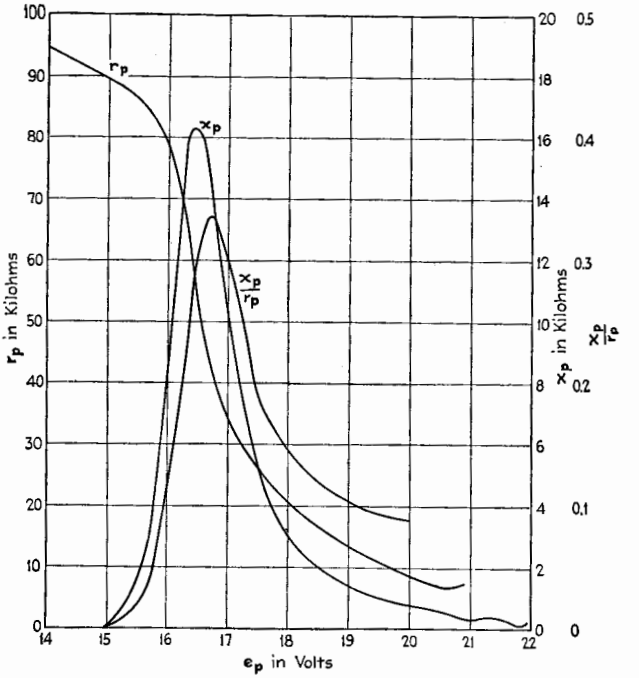


FIG. 131.—Components of z_p for a triode containing gas. $z_p = r_p + jx_p$. $e_g = 0$.

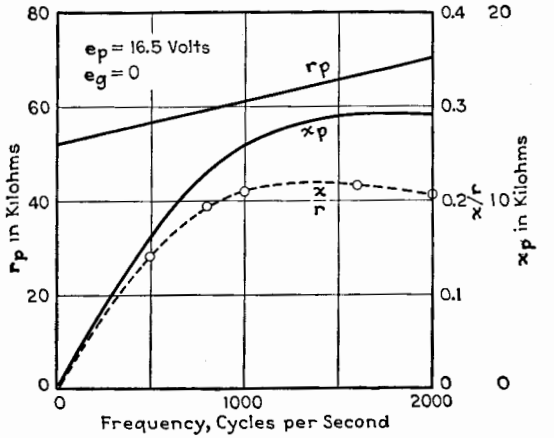


FIG. 132.—Variation of r_p and x_p with frequency. $z_p = r_p + jx_p$.

When the grid voltage is positive and the plate voltage zero or negative with respect to the cathode, electrons are drawn toward the grid. Some strike the grid directly, while others shoot through the grid mesh to be turned back by the opposing field between grid and plate and arrive at the grid after making one or more excursions back and forth through the grid. If the grid voltage just exceeds the ionization potential of the gas or vapor, ions are formed only in a thin layer near the grid. Some of the positive ions drift to the cathode, while others, especially those formed on the plate side of the grid, are attracted to the negative plate. If the grid voltage is much greater than the ionization voltage, ionization can take place in a much larger region between the cathode and grid and also between the grid and plate.

The total number of positive ions produced is proportional to the pressure of the gas or vapor, to the magnitude of the grid current, to the volume of the space traversed by the electrons, and depends upon the kind of gas or vapor present. The fraction of the number of positive ions which are collected by the negatively charged plate depends upon the plate potential and the geometry of the tube. The positive ions which are formed near the cathode are in a field which urges them to the cathode. Only those positive ions which are formed in a region where the electric field is toward the plate will be collected by the plate. The limit of this region lies somewhere between the grid and the cathode, its position depending upon the geometry of the tube, the potentials of the electrodes, and the space charge around the grid. This limit usually lies fairly close to the grid.

These statements are verified by the following experimental curves of Figs. 133 to 135. Figure 133 shows the rapid increase in positive-ion current to the plate of a triode containing mercury vapor, as the grid potential was increased. The ionization potential of mercury is 10.4 volts but, owing to a contact potential in the circuit, ionization began at 11.8 volts. The left-hand curve of the figure is plotted to the left-hand scale, and the curve on the right to the right-hand scale. In this experiment the temperature of the tube was 23.5°C., corresponding to a vapor pressure of about 1.6 microns, and the plate potential was −3 volts.

Figure 134 shows curves of positive-ion current plotted against plate voltage when the grid voltage is held constant at values not exceeding 18.5 volts. Under these conditions only mercury is

ionized. Another test showed that, if the grid voltage exceeded about 20 volts, the plate current showed a rapid increase due to other ions. The curves of Fig. 134 show first a rapid *rise* due to a rapid increase in the fraction of positive ions that are drawn to the plate. The curves show also a *decrease* for plate voltages more negative than -2 volts. This decrease is due, first, to the decrease in the grid current occasioned by the negative plate and,

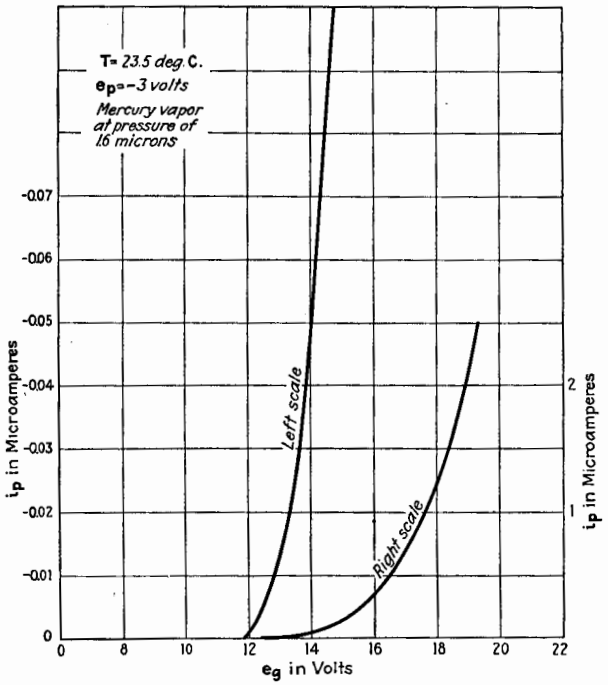


FIG. 133.—Positive-ion current *vs.* grid voltage for constant plate voltage in a triode containing mercury vapor at a pressure of 0.0016 mm.

second, to the decrease of the volume in which the electrons produce ions. This decrease in volume is a result of the repelling action of the negative plate and the consequent reduction in the distance the electrons travel toward the plate before they are reversed in direction.

The first of these causes for the downward slope of the curves of Fig. 134 can be eliminated by keeping the grid current constant. This is done by controlling the filament voltage. The curves of Fig. 135 were thus obtained and exhibit a slope which is less than that in Fig. 134.

If the grid voltage is very large compared to the plate voltage, the plate current is nearly constant when the grid current is maintained constant.

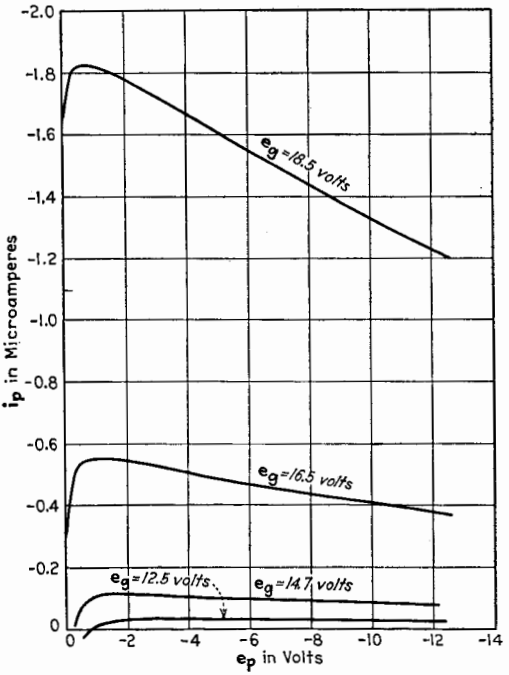


FIG. 134.—Positive-ion current *vs.* plate voltage for constant grid voltage in a triode containing mercury vapor at a pressure of 0.0016 mm.

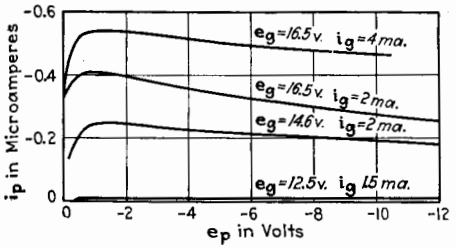


FIG. 135.—Positive-ion current *vs.* plate voltage for constant grid voltage and grid current in a triode containing mercury vapor at a pressure of 0.0016 mm.

The positive-ion plate current is practically linear with respect to grid current over a considerable range, provided the grid voltage is held constant at a large value and the plate voltage has

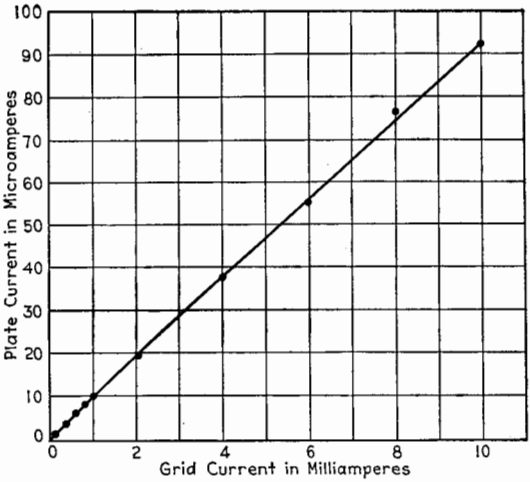


FIG. 136.—Positive-ion current *vs.* grid current for constant grid and plate voltages.

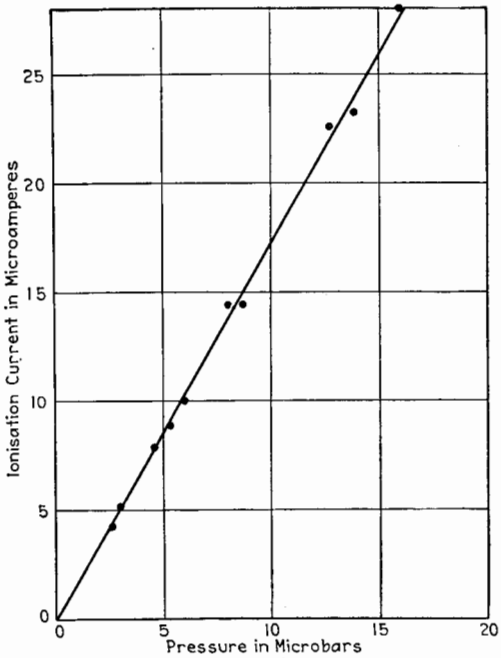


FIG. 137.—Positive-ion current *vs.* pressure.

a constant negative value. This linearity is shown in Fig. 136, for which the conditions are as follows: $e_g = 250$ volts, $e_p = -22$ volts, grid current controlled by temperature of the cathode. The graph of Fig. 136 might deviate from linearity for larger grid currents, owing to the increasing space charge around the plate and the consequent shortening of the paths of those electrons that shoot through the grid toward the plate.

When the grid potential is low, the curve of plate current against grid current may be far from linear. This is the case for the mercury tube using grid voltages of less than 20 volts.

Over a certain range of pressure, the positive-ion current to the plate may be proportional to the pressure, other factors being constant, as shown in Fig. 137. At higher pressure the law is not linear, owing to cumulative ionization. The linear law is not followed if the grid voltage is very low. This deviation exists because the filament voltage, which is varied to maintain the grid current constant, may then be an appreciable fraction of the grid voltage, and hence the volume in which ionization takes place is no longer constant.

111. The Ionization Gauge.—Buckley[5] was the first to study the characteristics of a triode as applied to the measurement of gas pressures. His method consists in measuring the positive-ion current to a negatively charged electrode or *collector*, as described in the preceding section.

Almost any triode will serve as an ionization gauge, although a linear relation between the positive-ion current and the pressure cannot be assumed. A triode used as an ionization gauge should have a pure-tungsten or an oxide-coated filament and be provided with a wide tubular connection to the vacuum system. Preferably the plate should surround the grid and filament to eliminate effects due to charges on the glass walls of the tube. The insulation from the collector electrode to the grid and filament should be especially thorough, because the collector current is very small at low pressures and a slight leakage would cause serious error. The tube should be unbased to aid in securing thorough insulation and because the tube must be subjected to high temperature during preparation.

Either the grid or the plate may be used as the collector for the positive ions. The sensitivity is considerably greater when the

[5] BUCKLEY, *Nat. Acad. Sci.*, **2**, 683 (1916).

plate is the collector, although the calibration may be linear over a greater range when the grid is the collector.

In order that the gauge shall indicate the true pressure, all gas occluded in the glass and electrodes of the gauge must be eliminated. This is done by first baking the gauge at a temperature of about 360°C. for an hour or so, while it is being exhausted by the pumps. The electrodes must then be heated to a bright red by electronic bombardment or by an induction furnace and must be maintained at this high temperature for at least 10 min., and longer if necessary, to drive off all gas.

The anode voltage is held constant at 100 to 200 volts and the collector potential is set at about -22 volts. The anode current is controlled by varying the temperature of the cathode and can be given any constant value according to the range of pressures being measured. In some of the commercial gauges the anode current is held constant at 20 milliamp. when low gas pressures are measured. Under these conditions a particular gauge with argon gas gave 1 microamp. for a pressure of 0.0132 bar or 0.00990 micron. For pressures greater than about 0.5 micron, the anode current for the gauge in question was reduced to 2 milliamp., which decreased the sensitivity of the gauge by a factor of 10.

Emphasis should be laid on the fact that the calibration of an ionization gauge is different for different gases. According to Dushman and Found,[6] the ionization current for the same pressure, filament current, and electrode voltages is roughly proportional to the number of electrons in the molecule of the gas. This number is 2 for H_2, 16 for O_2, 14 for N_2, 18 for A, 80 for Hg, 53 for I, 10 for H_2O, etc. The variation of sensitivity of the ionization gauge with composition of the gas reduces its value as an absolute gauge. It is of considerable value, however, when the composition of the gas is known and is useful in giving relative values of different pressures in the same vacuum system.

[6] DUSHMAN and FOUND, *J. Franklin Inst.*, **188**, 819 (1919).

CHAPTER XI

INPUT AND OUTPUT ADMITTANCE OF A TRIODE

The static characteristic curves of a high-vacuum triode show a conduction current to the grid only when the grid voltage is positive or very slightly negative. When ionization takes place in a triode containing gas, a conduction current flows to the grid whenever a plate current flows. If the plate voltage is constant, the conductance of the grid-to-filament path within the triode is k_g. When the triode is operated with its associated circuits, the plate voltage does not remain constant, and the reaction of the plate voltage on the grid current causes the *equivalent input admittance* or *input impedance* of the grid-to-filament path to be different from k_g or r_g. The equivalent input admittance or input impedance of a triode is defined as the simple admittance or series impedance which could be substituted for the triode and the associated circuits connected to the plate and would take from the circuits connected to the grid an alternating current which is the same in magnitude and phase as that which actually flows to the grid of the triode.

The alteration of k_g or r_g, as affected by the nature of the plate load, will first be studied. This preliminary study is of value only when the frequency of the alternating currents is very low. At high frequencies the capacitances between the electrodes of the triode offer additional current paths which greatly affect the equivalent input admittance or impedance of the triode. Following the preliminary study, a complete analysis of the input admittance of the triode will be made. This study will show that even when the conduction current between grid and filament is zero, as in a high-vacuum triode with a negatively polarized grid, the input admittance is appreciable and must be considered in the design of all radio-frequency systems using triodes. The importance of this study cannot be overemphasized.

In the following analysis the same restrictions are imposed as in the preceding chapters: the alternating variations are

assumed to be sinusoidal and of so small an amplitude that the tube coefficients may be assumed constant.

***112. Input Admittance for Very Low Frequencies, and When $k_g \neq 0$.**—The input admittance of a triode when the currents through the interelectrode capacitances are negligible, and as affected by the reaction of the plate voltage upon the conduction current from grid to filament, can be readily deduced by combining the e-p-c. theorem and the e-g-c. theorem.

According to the e-p-c. theorem,

$$r_p \Delta \mathbf{I}_p = u_p \Delta \mathbf{E}_g + \Delta \mathbf{E}_p \tag{241}$$

Since $\Delta \mathbf{I}_p = -\dfrac{\Delta \mathbf{E}_p}{\mathbf{Z}_b}$, Eq. (241) becomes

$$\frac{r_p + \mathbf{Z}_b}{\mathbf{Z}_b} \cdot \Delta \mathbf{E}_p = -u_p \Delta \mathbf{E}_g \tag{242}$$

According to the e-g-c. theorem

$$r_g \Delta \mathbf{I}_g = \Delta \mathbf{E}_g + u_g \Delta \mathbf{E}_p \tag{243}$$

Eliminating $\Delta \mathbf{E}_p$ from Eqs. (242) and (243), we obtain as the equivalent grid-circuit impedance of a triode

$$\mathbf{z}_g = \frac{r_g}{1 - \dfrac{u_g u_p \mathbf{Z}_b}{r_p + \mathbf{Z}_b}} = r_g \cdot \frac{r_p + \mathbf{Z}_b}{r_p + \alpha \mathbf{Z}_b} \tag{244}$$

where

$$\alpha = 1 - u_g u_p \tag{245}$$

From Eqs. (242) and (243), the equivalent grid-circuit admittance is

$$\left.\begin{aligned} \mathbf{y}_g &= k_g \left(1 - \frac{u_g u_p \mathbf{Z}_b}{r_p + \mathbf{Z}_b}\right) \\ &= k_g \cdot \frac{r_p + \alpha \mathbf{Z}_b}{r_p + \mathbf{Z}_b} \end{aligned}\right\} \tag{246}$$

$$= k_g \cdot \frac{\alpha k_p + \mathbf{Y}_b}{k_p + \mathbf{Y}_b} \tag{247}$$

where $\mathbf{Y}_b$ is the admittance of the plate-circuit external load.

Examination of Eqs. (244), (246), and (247) shows that the equivalent grid-circuit impedance and admittance at low frequencies are functions of the quantity α, which is defined by Eq. (245). If u_g and u_p are real quantities, as they are in a hard tube at all frequencies except those of the order of 10^7 cycles

per second and higher, α is a real quantity. Usually u_g is negative, making α positive. It is clear that in general the equivalent grid-circuit impedance is *less* than r_g, and the admittance is *greater* than k_g, under the initial assumption of negligible capacitance currents.

A rough idea of the value of $u_g u_p$ can be obtained from the static characteristic curves. Figure 138 shows two $i_p - e_g$ curves for two constant values of plate voltage, $\bar{E}_p$ and $\bar{E}_p + \Delta\bar{E}_p$.

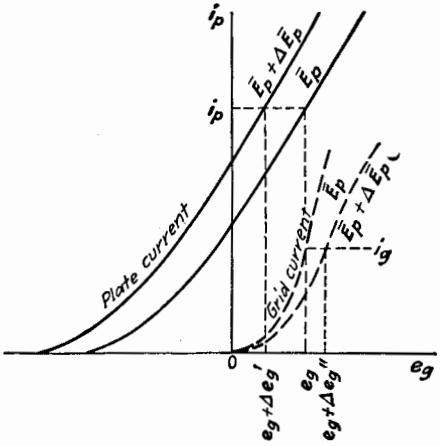


FIG. 138.—Evaluation of α.

The corresponding grid-current curves are also shown by dash lines. The product $u_g u_p$ is equal to

$$u_g u_p = \left(\frac{\partial e_g}{\partial e_p}\right)_{i_g} \cdot \left(\frac{\partial e_p}{\partial e_g}\right)_{i_p} \tag{248}$$

Using finite differences and making the increment in plate voltage the same for both partial derivatives and equal to $\Delta\bar{E}_p$, Eq. (248) takes the approximate form

$$\left.\begin{aligned} u_g u_p &= \left(\frac{\Delta e_g''}{\Delta\bar{E}_p}\right)_{i_g} \cdot \left(\frac{\Delta E_p}{\Delta e_g'}\right)_{i_p} \\ &= \frac{\Delta e_g''}{\Delta e_g'} \end{aligned}\right\} \tag{249}$$

The value of α as calculated from the measured values of u_g and u_p for a certain hard tube is plotted against e_g in Fig. 139.

In the graphical solution shown in Fig. 108, page 212, the slope of the path line uv at any point is the value of y_g at that point,

y_g being in this case a real quantity, since the plate-circuit impedance is a pure resistance.

The complex expression for $\mathbf{z}_g$, as given in Eq. (244), can be separated into a real part, the equivalent resistance, and an imaginary part, the equivalent reactance. The steps are not shown but the final result is

$$\mathbf{z}_g = r_g \cdot \frac{r_p^2 + (1+\alpha) r_p \tilde{R}_b + \alpha Z_b^2}{r_p^2 + 2\alpha r_p \tilde{R}_b + \alpha^2 Z_b^2} - j r_g \cdot \frac{(\alpha - 1) r_p X_b}{r_p^2 + 2\alpha r_p \tilde{R}_b + Z_b^2} \tag{250}$$

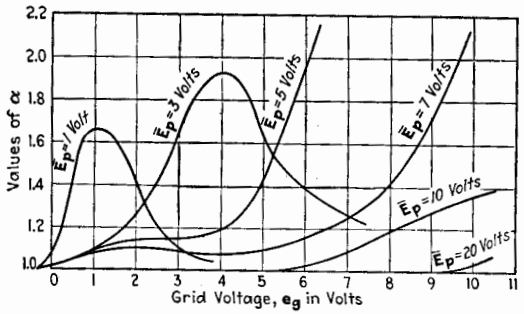


FIG. 139.—Values of α for a high-vacuum triode (UV201).

From Eq. (250) we can express as follows the equivalent grid-circuit resistance $[r_g]$ and reactance $[x_g]$, where $\mathbf{z}_g = [r_g] + j[x_g]$.

$$[r_g] = r_g \cdot \frac{r_p^2 + (1+\alpha) r_p \tilde{R}_b + \alpha Z_b^2}{r_p^2 + 2\alpha r_p \tilde{R}_b + \alpha^2 Z_b^2} \tag{251}$$

$$[x_g] = -r_g \cdot \frac{(\alpha - 1) r_p X_b}{r_p^2 + 2\alpha r_p \tilde{R}_b + Z_b^2} \tag{252}$$

The equivalent admittance $\mathbf{y}_g$ can be similarly expanded to give the equivalent grid-circuit conductance g_g and susceptance b_g, where $\mathbf{y}_g = g_g - jb_g$.

$$g_g = k_g \cdot \frac{r_p^2 + (1+\alpha) r_p \tilde{R}_b + \alpha Z_b^2}{r_p^2 + 2 r_p \tilde{R}_b + Z_b^2} \tag{253}$$

$$b_g = -k_g \cdot \frac{(\alpha - 1) r_p X_b}{r_p^2 + 2 r_p \tilde{R}_b + Z_b^2} \tag{254}$$

Since the grid circuit is more often a series than a parallel circuit, the equivalent impedance is more useful than the admittance. Examine somewhat more in detail the expressions of Eqs. (251) and (253).

Special Case 1. $X_b = 0$. *Low Frequency.*—Assume first that the plate-circuit reactance is zero. This condition corresponds to a pure resistance in the plate circuit and is very nearly true for a resonant tuned circuit, a very common type of plate-circuit load. Equations (251) and (252) for $X_b = 0$ reduce to

$$\frac{[r_g]}{r_g} = \frac{1 + \frac{\bar{R}_v}{r_p}}{1 + \alpha\frac{\bar{R}_b}{r_p}} \tag{255}$$

and

$$[x_g] = 0 \tag{256}$$

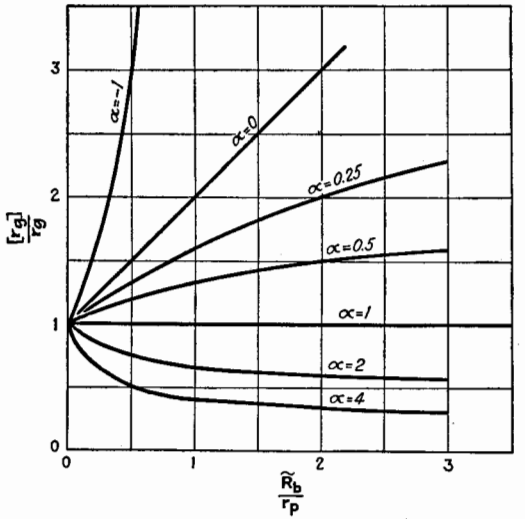


FIG. 140.—Values of $[r_g]/r_g$ for $X_b = 0$ and negligible tube capacitances.

Figure 140 is a graphical representation of Eq. (255). If α is negative, the equivalent resistance can become negative by passing first through infinity. The plot of g_g shows somewhat better than does that of r_g the change from a positive to a negative grid loss. For $X_b = 0$, expression (253) reduces to

$$\frac{g_g}{k_g} = \frac{1 + \alpha\frac{R_b}{r_p}}{1 + \frac{\bar{R}_b}{r_p}} \tag{257}$$

Equation (257) is shown plotted in Fig. 141. Whenever $[r_g]$ or g_g is negative, the triode system, instead of absorbing power,

supplies power to the grid circuit. A resonant circuit in the grid circuit of a triode oscillates if the input conductance is sufficiently negative to cancel the positive conductance of the circuit. Such oscillations are often observed when a soft tube is used.

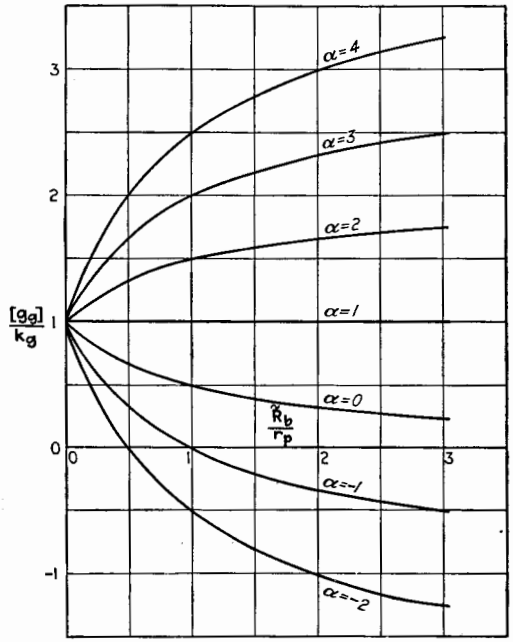


FIG. 141.—Values of $[g_g]/k_g$ for $X_b = 0$ and negligible tube capacitances.

Special Case 2. $\tilde{R}_b = 0$. *Low Frequency.*—Consider now the case when the external plate-circuit impedance is a pure reactance with negligible resistance. Equation (251) reduces to

$$\frac{[r_g]}{r_g} = \frac{1 + \alpha\dfrac{X_b^2}{r_p^2}}{1 + \alpha^2\dfrac{X_b^2}{r_p^2}} \tag{258}$$

and

$$\frac{[x_g]}{r_g} = -\frac{(\alpha - 1)\dfrac{X_b}{r_p}}{1 + \alpha^2\dfrac{X_b^2}{r_p^2}} \tag{259}$$

Equation (258) is plotted in Fig. 142 and should be compared with Fig. 140. Since the most usual values of α lie between 1 and 1.5,

the equivalent input resistance $[r_g]$ is in this case less than r_g. Negative values of α, however, give negative input resistances over certain ranges of X_b/r_p, as shown by Fig. 142.

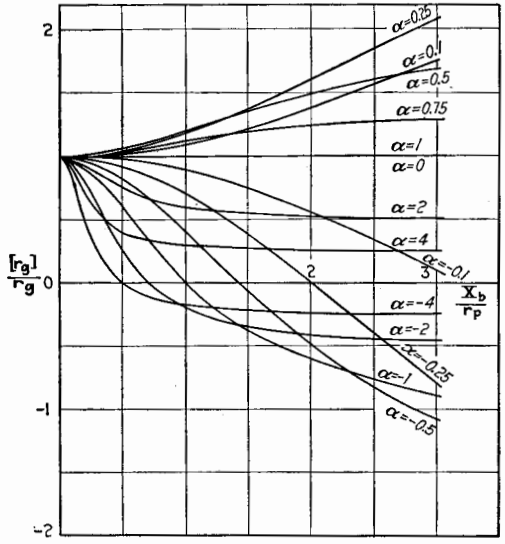


FIG. 142.—Values of $[r_g]/r_g$ for $R_b = 0$ and negligible tube capacitances.

Figure 143 is a plot of Eq. (259) for the equivalent input reactance. If X_b is positive, *i.e.*, inductive, the equivalent input reactance for values of α greater than 1 is negative or capacitive. If X_b is negative, the sign of all values of $[x_g]$ plotted in Fig. 143 must be reversed.

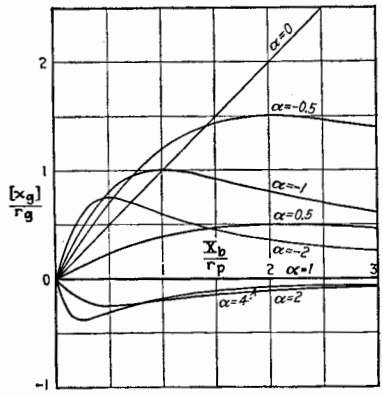


FIG. 143.—Values of $[x_g]/r_g$ for $R_b = 0$ and negligible tube capacitances.

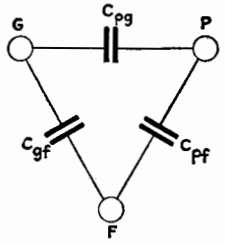


FIG. 144.—Interelectrode tube capacitances.

113. Tube Capacitances.—Before taking up the general solution of the input admittance, which depends not only on the grid

conductance but also on the interelectrode capacitances, a digression will be made to discuss tube capacitances.

A triode contains three separated electrodes, between each pair of which there is a capacitance. There are consequently three capacitances in a triode, as indicated in Fig. 144. The capacitance between grid and filament is denoted by C_{gf}, between plate and filament by C_{pf}, and between plate and grid by C_{pg}, the order of the subscript letters having no significance.

The values of the interelectrode capacitances for several standard tubes are given in Table XIII.

When the triode is fixed in a socket and connected in a system, the effective capacitances C_{pf}, C_{gf}, and C_{pg} are all increased by the capacitances within the socket and between the connecting wires. Consequently, the effective values of C_{pf}, C_{gf}, and C_{pg} are then greater than the values of the tube capacitances alone by from 2 or 3 to 10 or 15 $\mu\mu$f or more. However, the same symbols, C_{pf}, C_{gf}, and C_{pg}, are used to denote the effective tube capacitances whether or not the effects of the socket and wires are included.

TABLE XIII.—INTERELECTRODE CAPACITANCES OF COMMERCIAL TRIODES

Tube	Capacitance, micromicrofarads		
	C_{pf}	C_{gf}	C_{pg}
WD 11 amplifier and detector	2.5	2.5	3.3
WX 12 amplifier and detector	2.5	2.5	3.3
UX 112A amplifier and detector	2.1	4.2	8.1
UX 120 power amplifier	2.3	2.0	4.1
UX 171A power amplifier	2.1	3.7	7.4
UX 199 detector and amplifier	2.5	2.5	3.3
UX 200A detector	2.0	3.2	8.5
UX 201A detector and amplifier	2.2	3.1	8.1
UX 210 power amplifier	4.0	5.0	8.0
UX 226 amplifier	2.2	3.5	8.1
UY 227 detector and amplifier	3.0	3.5	3.3
RC A230 detector and amplifier	2.1	3.7	6.4
RC A231 power amplifier	2.4	3.5	5.9
RC A237 detector and amplifier, d-c. heater	2.5	3.3	2.1
UX 240 amplifier	1.5	3.4	8.8
UX 245 power amplifier	3.0	5.0	8.0
UX 250 power amplifier	3.0	5.0	9.0
RCA 852 power amplifier	1.0	2.0	3.0

114. Measurement of Tube Capacitances.—One method of measuring at an audio frequency the capacitances between the electrodes of a triode, using an ordinary capacitance bridge, is as follows: If grid and plate are connected and the capacitance between filament and plate is measured, the result is $C_{pf} + C_{gf}$. Again, if the plate and filament are connected and the capacitance between grid and filament is measured, the result is $C_{gf} + C_{pg}$. Finally, by measuring the capacitance between plate and filament with grid and filament connected, $C_{pf} + C_{pg}$ is obtained. From these three results the separate capacitances can be calculated.

One form of capacitance bridge suitable for the measurement of the capacitances as described is shown in Fig. 145. Usually the most accurate and convenient procedure is, first, to balance the bridge of Fig. 145 with a small gap, shown at a, in the connection to the tube. Then the gap is closed, and C_4 and R_3 are adjusted to restore balance. The capacitance, being measured, say $C_{pf} + C_{pg}$, is then equal to the change in capacitance of C_4. A two-stage amplifier is usually necessary to secure sufficient sensitivity. The resistances R_1 and R_2 should preferably be equal and of the order of 1,000 or 10,000 ohms. If the capacitance of the tube alone is desired, the change in C_4 may be obtained when the tube is inserted in the socket, the socket being permanently connected across C_4.

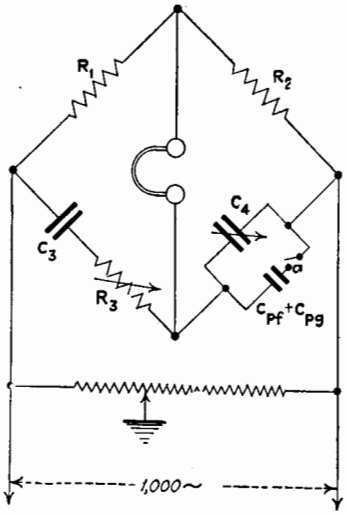


FIG. 145.—Capacitance bridge.

Very often there is a high-resistance path through a film of sputtered metal or of "getter" deposited on the "squash" of the tubes between the wires leading to the electrodes, or in some cases there is a high dielectric loss in the glass and other insulating substances. When such effects exist, the capacitance being measured is equivalent to a pure capacitance shunted by a resistance. This resistance causes an error in the measured capacitance obtained from the bridge of Fig. 145, as is evident from the conditions of balance for the bridge when R_4 is included. The conditions of balance are

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} + \frac{C_4}{C_3} \tag{260}$$

$$C_4\omega R_4 = \frac{1}{C_3\omega R_3} \tag{261}$$

where R_4 is the shunt resistance across capacitance C_4, the latter now representing the total capacitance in the arm. If R_4 is neglected, the fractional error in the measured value of a small capacitance, denoted by ΔC_4, is

$$\text{Fractional error} = \frac{C_4}{\Delta C_4} \cdot \frac{1}{R_4^2 C_4^2 \omega^2} \tag{262}$$

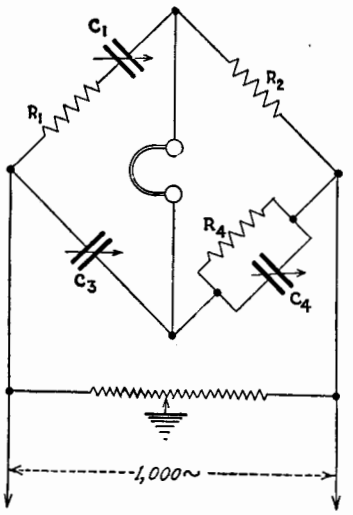


FIG. 146.—Capacitance bridge for the measurement of the capacitance of a leaky condenser.

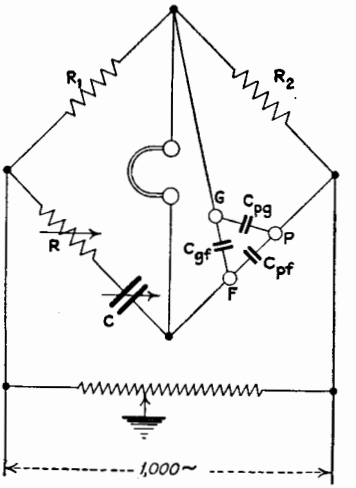


FIG. 147.—Capacitance bridge for directly measuring tube capacitances.

If ΔC_4 is $10\mu\mu$f, C_4 is $200\mu\mu$f, R_4 is 10 megohms, and ω is $2\pi 1{,}000$, this error amounts to 12.6 per cent. Evidently the error is reduced by increasing the frequency used on the bridge and by increasing C_4.

The arrangement of the bridge shown in Fig. 146 eliminates this error due to the resistance R_4. The conditions of balance of this bridge are

$$\frac{R_1}{R_2} = \frac{C_4}{C_3} \tag{263}$$

$$\frac{R_4}{R_2} = \frac{C_1}{C_3} \tag{264}$$

In order that C_1 be not inconveniently large, C_3 and C_4 should be small, of the order of 100$\mu\mu$f, and R_1 and R_2 large. The bridge is first balanced by adjusting C_3, when R_4 and the added capacitance are out of circuit. When ΔC_4 and R_4 are added, the bridge balance is restored by resetting C_4 and C_1. The added capacitance ΔC_4 is given by the change in C_4, the resistance R_4 being balanced by changing C_1. It may be of advantage to add across C_4 a high resistance, such as a grid leak, in order that C_1 may be of convenient size. The arrangement of Fig. 146 can be used to determine the tube capacitances when the filament is operating.

A method of directly measuring the tube capacitance is shown in Fig. 147. With the connections shown, C_{gf} has no effect upon the conditions of balance because it is across the telephone receivers. The capacitance C_{pg} is balanced by resistance R, as shown by the conditions of balance given for the connections of Fig. 147.

$$\frac{R_1}{R_2} = \frac{C_{pf}}{C} \tag{265}$$

$$\frac{R_1}{R} = \frac{C_{pf}}{C_{pg}} \tag{266}$$

Here again, if there is conduction or its equivalent between the terminals of the condenser C_{pf}, this conduction being determined by R_{pf}, the conditions of balance become

$$\frac{R_1}{R_2} = \frac{C_{pf}}{C} + \frac{R}{R_{pf}} \tag{267}$$

$$R_1 C_{pg}\omega = RC_{pf}\omega - \frac{1}{R_{pf}C\omega} \tag{268}$$

The same sort of error as in the bridge of Fig. 145 enters here, owing to the conduction indicated by R_{pf}. This fractional error, when C_{pf} is measured by the method shown in Fig. 147 and calculated by Eq. (265), is

$$\text{Fractional error} = \frac{1}{R_{pf}^2 C_{pf}^2 \omega^2} + \frac{R_1 C_{pg} C}{R_{pf} C_{pf}^2} \tag{269}$$

If R_{pf} is 100 megohms, C_{pf} is 10$\mu\mu$f, C_{pg} is 10$\mu\mu$f, R_1 is 10 ohms, C is 1,000$\mu\mu$f, and ω is 2π1,000, this error amounts to 2.5 per cent. If R_{pf} is as low as 10 megohms, the result is over 100 per cent in error. If there is any appreciable conduction in the triode, caution is necessary in the use of this bridge for accurate work.

Methods of measuring the tube capacitances at radio frequencies are less affected by conduction within the triode but are usually far less convenient in manipulation. Several methods may be found in handbooks. One of these methods is shown in Fig. 148. An oscillatory circuit, consisting of a relatively small inductance L and a large capacitance C, is enclosed in a shielding box S. A thermocouple T is connected in the oscillatory circuit as shown and to a suitable galvanometer or microammeter. A small standard variable condenser C_s, enclosed within a shield, is graduated in micromicrofarads. A socket, the terminals of which are indicated by letters G, P, and F, is con-

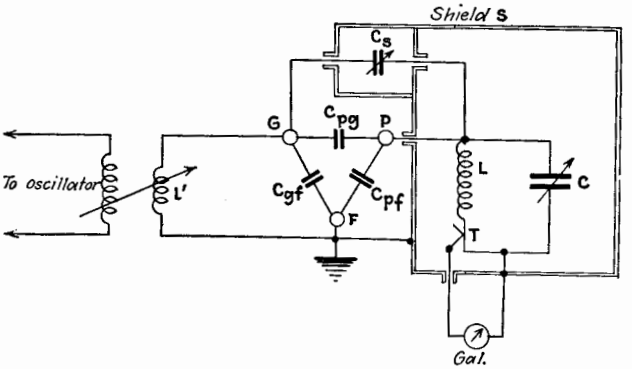


FIG. 148.—Method of measuring tube capacitances at a radio frequency.

nected as shown. The small coil L' couples the circuit loosely to a radio-frequency oscillator. With the tube in the socket and C_s set at a low value, C is tuned to resonance as indicated by a maximum deflection of the galvanometer. The tube is then withdrawn from the socket and C_s is adjusted to give the same galvanometer reading after C has been readjusted to compensate for the removal of C_{pf}. The difference in the values of C_s gives the value of C_{pg}. The other triode capacitances can be measured in the same manner.

115. Total Input Admittance of a Triode.[1]—The complete expression for the input admittance, as affected both by the grid-conduction current and by the tube capacitances, will now be derived. The final result is simpler and more convenient

[1] The equivalent input impedance of a triode, neglecting the effects of $u_g \Delta E_p$, was derived by J. M. Miller in 1919 (see *Bur. Standards Bull.* 351); by Stuart Ballantine, 1920 (see *Phys. Rev.*, **15**, 409); by H. W. Nichols in 1919 (see *Phys. Rev.*, **13**, 405).

to use when expressed as admittance rather than as impedance. The tube and its associated circuits are represented in Fig. 149. G, P, and F represent the grid, plate, and filament terminals of the tube. C_{gf}, C_{pf}, and C_{pg} are the grid-to-filament, plate-to-filament, and plate-to-grid tube capacitances. The conductive path from plate to filament within the tube is replaced by the fictitious e.m.f. $u_p\Delta E_g$ in series with k_p. Similarly, the grid-to-filament path consists of the fictitious e.m.f. $u_g\Delta E_p$ in series with the conductance k_g. $\mathbf{Y}_b = G_b - jB_b$ is the admittance of the external plate circuit, and $\mathbf{Y}_c = G_c - jB_c$ is the admittance of the external grid circuit. The currents and potentials are indicated on the circuit diagram.

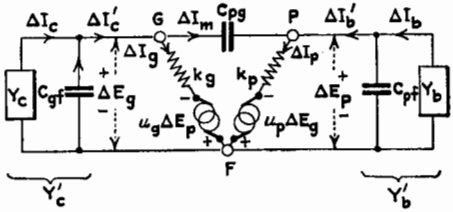


FIG. 149.—Complete equivalent circuits of a triode.

For simplicity in derivation, $\mathbf{Y}'_b$ is defined by the equation $\mathbf{Y}'_b = \mathbf{Y}_b + jC_{pf}\omega$, and $\mathbf{Y}'_c$ by the equation $\mathbf{Y}'_c = \mathbf{Y}_c + jC_{gf}\omega$.

The actual input admittance of the triode is

$$\mathbf{y}_g = \frac{\Delta\mathbf{I}_c}{\Delta\mathbf{E}_g} = \frac{\Delta\mathbf{I}'_c}{\Delta\mathbf{E}_g} + jC_{gf}\omega \tag{270}$$

The admittance $\Delta\mathbf{I}'_c/\Delta\mathbf{E}_g$ is obtained by solving the following equations for the network of Fig. 149:

$$\left.\begin{aligned} -\Delta\mathbf{I}_m + \Delta\mathbf{I}'_c - \Delta\mathbf{I}_g &= 0 \\ \Delta\mathbf{I}_m + \Delta\mathbf{I}'_b - \Delta\mathbf{I}_p &= 0 \\ -s_g\Delta\mathbf{E}_p + \Delta\mathbf{I}_g &= k_g\Delta\mathbf{E}_g \\ -k_p\Delta\mathbf{E}_p + \Delta\mathbf{I}_p &= s_p\Delta\mathbf{E}_g \\ \Delta\mathbf{I}_m + jC_{pg}\omega\Delta\mathbf{E}_p &= jC_{pg}\omega\Delta\mathbf{E}_g \\ \Delta\mathbf{I}'_b + \mathbf{Y}'_b\Delta\mathbf{E}_p &= 0 \end{aligned}\right\} \tag{271}$$

Attention is called to the fact that the admittance between grid and plate is here assumed to consist of a capacitance only. If conductance also exists between these terminals, or, in general,

if the admittance is $\mathbf{Y}_{pg}$, then $\mathbf{Y}_{pg}$ should be substituted for $jC_{pg}\omega$ in the fifth equation in group (271).

The input admittance, derived from the relations (271), is

$$\mathbf{y}_g = \left\{k_g + \frac{(C_{pg}^2\omega^2 - s_p s_g)(k_p + G_b) + C_{pg}\omega(s_p + s_g)[(C_{pg} + C_{pf})\omega - B_b]}{[k_p + G_b]^2 + [(C_{pg} + C_{pf})\omega - B_b]^2}\right\}$$
$$+ j\left\{(C_{gf} + C_{pg})\omega + \frac{C_{pg}\omega(s_p + s_g)(k_p + G_b) - (C_{pg}^2\omega^2 - s_p s_g)[(C_{pg} + C_{pf})\omega - B_b]}{[k_p + G_b]^2 + [(C_{pg} + C_{pf})\omega - B_b]^2}\right\} \tag{272}$$

If C_{pg}, C_{pf}, and C_{gf} are made zero, the two braces of Eq. (272) become Eqs. (253) and (254) of Sec. 112.

116. Total Internal Output Admittance of a Triode.—Sometimes, in order to balance telephone networks fed from the plate circuit of a triode, it is necessary to know the admittance of the triode looking back into it from the plate circuit. This is called the *internal output admittance.* This quantity is useful in other problems, as well as in the one cited.

Because of the symmetry of the network of Fig. 149, the equivalent internal output admittance $\mathbf{y}_p = \Delta\mathbf{I}_b/\Delta\mathbf{E}_p$ can be obtained from Eq. (272) by interchanging all subscripts p and g and all subscripts b and c.

The resulting expression for the internal output admittance is

$$\mathbf{y}_p = \left\{k_p + \frac{(C_{pg}^2\omega^2 - s_p s_g)(k_g + G_c) + C_{pg}\omega(s_p + s_g)[(C_{pg} + C_{gf})\omega - B_c]}{[k_g + G_c]^2 + [(C_{pg} + C_{gf})\omega - B_c]^2}\right\}$$
$$+ j\left\{(C_{pf} + C_{pg})\omega + \frac{C_{pg}\omega(s_p + s_g)(k_g + G_c) - (C_{pg}^2\omega^2 - s_p s_g)[(C_{pg} + C_{gf})\omega - B_c]}{[k_g + G_c]^2 + [(C_{pg} + C_{gf})\omega - B_c]^2}\right\} \tag{273}$$

117. Special Examples of Input and Output Admittances of a Triode.—The magnitudes of the components $\mathbf{y}_g$ and $\mathbf{y}_p$ depend upon the four tube coefficients k_p, k_g, s_p, and s_g, all of which are functions of the plate and grid potentials. The expressions are

so complicated that little information is conveyed by mere examination of the complete equations. Some special cases will be considered to give a general idea of the way in which the admittances vary with frequency and with variations in the character of the plate and grid external circuits.

Case 1. *Input Admittance for* $k_g = 0$, $B_o = 0$.—This case applies to an amplifier whose grid is polarized negatively and whose plate load is a pure resistance. Equation (272), giving the equivalent input admittance, reduces to the following form on putting $k_g = 0$, and $B_b = 0$.

$$\left.\begin{array}{l} k_g = 0 \\ B_b = 0 \end{array}\right| \quad \frac{g_g}{k_p} = A^2 \frac{\left(1 + \frac{1}{x_b}\right) + u_p\left(1 + \frac{C_{pf}}{C_{pg}}\right)}{\left(1 + \frac{1}{x_b}\right)^2 + A^2\left(1 + \frac{C_{pf}}{C_{pg}}\right)^2} \tag{274}$$

$$\left.\begin{array}{l} k_g = 0 \\ B_b = 0 \end{array}\right| \quad \frac{b_g}{k_p} = -A\left[1 + \frac{C_{gf}}{C_{pg}} + \frac{u_p\left(1 + \frac{1}{x_b}\right) - A^2\left(1 + \frac{C_{pf}}{C_{pg}}\right)}{\left(1 + \frac{1}{x_b}\right)^2 + A^2\left(1 + \frac{C_{pf}}{C_{pg}}\right)^2}\right] \tag{275}$$

where $y_g = g_g - jb_g$; $x_b = R_b/r_p$; and $A = C_{pg}\omega/k_p$.

The expressions for the input and output admittances are much simpler of interpretation when expressed in ratio form, as above. By dividing through by k_p, as in Eqs. (274) and (275), the ratios of the equivalent conductance and equivalent susceptance to k_p are functions of the simple ratios R_b/r_p, $C_{pg}\omega/k_p$, u_p, C_{pf}/C_{pg}, and C_{gf}/C_{pg}. The last two ratios are constants, and u_p is nearly constant for any given tube.

In Fig. 150, g_g/k_p and b_g/k_p are plotted against $R_b/r_p = x_b$ as abscissas for various values of the parameter A. The values of the ratios u_p, C_{pf}/C_{pg}, and C_{gf}/C_{pg}, used in the calculations correspond approximately to a standard type of triode. The numerical values used for the voltage ratio and the capacitances are as follows: u_p is assumed to be 8, and the values of C_{pg}, C_{pf} and C_{gf}, including the capacitances of the socket, are taken as 10, 5, and 5$\mu\mu$f. The parameter A has the value unity if r_p is 16,000 ohms, C_{pg} is 10$\mu\mu$f, and the frequency is 10^6 cycles per second.

Examination of Fig. 150 shows that, when the plate load of an amplifier is a pure resistance or its equivalent, the triode acts

upon the input circuit as if there were connected from grid to filament a resistance and a capacitive reactance in parallel. Both the resistance and the capacitive reactance may be much less than r_p, especially if the frequency is high.

*Case 2. *Internal Output Admittance for $k_g = 0$ and $B_c = 0$.*— This case applies to an amplifier whose grid is polarized negatively and whose grid-circuit load is a pure resistance. Under these conditions Eq. (273) reduces to

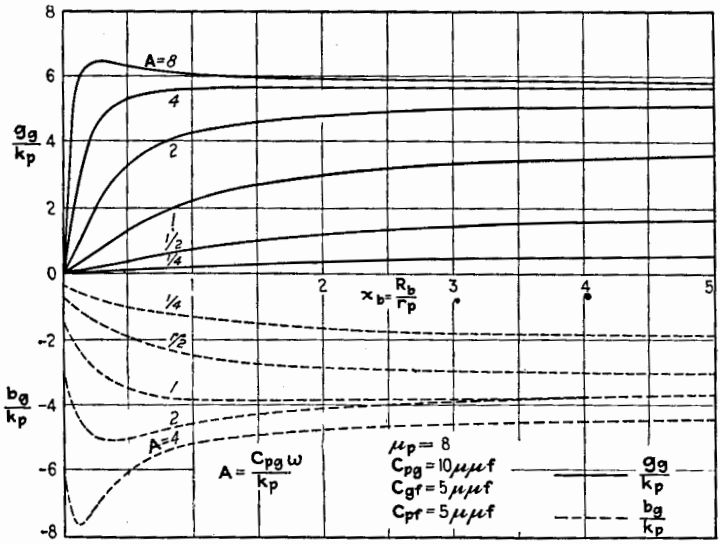


FIG. 150.—Components of the input admittance of a triode when $k_g = 0$ and $B_b = 0$.

$$k_g = 0,\quad B_c = 0 \begin{cases} \dfrac{g_p}{k_p} = 1 + A^2 \dfrac{\dfrac{1}{x_c} + u_p\left(1 + \dfrac{C_{gf}}{C_{pg}}\right)}{\left(\dfrac{1}{x_c}\right)^2 + A^2\left(1 + \dfrac{C_{gf}}{C_{pg}}\right)^2} & (276) \\ \dfrac{b_p}{k_p} = -A\left[1 + \dfrac{C_{pf}}{C_{pg}} + \dfrac{\dfrac{u_p}{x_c} - A^2\left(1 + \dfrac{C_{gf}}{C_{pg}}\right)}{\left(\dfrac{1}{x_c}\right)^2 + A^2\left(1 + \dfrac{C_{gf}}{C_{pg}}\right)^2}\right] & (277) \end{cases}$$

where $\mathbf{y}_p = g_p - jb_p$, $x_c = R_c/r_p$, and $A = C_{pg}\omega/k_p$.

Figure 151 illustrates the shape of the curves obtained by plotting Eqs. (276) and (277) against R_c/r_p as abscissas for a tube having the same constants as in Case 1.

Looking back into the triode from the plate load, the triode, with its associated grid load of a pure resistance, is equivalent to a resistance and capacitive reactance in parallel. Both resistance and reactance may be much less than r_p, as shown by the curves of Fig. 151.

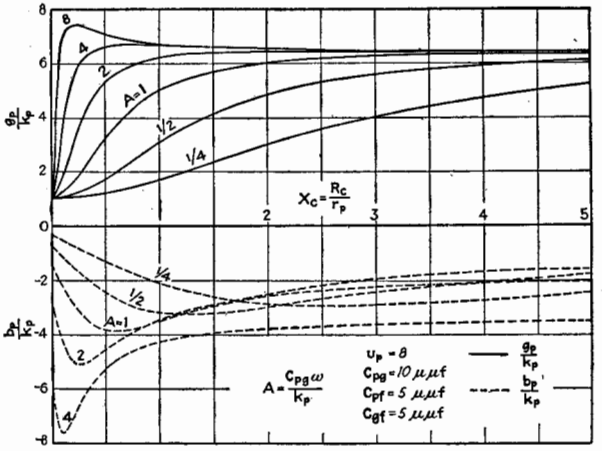


FIG. 151.—Components of the internal output admittance of a triode when $k_g = 0$ and $B_c = 0$.

Case 3. *Input Admittance for Reactive Plate Load, and When* $k_g = 0$.—This case applies to an amplifier whose grid is polarized negatively and whose plate load has both resistance and reactance. Another parameter η_b is now introduced, its absolute value giving the ratio of the conductance to the susceptance of the load; *i.e.*, $\eta_b = G_b/|B_b|$. Equation (272) is simplified for this case by substituting for G_b the quantity $\eta_b|B_b|$ and by putting $k_g = 0$. The result is

$$k_g = 0 \qquad \eta_b = \frac{G_b}{|B_b|}$$

$$\frac{g_g}{k_p} = A^2 \frac{1 + \frac{\eta_b}{|x_b'|} + u_p\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{Ax_b'}\right)}{\left(1 + \frac{\eta_b}{|x_b'|}\right)^2 + A^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{Ax_b'}\right)^2} \tag{278}$$

$$\frac{b_g}{k_p} = -A\left[1 + \frac{C_{gf}}{C_{pg}} + \frac{u_p\left(1 + \frac{\eta_b}{|x_b'|}\right) - A^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{Ax_b'}\right)}{\left(1 + \frac{\eta_b}{|x_b'|}\right)^2 + A^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{Ax_b'}\right)^2}\right] \tag{279}$$

where $y_g = g_g - jb_g$; $x_b' = k_p/B_b$, and $A = C_{pg}\omega/k_p$. The coordinate x_b is positive if the load is inductive and negative if the load is capacitive.

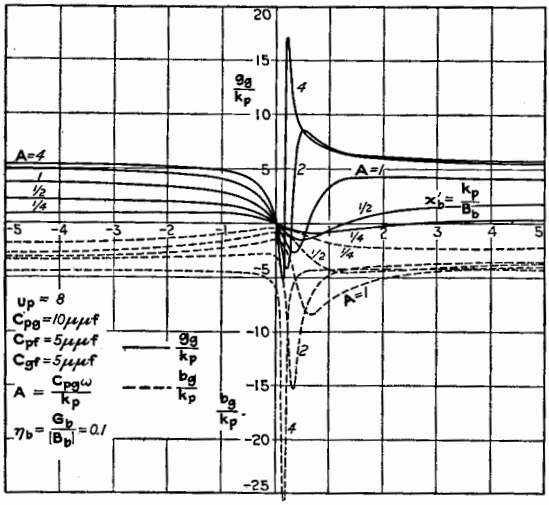


FIG. 152.—Components of the input admittance of a triode when $k_g = 0$. $\eta_b = G_b/[B_b]$ for various values of A.

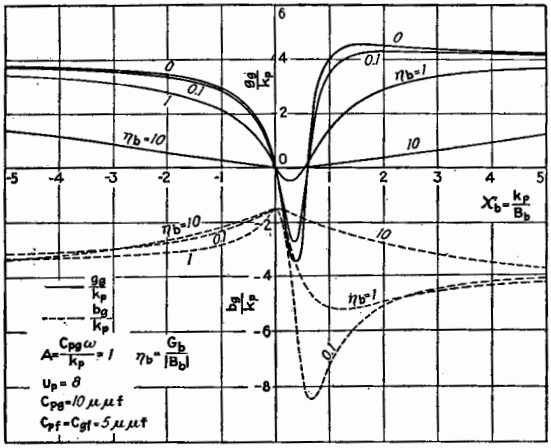


FIG. 153.—Components of the input admittance of a triode for various values of η_b, when $k_g = 0$. $\eta_b = G_b/[B_b]$.

In Fig. 152, Eqs. (278) and (279) are plotted for η_b equal to 0.1 and for several values of A. The values of u_p, C_{pf}, C_{pg}, and C_{gf} are the same as in Case 1, Fig. 150.

In Fig. 153, Eqs. (278) and (279) are plotted for $A = 1$ and for several values of η_b. The constants of the tube are the same as in Fig. 152.

Referring now to Figs. 152 and 153, it is seen that the susceptance of the input admittance is always capacitive, no matter what the type of plate load is. The input conductance is positive for all plate loads which have a capacitive susceptance, but the input conductance is negative for a plate load having certain ranges of inductive susceptance.

As far as effects in the grid circuit are concerned, the triode and its plate load may be replaced by its equivalent input admittance $\mathbf{y}_g$ in parallel with $\mathbf{Y}_c$. The system oscillates if $\mathbf{Y}_c + \mathbf{y}_g = 0$. Oscillation occurs when the grid circuit contains an inductance and capacitance in parallel, if

$$\frac{R_c}{R_c^2 + L_c^2\omega^2} + g_g \leqq 0$$

and

$$C_c\omega - \frac{L_c\omega}{R_c^2 + L_c^2\omega^2} - b_g = 0$$

**Case* 4. *Internal Output Admittance for Reactive Grid Circuit, and When* $k_g = 0$. This case applies to the internal output admittance of an amplifying triode whose grid is polarized negatively and whose grid-circuit load contains both resistance and reactance, the latter being either inductive or capacitive. As in Case 3, a new parameter $\eta_c = G_c/|B_c|$ is introduced. Formula (273) reduces to

$$k_g = 0 \begin{cases} \dfrac{g_p}{k_p} = 1 + A^2 \cdot \dfrac{\dfrac{\eta_c}{|x_c'|} + u_p\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{Ax_c'}\right)}{\left(\dfrac{\eta_c}{x_c'}\right)^2 + A^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{Ax_c'}\right)^2} & (280) \\ \dfrac{b_p}{k_p} = -A\left[1 + \dfrac{C_{pf}}{C_{pg}} + \dfrac{\dfrac{u_p\eta_c}{|x_c'|} - A^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{Ax_c'}\right)}{\left(\dfrac{\eta_c}{x_c'}\right)^2 + A^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{Ax_c'}\right)^2}\right] & (281) \end{cases}$$

where $\mathbf{y}_p = g_p - jb_p$, $x_c' = k_p/B_c$, and $A = C_{pg}\omega/k_p$.

In Fig. 154, Eqs. (280) and (281) are plotted for η_c equal to 0.1 and for several values of A. The values of u_p, C_{pf}, C_{pg}, and C_{gf} are the same as in Case 1.

In Fig. 155, Eqs. (280) and (281) are plotted for A equal to 1 and for several values of η_c.

The results in this case are analogous to those in Case 3, in that the equivalent output susceptance, like the input susceptance, is always capacitive. Further, the output conductance is positive if the grid circuit has a capacitive susceptance, but

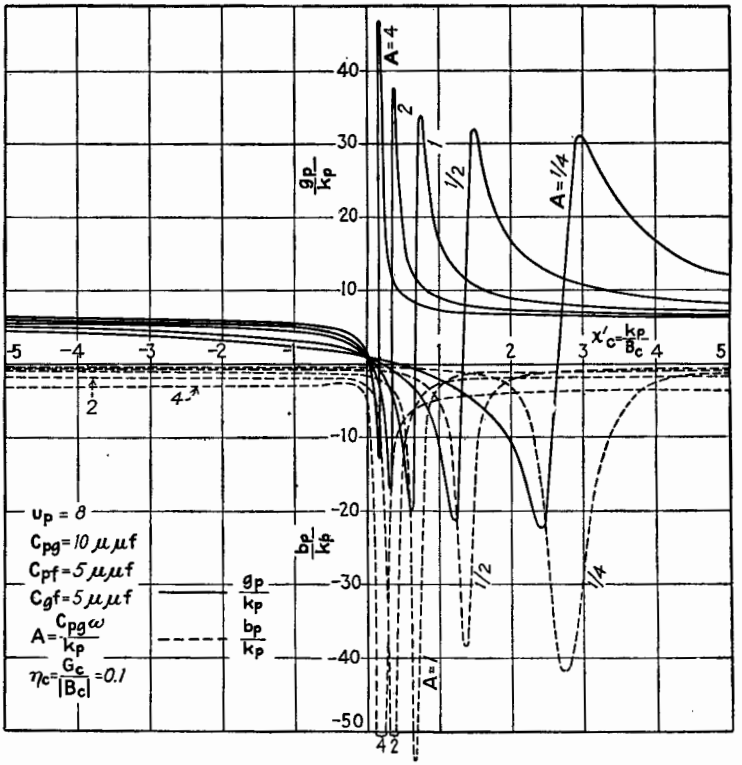


FIG. 154.—Components of the internal output admittance of a triode for various values of A when $k_g = 0$.

may be negative if the grid circuit is inductive. The conditions of oscillation are expressed by the relation $\mathbf{Y}_b + \mathbf{y}_p = 0$ or

$$\frac{R_b}{R_b^2 + L_b^2\omega^2} + g_p \leqq 0$$

and

$$C_b\omega - \frac{L_b\omega}{R_b^2 + L_b^2\omega^2} - b_p = 0$$

118. Extension of the Equivalent-plate-circuit Theorem.—The e-p-c. theorem, presented in Chap. VIII, was derived with no regard to the interelectrode capacitances. Consideration of these capacitances does not in any way affect the validity of the theorem as expressed in Eq. (159) (also given in the fourth equation of group (271)), provided ΔI_p is considered as the current through r_p or k_p only, as shown in Fig. 149. The capacitances

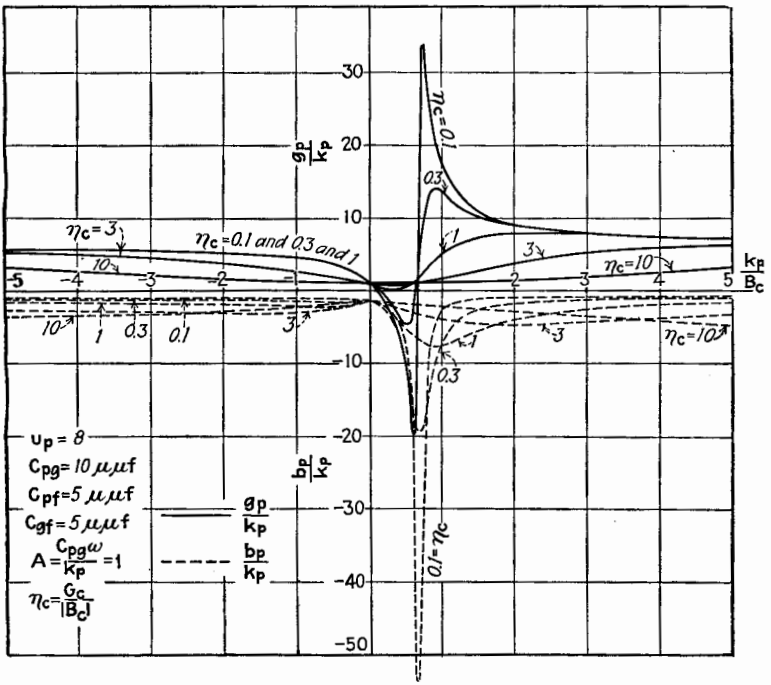


FIG. 155.—Components of the internal output admittance of a triode for various values of η_c when $k_g = 0$. $\eta_c = G_c/[B_c]$.

C_{pg}, C_{gf}, and C_{pf} merely add extra current paths to the simple network of Fig. 94, page 199. In the calculation of the current ΔI_b through the load, these paths can readily be taken into account by solution of the network equations for Fig. 149, using the form of the equivalent-circuit theorem given in Chap. VIII. Since, however, the network is practically always of the form shown in Fig. 149, the problem may be solved once for all. The current ΔI_b or $\Delta I_b'$ can be expressed in terms of ΔE_g and the constants of the circuit, in the form of an extension of the e-p-c. theorem.

The procedure to obtain the new form of the e-p-c. theorem is to solve the equations of group (271) for ΔI_b in terms of ΔE_g. The result is

$$\mathbf{\Delta I}_b = \frac{(s_p - jC_{pg}\omega)\mathbf{Y}_b}{k_p + \mathbf{Y}_b + j(C_{pg} + C_{pf})\omega} \cdot \mathbf{\Delta E}_g \tag{282-1}$$

$$= \frac{(u_p - jA)\mathbf{Y}_b}{1 + \dfrac{\mathbf{Y}_b}{k_p} + jA\left(1 + \dfrac{C_{pf}}{C_{pg}}\right)} \cdot \mathbf{\Delta E}_g \tag{282-2}$$

$$= \frac{(u_p - jr_pC_{pg}\omega)}{r_p + \mathbf{Z}_b + jr_p\mathbf{Z}_b(C_{pg} + C_{pf})\omega} \cdot \mathbf{\Delta E}_g \tag{282-3}$$

An equivalent network which is consistent with Eq. (282-3) is given in the right-hand half of Fig. 156.

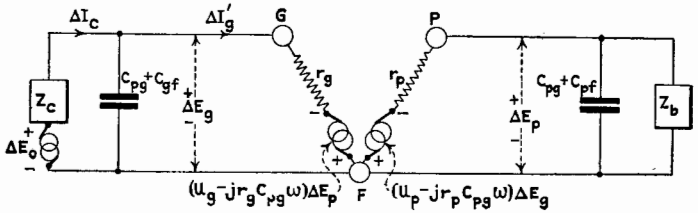


FIG. 156.—Equivalent circuits of a triode including effects of currents through tube capacitances.

In a similar manner, the current ΔI_c of Fig. 149 can be found and is

$$\mathbf{\Delta I}_c = \frac{[1 + jr_g(C_{pg} + C_{gf})\omega]\mathbf{\Delta E}_0 + [u_g - jr_gC_{pg}\omega]\mathbf{\Delta E}_p}{r_g + \mathbf{Z}_c + jr_g\mathbf{Z}_c(C_{pg} + C_{gf})\omega} \tag{283}$$

The equivalent network for the grid circuit shown in the left-hand half of Fig. 156 is consistent with this equation.

Another equivalent network, similar to that shown in Fig. 96, page 200, for the simple e-p-c. theorem, can be drawn for the case under consideration. Expressing ΔE_b in terms of ΔI_b, Eq. (282-1) becomes

$$(s_p - jC_{pg}\omega)\mathbf{\Delta E}_g = [k_p + \mathbf{Y}_b + j(C_{pg} + C_{pf})\omega]\mathbf{\Delta E}_b \tag{284}$$

Dividing by k_p,

$$(u_p - jr_pC_{pg}\omega)\mathbf{\Delta E}_g = \left[1 + \frac{\mathbf{Y}_b}{k_p} + jr_pC_{pg}\omega\left(1 + \frac{C_{pf}}{C_{pg}}\right)\right]\mathbf{\Delta E}_b \tag{285}$$

These equations correspond to Eq. (176), page 201, for the simple case, excluding the effects of interelectrode capacitances. The

equivalent network corresponding to Eq. (285) is shown in the right-hand half of Fig. 157.

Similarly, we can draw an equivalent parallel-circuit network for the grid circuit, *provided only* that ΔE_0 is zero. Such a network is shown in the left-hand half of Fig. 157.

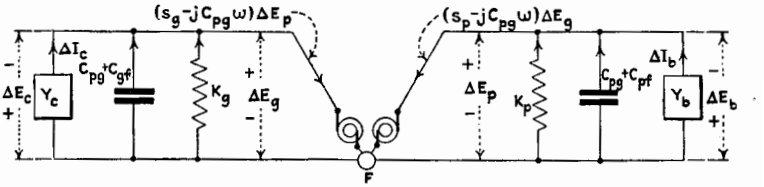


FIG. 157.—Equivalent circuits of a triode including effects of tube capacitances and expressed in terms of resistanceless generators.

Equations (282) to (285) will be found more generally useful than the simple expression of the e-p-c. theorem when the triode system is operated at frequencies for which the capacitance currents are appreciable. Equation (284) is especially convenient in calculating the voltage amplification of an amplifier.

CHAPTER XII

LOW-POWER AMPLIFIER

The most characteristic property of a triode is its ability to act as an amplifier. This ability is due to the fact that a small amount of electrical power expended in the grid circuit of a triode is able to control a relatively large amount of power in the plate circuit. Most of the uses of a triode depend directly or indirectly on this most important property.

119. Classification of Amplifiers.—Vacuum-tube amplifiers may conveniently be divided into three classes according to the amplitude of the electrical variations or, more specifically, according to the relative size of the path of operation on the characteristic surface of the triode.

Class A includes amplifiers which are worked at such low power that the several tube coefficients, such as u_p and k_p, are essentially constant over the entire cycle of electrical variation. Audio-frequency amplifiers are representative of this class.

Class B includes those amplifiers for which the path of operation is relatively larger than in amplifiers of Class A. The path of operation is not so large, however, as to pass off the characteristic surface. In other words, the tube coefficients may vary during a cycle of variation of plate current, but at no time is the variation of plate current so large as to cause the plate current or plate potential to become zero.

Most low-power, radio-frequency amplifiers belong to this class. Although truly distortionless audio-frequency amplifiers belong to Class A, audio-frequency amplifiers of Class B have been receiving considerable attention. In order partially to eliminate distortion, two triodes, working as Class B amplifiers, are operated so that one triode amplifies the potential variations on one side of the time axis and the other triode amplifies the potentials on the other side of the time axis. A form of Class B audio-frequency amplifier will be described in Chap. XXIII.

Class C includes those amplifiers for which the path of operation is so large that the plate current is reduced to zero during

part of each cycle. Radio-frequency power amplifiers operated at high efficiency belong to this class.

The discussion in this chapter is confined to low-power amplifiers of Class A.

120. Causes of Wave-form Distortion.—An ideal amplifier of Class A is one which gives an output similar in all respects to the input, except for being magnified. In other words, in an ideal amplifier the wave form of the plate current is exactly similar to the wave form of the impressed grid potential. Such an amplifier has no *wave-form distortion.* A few causes of wave-form distortion will now be considered.

Any complex wave form can be analysed into component pure sine-wave variations plus a constant. With this in mind, there are three possible causes of wave-form distortion which may be present singly or together in any amplifying system. These three causes of wave-form distortion are:

1. Shift of phase of the components of different frequencies.
2. Unequal amplification at different frequencies.
3. Introduction of frequencies in the output not present in the input.

The first cause is of no practical importance, at least when dealing with audio-frequency amplifiers, because the ear does not detect phase shifts among the several component vibrations. The ear acts as a harmonic analyzer, taking note only of the intensity and frequency, or pitch, of the components of any complex wave.

Phase shift is fatal in an amplifier designed for increasing the sensitivity of an oscillograph, for in such an instrument the wave form must be recorded unchanged. At low frequencies the high-vacuum triode introduces no lag and hence no phase shift, and therefore any phase shift present must be due to the character of the input and output circuits. To secure the absence of phase shift, all series reactances and shunt susceptances must be negligible.

The second cause of wave-form distortion, often known as *frequency distortion,* is not due to any characteristic such as lag of the electron streams within the high-vacuum triode. It is assignable to the characteristics of the input and output circuits when these characteristics are determined by taking into account the equivalent input and output admittances of the triode.

Distortion due to the third cause, often known as *amplitude distortion*, is due to a nonlinear relation between the instantaneous output current or voltage and the instantaneous input voltage. This nonlinearity may be due to the curvature of the characteristic surface of the tube, or it may be due to the characteristics of the output device. For example, iron in an output device, if worked over a large amplitude of magnetization, may have sufficient hysteresis to cause distortion of this third type.

We shall now show mathematically that a nonlinear relation between output and input inevitably results in the introduction of new sinusoidal components. Assume that the relation between the output current and the input voltage is given by the series

$$i = a + be + ce^2 + \cdots \tag{286}$$

For simplicity suppose that e has only two components, of frequencies $\omega_1/2\pi$ and $\omega_2/2\pi$, represented by the relation

$$e = \hat{E}_1 \sin \omega_1 t + \hat{E}_2 \sin \omega_2 t \tag{287}$$

Substituting Eq. (287) in Eq. (286),

$$\begin{aligned} i &= a + b\hat{E}_1 \sin \omega_1 t + b\hat{E}_2 \sin \omega_2 t + c\hat{E}_1{}^2 \sin^2 \omega_1 t \\ &\quad + c\hat{E}_2^2 \sin^2 \omega_2 t + 2c\hat{E}_1\hat{E}_2 \sin \omega_1 t \sin \omega_2 t + \cdots \\ &= a + \frac{c\hat{E}_1^2}{2} + \frac{c\hat{E}_2^2}{2} + b\hat{E}_1 \sin \omega_1 t + b\hat{E}_2 \sin \omega_2 t - \frac{c\hat{E}_1^2}{2} \cos 2\omega_1 t \\ &\quad - \frac{c\hat{E}_2^2}{2} \cos 2\omega_2 t + c\hat{E}_1\hat{E}_2 \cos (\omega_1 - \omega_2)t - c\hat{E}_1\hat{E}_2 \cos (\omega_1 + \omega_2)t \\ &\quad + \cdots \end{aligned} \tag{288}$$

Examination of Eq. (288) shows that a nonlinear amplifier introduces harmonics of all frequencies present in the input, together with new components having frequencies equal to the sum and difference of each pair of frequencies in the input. If terms of powers of e higher than the second had been included in Eq. (286), more frequencies would be present in Eq. (288). The sum and difference tones often cause much more annoyance in audio amplification than the harmonics, because the harmonics in general are harmonious, whereas the sum and difference tones may be discordant. The sum and difference tones are the cause of a disagreeable fuzziness or roughness in the output of

many amplifiers, noticeable especially when listening to complex sounds as from a chorus or orchestra.

It is apparent that, in order to obtain distortionless amplification, the path of operation of the triode must be located on the plane portion of the characteristic surface of the tube. The amplitude of variation of plate and grid potentials must be so small that u_p and r_p are essentially constant over the cycle of operation. Only amplifiers of Class A approach fulfillment of the rigid requirements for no wave-form distortion.

In the adjustment of amplifiers of Class A, it is not enough merely to be assured that the path of operation is on the plane portion of the plate-current surface. If the grid circuit contains an impedance in series with the impressed voltage, and if the grid draws conduction current, the grid potential is distorted because of the nonlinear relation between the grid current and grid potential. It is highly essential that the grid be polarized negatively to an amount which precludes the possibility of the grid potential's becoming positive at any time. This usually demands a high plate-battery voltage in order to maintain the path of operation in a plane portion of the plate-current surface.

The amount of amplification of a triode with its plate load can be expressed in several ways.

Voltage amplification is the most usual measure of amplification and is the ratio of the alternating output voltage to the alternating input voltage. Usually the output voltage is the voltage across the plate load or is the secondary voltage of a transformer connected in the plate circuit. The input voltage is usually taken as the grid voltage.

Current amplification is defined as the ratio of the current through a device connected in the plate circuit to the current that would flow through the same device if the amplifier tube were not used. Current amplification depends upon the impedance in the grid circuit in series with the impressed voltage. Except in certain very special cases, current amplification is seldom used as a measure of the performance of a vacuum-tube amplifier. For example, it is usually unfair to assume that the same impedance would be chosen for the device when used both with and without the triode. Some examples will be given later to illustrate this point.

Another definition sometimes used for current amplification is the ratio of the alternating current in the output circuit to the alternating current in the input circuit.

Power amplification is the ratio of the alternating power output to the alternating power input. If, as in the case of amplifiers of Class A, there is no grid conduction current, the power input at very low frequencies is practically zero. On the other hand, because of the tube capacitances there is a real component of the equivalent input admittance as given in Chap. XI, and an appreciable input power is demanded even at audio frequencies. The input power is given by

$$\text{Input power} = (\Delta E_g)^2 g_g \tag{289}$$

With this as the true input power for amplifiers of Class A, the true power amplification can be expressed in any special case.

Power amplification has sometimes been defined as the square of the voltage amplification. This is the ratio of power in the plate load to the power which would be developed in the same device if ΔE_g acted directly upon the device without the amplifier. This definition of power amplification is open to the same objection as has been offered already to current amplification.

I. AMPLIFIERS WITH SPECIAL PLATE LOADS

121. Amplifier with Resistance Load.—Figure 158 shows an amplifier with a pure resistance as plate load. By the approximate form of the e-p-c. theorem,

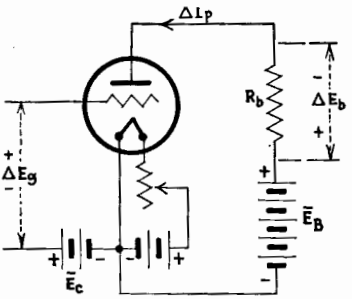


FIG. 158.—Triode with pure resistance as plate load.

$$\Delta I_p = \frac{u_p \Delta E_g}{r_p + R_b} \tag{290}$$

The output voltage is ΔE_b, which is equal to $(\Delta I_p)R_b$. The voltage amplification, denoted by (V.A.), is

$$\text{(V.A.)} = \frac{u_p R_b}{r_p + R_b} \tag{291}$$

Equation (291) can be obtained directly from Eq. (284) by neglecting the terms containing ω.

From Eq. (291) it is evident that, the larger R_b is, the greater (V.A.) is, but the gain is slow after R_b becomes greater than r_p, as is shown by the plot of Fig. 159. It would be advantageous to make R_b very large except for the resulting disadvantage

due to the steady-voltage drop through R_b which demands a higher plate-battery voltage in order to maintain the proper plate voltage.

If R_c is a resistance in the grid circuit in series with an impressed alternating e.m.f. ΔE_0, the current amplification, denoted by (I.A.), is

$$(\text{I.A.}) = u_p \frac{R_c + R_b}{r_p + R_b} \tag{292}$$

The power amplification, denoted by (P.A.), is

$$(\text{P.A.}) = \frac{(\Delta I_p)^2 R_b}{(\Delta E_g)^2 g_g} \tag{293}$$

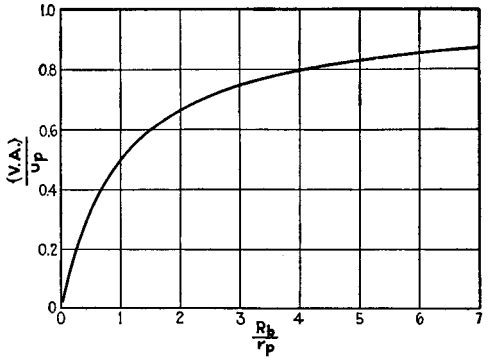


FIG. 159.—Variation of (V.A.) of an amplifier with the resistance of the plate load.

where g_g is given by Eq. (274), page 275, and ΔI_p by the more exact Eq. (284). Substituting the values of g_g and ΔI_p in Eq. (293) gives, for pure resistance load,

$$(\text{P.A.}) = \frac{1 + \dfrac{u_p^2}{(r_p C_{pg}\omega)^2}}{1 + \dfrac{R_b}{r_p}\left[1 + u_p\left(1 + \dfrac{C_{pf}}{C_{pg}}\right)\right]} \tag{294}$$

As an example, assume $u_p = 10$, $r_p = 15{,}000$ ohms, $R_b = 45{,}000$ ohms, $C_{pg} = C_{pf} = 10\mu\mu\text{f}$, and $\omega = 2\pi 5{,}000$. The voltage amplification by Eq. (291) is 7.5, and the power amplification by Eq. (294) is 70,300.

It is instructive to examine the $i_p - e_p$ diagram for the resistance-loaded amplifier, shown in Fig. 160. Since there must be no grid current, only that portion of the diagram to the right of the

curve for $e_g = 0$ can be used. The path of operation lies along the resistance line making with the vertical an angle whose tangent is R_b. In Fig. 160, R_b is assumed to be about equal to r_p. The quiescent point, determined by the grid-battery voltage $\bar{E}_c$, is so chosen that, as the grid potential varies, the path of operation ph does not pass to the left of the $e_g = 0$ curve. The amplitude of the output voltage ΔE_b is the horizontal projection of *half* of the path of operation. The path of operation is on a part of the diagram where the graphs are nearly straight, parallel,

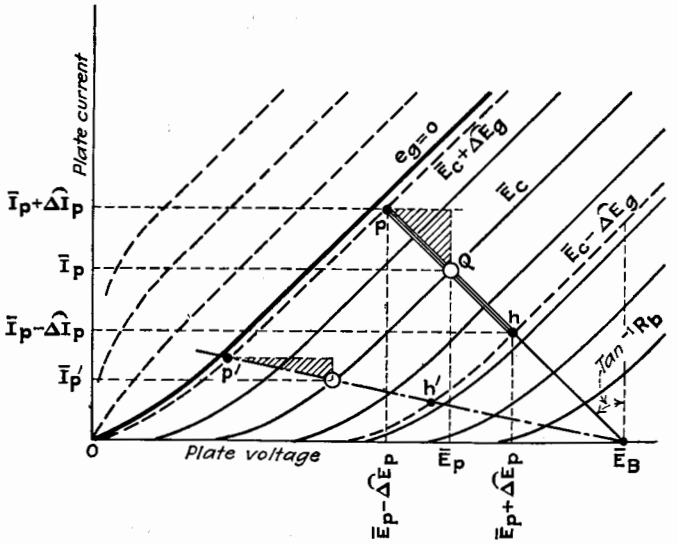


FIG. 160.—Path of operation of a triode used as an amplifier with a resistance load.

and equally spaced, showing that u_p and r_p are practically constant.

The effects of employing different values of $\bar{E}_B$ and R_b are easily understood from the diagram. For example, if a larger value of R_b is used, while the plate-battery voltage is kept the same, the resistance line takes on some new position such as that shown by the dot-and-dash line, and if the grid-polarizing potential is the same, the new path of operation is $p'h'$. This new path of operation cuts across the curved portions of the plate-current curves, which might give the impression that considerable distortion would result. As a matter of fact, since the curves for constant grid potential are nearly equally spaced as they cut

across the path $p'h'$, the distortion is less than might be inferred at first. The higher resistance gives greater voltage amplification and draws less steady current $\bar{I}'_p$ from the plate battery. Therefore it is generally advantageous to employ a plate-load resistance which is several times as large as r_p.

When properly adjusted, the resistance-loaded amplifier, shown in Fig. 158, is the most nearly distortionless form. The principal point of advantage of the resistance-loaded amplifier lies in the fact that all frequencies are equally amplified, if the capacitance effects are negligible. Its disadvantages are the necessity of using high plate-battery voltages and a lower value of voltage amplification than is attainable with other forms of amplifier connections to be discussed. The resistance-loaded

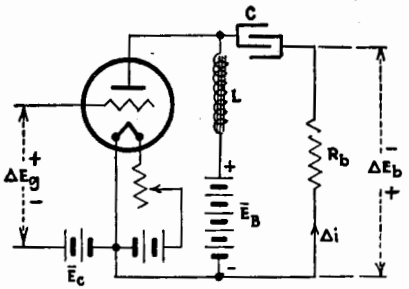


FIG. 161.—Resistance-loaded amplifier using parallel feed.

amplifier is the only form which can be used at zero frequency, *i.e.*, for abrupt changes in steady grid potentials. Such an amplifier is sometimes spoken of as a *direct-current amplifier*, clearly a misnomer.

Figure 161 shows a modified form of amplifier which can be considered as a pure-resistance-loaded amplifier for frequencies so high that C has negligible reactance compared with R_b, and L has a very large reactance compared with R_b. The advantage of this arrangement is the elimination of the steady-current loss in R_b. Except for a small drop in the resistance of L, the steady plate voltage is equal to $\bar{E}_B$. $\bar{E}_B$, therefore, need not be so large as in the simple case previously considered. The plate load of the amplifier of Fig. 161 has for the steady current a low value of resistance, $\bar{R}_b$, equal to the resistance of L, and has for the alternating current a high value of resistance, $\tilde{R}_b$, equal to R_b. The $i_p - e_p$ diagram for this case is shown in Fig. 162.

Before leaving this brief discussion of the resistance-loaded amplifier, consider the effects of the tube capacitances. Equation (291) was deduced by neglecting terms containing capacitance. The complete expression for the voltage amplification of a

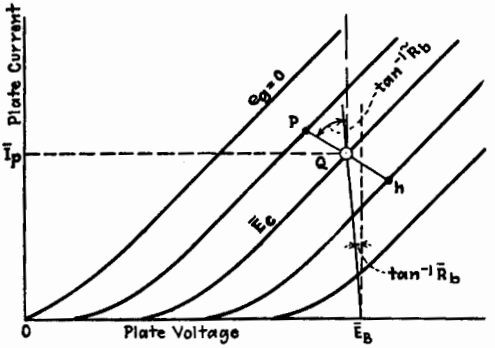


FIG. 162.—Path of operation of resistance-loaded amplifier when $\bar{R}_b \neq \bar{R}_b$.

resistance-loaded amplifier is obtained directly from Eq. (284), and is

$$(\text{V.A.}) = \frac{\sqrt{u_p^2 + (r_p C_{pg}\omega)^2}}{\sqrt{\left[1 + \frac{r_p}{\bar{R}_b}\right]^2 + \left[r_p(C_{pg} + C_{pf})\omega\right]^2}} \qquad (295)$$

Evidently the capacitances have appreciable effects if $[r_p(C_{pg} + C_{pf})\omega]^2$ is appreciable compared to unity. If $C_{pg} + C_{pf}$ is $20\mu\mu$f, r_p is 20,000 ohms, ω is $2\pi 5{,}000$, then $[r_p(C_{pg} + C_{pf})\omega]^2$ is only 0.0158. This value is negligible in comparison with unity. Hence, the capacitance or frequency terms are of no consequence at frequencies below 5,000 cycles per second unless r_p is of the order of 100,000 ohms or more, or unless $C_{pg} + C_{pf}$ is very much greater than $20\mu\mu$f. At frequencies well above 5,000 cycles per second, the frequency terms are not negligible, frequency distortion exists, and the value of the voltage amplification is less than at lower frequencies. These effects will be discussed more in detail in Chap. XV.

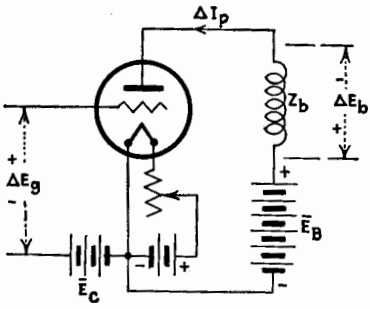


FIG. 163.—Amplifier with inductive load.

122. Amplifier with Inductive Load.—We shall now consider a triode with an inductance, possessing resistance, as plate load. In practice, the resistance is usually made as small as possible, and the inductance as large as possible. The connections are indicated in Fig. 163.

The alternating component of plate current of frequency $\omega/2\pi$ is

$$\Delta \mathbf{I}_p = \frac{u_p \Delta \mathbf{E}_g}{r_p + \mathbf{Z}_b} \tag{296}$$

where $\mathbf{Z}_b = R_b + jL_b\omega$. The voltage amplification is given in numerical terms by

$$(\text{V.A.}) = \frac{u_p \sqrt{R_b^2 + L_b^2\omega^2}}{\sqrt{(r_p + R_b)^2 + L_b^2\omega^2}} \tag{297}$$

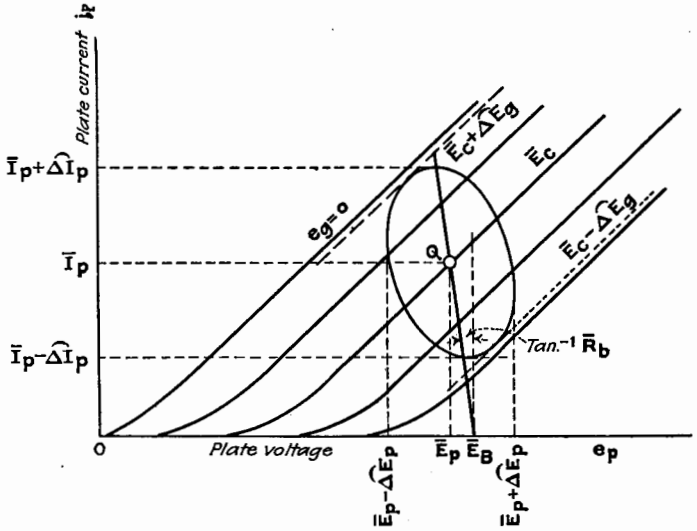


FIG. 164.—Path of operation for amplifier having an inductive plate load.

The voltage amplification as given by Eq. (297) is evidently a function of frequency. If, however, $L_b^2\omega^2$ is made very large compared with $(r_p + R_b)^2$, then over a limited range of frequencies the amplifier has practically no frequency distortion. To make $L_b^2\omega^2$ large compared with $(r_p + R_b)^2$, *i.e.*, about ten times as great, requires a very large inductance. As an example, if $(r_p + R_b)$ is 20,000 ohms, $L_b\omega$ must be at least 63,000 ohms. If the frequency is 100 cycles, L_b must be at least 100 henries. With

such a large inductance, the distributed capacitance of the coil, together with the tube capacitance, is usually sufficient to tune the inductance for some frequency within the audible range of frequencies. Under this condition the amplifier is a tuned amplifier, a type to be considered later.

The $i_p - e_p$ diagram for the amplifier of Fig. 163 is shown in Fig. 164. If we assume that R_b has the same value for steady and alternating currents, the path of operation is an ellipse, as shown in Fig. 164. (The reader may well refer back to Chap. VIII where the path of operation for an inductive load is deduced.) Here, too, the path of operation must always remain on the portion of the diagram where there is no grid current, a region which is usually bounded by the $e_g = 0$ curve. Although the path of operation is an ellipse, it falls on a portion of the characteristic surface where u_p and r_p are essentially constant. Hence all circuit elements are constant and there is no distortion of the character of Types 2 and 3. A phase shift is introduced which is of no importance for audio-frequency amplification.

The power output of this type of amplifier is $I_p^2 R_b'$, where R_b' is that portion of R_b which determines the useful power output. If the output device is a loud-speaker or telephone receiver, the motional resistance times the square of the current gives the output power, most of which is converted into sound.

$$\text{Power output} = \frac{u_p^2 R_b' (\Delta E_g)^2}{(r_p + R_b)^2 + L_b^2 \omega^2} \tag{298}$$

The power amplification can be expressed in terms of the equivalent input conductance

$$(\text{P.A.}) = \frac{u_p^2 R_b'}{[(r_p + R_b)^2 + L_b^2 \omega^2] g_g} \tag{299}$$

where g_g is given by Eq. (278), page 277.

123. Amplifier with Tuned Load.—Consider the case for which the plate load consists of an inductive resistance in parallel with a tuning condenser C', as shown in Fig. 165. The admittance of the load is

$$\mathbf{Y}_b = \frac{R}{R^2 + L^2\omega^2} - j\left(\frac{L\omega}{R^2 + L^2\omega^2} - C'\omega\right) \tag{300}$$

Substituting this expression in Eq. (284) gives for the voltage amplification of the amplifier of Fig. 165

$$(\text{V.A.}) = \frac{\sqrt{u_p^2 + (r_p C_{pg}\omega)^2}}{\sqrt{\left[1 + \frac{r_p R}{R^2 + L^2\omega^2}\right]^2 + r_p^2\left[\frac{L\omega}{R^2 + L^2\omega^2} - C\omega\right]^2}} \tag{301}$$

where $C = C' + C_{pg} + C_{pf}$.

The voltage amplification given by Eq. (301) varies rapidly with frequency, rising to a maximum for a value of ω which is close to the value which gives resonance in the load. For practical purposes, the condition of maximum voltage amplification is obtained by equating to zero the second bracket in the denominator of Eq. (301).

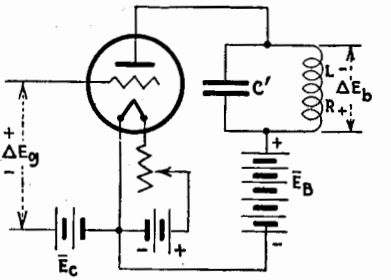


FIG. 165.—Amplifier with a tuned plate load.

$$\text{Condition for max. (V.A.)} \qquad \frac{L\omega}{R^2 + L^2\omega^2} = C\omega \tag{302}$$

The value of the maximum voltage amplification is

$$\left.\begin{aligned}(\text{V.A.})_{\max.} &= \frac{\sqrt{u_p^2 + (r_p C_{pg}\omega)^2}}{1 + \frac{r_p R}{R^2 + L^2\omega^2}} \\ &= \frac{\sqrt{u_p^2 + (r_p C_{pg}\omega)^2}}{1 + \frac{r_p RC}{L}} \\ &= \frac{\sqrt{u_p^2 + (r_p C_{pg}\omega)^2}}{1 + \frac{r_p}{(R_b)_{\text{res.}}}}\end{aligned}\right\} \tag{303}$$

In the last form of the expression for maximum voltage amplification $(R_b)_{\text{res.}} = L/CR$ denotes the equivalent resistance of the load when resonated by capacitance $C' + C_{pg} + C_{pf}$. To give a large value of the maximum voltage amplification, $(R_b)_{\text{res.}}$ should be at least equal to r_p, and preferably several times as large. The term $(r_p C_{pg}\omega)^2$ is negligible in comparison with u_p^2 at audio frequencies, but is appreciable at radio frequencies.

An idea of the sharpness of resonance, that is, the variation of voltage amplification with frequency, can be obtained conveniently by forming the ratio

$$\left[\frac{(\text{V.A.})_\omega}{(\text{V.A.})_{\max.}}\right]^2 = \frac{u_p^2 + (r_p C_{pg}\omega)^2}{u_p^2 + (r_p C_{pg}\omega_m)^2} \cdot \frac{\left[1 + \dfrac{r_p R}{R^2 + L^2\omega_m^2}\right]^2}{\left[1 + \dfrac{r_p R}{R^2 + L^2\omega^2}\right]^2 + r_p^2\left[\dfrac{L\omega}{R^2 + L^2\omega^2} - C\omega\right]^2} \tag{304}$$

where ω_m is the angular velocity that gives maximum voltage amplification and is found by solving Eq. (302) for ω. The first ratio on the right-hand side of Eq. (304) varies slowly with ω and without appreciable error can be taken as unity. If we make the further approximation that the ω in the first term in the denominator can be taken with little error as ω_m, then, by equating the two terms of the denominator, we can find the two values of ω, one greater and the other less than ω_m, for which the ratio of

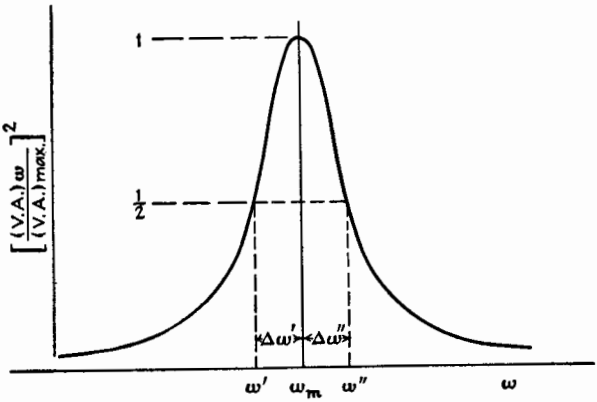


FIG. 166.—Variation of $(\text{V.A.})_\omega$ with frequency for an amplifier having a tuned plate load.

Eq. (304) is reduced to half value. Let $\omega' = \omega_m - \Delta\omega'$ and $\omega'' = \omega_m + \Delta\omega''$ be the two values of ω just referred to and indicated in Fig. 166. Then

$$\left.\begin{aligned} \Delta\omega' &= \frac{K}{2C\left(1 + K\sqrt{\frac{L}{C}}\right)} + \frac{K^2\sqrt{LC}}{4C^2\left(1 + K\sqrt{\frac{L}{C}}\right)^3} + \cdots \\ \Delta\omega'' &= \frac{K}{2C\left(1 - K\sqrt{\frac{L}{C}}\right)} - \frac{K^2\sqrt{LC}}{4C^2\left(1 - K\sqrt{\frac{L}{C}}\right)^3} - \cdots \end{aligned}\right\} \tag{305}$$

where $K = k_p + \dfrac{RC}{L}$.

In the derivation of Eqs. (305), $\Delta\omega^2$ is neglected in comparison with $2\omega_m\Delta\omega$, and ω_m is given its approximate value of $1/\sqrt{\bar{L}\bar{C}}$.

The fractional width of the resonance curve for

$$\left[\frac{(\text{V.A.})_\omega}{(\text{V.A.})_{\text{max.}}}\right]^2 = \frac{1}{2}$$

is a convenient measure of the sharpness of the response curve. As obtained from Eqs. (305),

$$\left.\begin{aligned}\text{Fractional band width} = \frac{\omega'' - \omega'}{\omega_m} &= K\sqrt{\frac{\bar{L}}{\bar{C}}}\cdot\frac{1}{1 - K^2\dfrac{\bar{L}}{\bar{C}}} \\ &= \sqrt{\frac{\bar{L}}{\bar{C}}}\left(k_p + \frac{RC}{\bar{L}}\right)\text{(approx.)} \\ &= k_p\sqrt{\frac{\bar{L}}{\bar{C}}} + \eta\ \text{(approx.)}\end{aligned}\right\} \quad (306)$$

where $\eta = R/L\omega_m$.

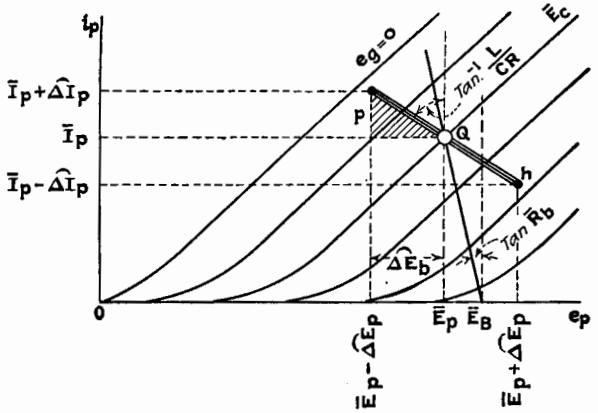


FIG. 167.—Path of operation of an amplifier having a tuned plate load set at resonance.

Equation (303) demonstrates that, for large amplification, L/RC should be at least as large as r_p and preferably two or more times as large. Under such circumstances, Eq. (306) shows that the sharpness of the resonance curve is controlled largely by the factor $\sqrt{\bar{L}/\bar{C}}$. The resonance curve is sharper the smaller the ratio L/C.

As an example, assume $k_p = 70$ micromhos, $C = 100\mu\mu\text{f}$, $L = 250\mu\text{h}$, and $\eta = 0.05$. The frequency at resonance is about

10^6 cycles per second, and the fractional width of the response curve is 0.161. Equation (303) gives as the maximum voltage amplification $0.816u_p$.

At resonance, the plate load is practically a pure resistance and the $i_p - e_p$ diagram is as shown in Fig. 167. The steady plate current is determined in value by the intersection of the curve corresponding to the grid-polarizing voltage and the resistance line making an angle with the vertical whose tangent is R, assuming R is the same for both steady and alternating current. The path ph passes through this intersection point and makes an angle with the vertical whose tangent is the equivalent resistance of the parallel circuit, *i.e.*, L/CR.

The a-c. power dissipated in the resistance R at resonance is given by the area of the shaded triangle of Fig. 167 and is equal to $\frac{u_p^2(\Delta E_g)^2 L}{\left(r_p + \frac{L}{CR}\right)^2 CR}$. The power input is $(\Delta E_g)^2 g_g$, where g_g is given by Eq. (274). The power amplification at resonance is given by the following equation which is the same as Eq. (294) with C_{pf} considered zero, this capacitance being part of the tuning capacitance.

$$(\text{P.A.})_{\text{res.}} = \frac{1 + \frac{u_p^2}{(r_p C_{pg}\omega)^2}}{1 + \frac{(R_b)_{\text{res.}}}{r_p}(1 + u_p)} \tag{307}$$

If we add to the data given in the example, $u_p = 10$ and $C_{pg} = 10\mu\mu\text{f}$, then $r_p C_{pg}\omega = 0.90$ and $(R_b)_{\text{res.}}/r_p = 2.22$. The power amplification, by Eq. (307), is 7.0.

124. Amplifier with Transformer Load.—For all of the amplifier connections so far considered, the voltage amplification is less than the voltage ratio u_p of the tube. The obvious way to increase the voltage amplification is to use a step-up transformer in the plate circuit of the amplifier tube. Because of the much larger values of voltage amplification obtained by the use of transformers, the great majority of both audio-frequency and radio-frequency amplifiers are of this type.

In studying this very important kind of amplifier it is convenient to consider two forms. The first form, used mostly for audio-frequency amplification, uses an untuned closely coupled transformer, as indicated in Fig. 168. The close coupling is

generally secured by winding the coils on a laminated iron core. The second form of transformer is used in radio-frequency amplifiers and consists of a tuned secondary winding, loosely coupled to a small primary winding, as indicated in Fig. 169. This transformer ordinarily has no iron and is called an *air-cored transformer*.

A clearer picture of the main characteristics of amplifiers with transformers is obtained if the theory is first developed neglecting certain capacitance effects. For example, the windings of the untuned transformer have distributed capacitances and these, together with the tube capacitances, have considerable effect, especially at the higher audio frequencies. The capacitances may be sufficient to transfer the so-called *untuned transformer* to the class of the *tuned transformer*. Furthermore, there is capacitance between the primary and secondary windings of the untuned transformer which in the final analysis should be taken into account. The plate-to-filament and grid-to-filament tube capacitances and the distributed capacitances of the windings can be taken into account more easily in the tuned transformer. This results because in the secondary circuit the capacitances are considered merely as part of the tuning capacitance, and because the primary winding is usually so small that its distributed capacitance and the plate-to-filament tube capacitance have negligible effects.

The simple theory now to be given will be extended in Chap. XVII to include those capacitance effects which are neglected here. Also the effect of connecting tubes in tandem to form multistage amplifiers will be discussed later.

a. Untuned Closely Coupled Transformer.—First, we shall consider briefly a closely coupled, untuned transformer such as is used for audio frequencies. The diagram of connections is shown in Fig. 168. The grid is polarized negatively so that no grid current flows. The voltage amplification is the ratio of the secondary voltage ΔE_2 to the grid voltage ΔE_g. Neglecting capacitance effects, we assume that the secondary coil is virtually open-circuited, so that the secondary voltage ΔE_2 divided by the voltage ΔE_1 across the primary is numerically equal to the ratio of turns N. Under these conditions

$$(\text{V.A.}) = \frac{u_p N \sqrt{R_1^2 + L_1^2\omega^2}}{\sqrt{(r_p + R_1)^2 + L_1^2\omega^2}} \tag{308}$$

If $L_1\omega$ is made very large compared with R_1 and r_p, the value of (V.A.) approaches u_pN and is nearly independent of frequency. Frequency distortion is thus considerably reduced, and all high-grade audio-frequency transformers are constructed to have a

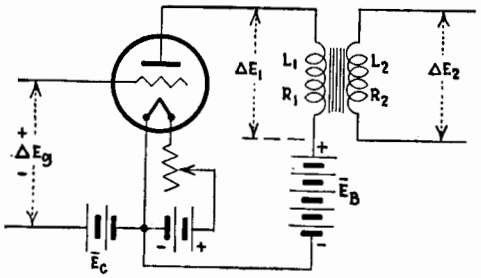


FIG. 168.—Amplifier having a plate load consisting of a closely coupled untuned transformer.

very large primary inductance. By making L_1 very large it becomes physically difficult to make L_2 much larger than L_1, so that N, the ratio of turns, is frequently not greater than 3 or 4 in high-grade transformers. When the inductances are made large, the distributed capacitances of the coils, together with other

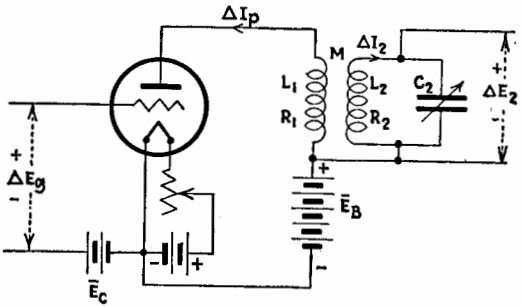


FIG. 169.—Amplifier having a plate load consisting of a loosely coupled tuned transformer.

stray capacitances, have a greater effect, and so the simple theory gives only an approximate idea of the performance.

b. Tuned Air-core Transformer.—A transformer very frequently found in radio-frequency amplifiers is the tuned transformer, as shown in Fig. 169. The following equations express the voltages in the circuits.

$$\left.\begin{aligned}(r_p + R_1 + jL_1\omega)\Delta\mathbf{I}_p + jM\omega\Delta\mathbf{I}_2 &= u_p\Delta\mathbf{E}_g \\ jM\omega\Delta\mathbf{I}_p + \left[R_2 + j\left(L_2\omega - \frac{1}{C_2\omega}\right)\right]\Delta\mathbf{I}_2 &= 0 \\ \Delta\mathbf{E}_2 &= -\frac{j\Delta\mathbf{I}_2}{C_2\omega}\end{aligned}\right\} \tag{309}$$

The solution of these equations gives, in complex form,

$$\text{(V.A.)} = \frac{-u_pM\omega}{C_2\omega\left\{\left[r_p + R_1 + jL_1\omega\right]\left[R_2 + j\left(L_2\omega - \frac{1}{C_2\omega}\right)\right] + M^2\omega^2\right\}} \tag{310}$$

Neglecting R_1 in comparison with r_p, the numerical value of the voltage amplification is

$$\text{(V.A.)} = \frac{u_pM\omega}{C_2\omega\sqrt{r_p^2 + L_1^2\omega^2}\sqrt{\left(R_2 + \frac{M^2\omega^2r_p}{r_p^2 + L_1^2\omega^2}\right)^2 + \left(X_2 - \frac{M^2\omega^2L_1\omega}{r_p^2 + L_1^2\omega^2}\right)^2}} \tag{311}$$

where $X_2 = L_2\omega - \dfrac{1}{C_2\omega}$.

The secondary capacitance C_2 is variable and in practice is adjusted to give maximum amplification. Since X_2 varies much more rapidly than $1/C_2\omega$ does when C_2 is varied, the voltage amplification is sufficiently near a maximum when the second bracket under the radical of Eq. (311) is zero. Hence:

$$C_2 \text{ for max. (V.A.)} \qquad X_2 = \frac{M^2\omega^2L_1\omega}{r_p^2 + L_1^2\omega^2} \tag{312}$$

For most practical cases, the second member of Eq. (312) is very small because of the large value of r_p, so that with little error maximum voltage amplification occurs when $X_2 = 0$ or $L_2\omega = 1/C_2\omega$. Equation (311) then reduces to

$$\text{Max. (V.A.)} = \frac{u_pM\omega}{C_2\omega\sqrt{r_p^2 + L_1^2\omega^2}\left(R_2 + \frac{M^2\omega^2r_p}{r_p^2 + L_1^2\omega^2}\right)} \tag{313}$$

If $L_1^2\omega^2$ is small in comparison with r_p^2, as is almost always the case in practice, a still further approximation can be made. Equation (313) becomes

$$\text{Approx. max. (V.A.)} = \frac{u_p M}{C_2 r_p\left(R_2 + \dfrac{M^2\omega^2}{r_p}\right)} \tag{314}$$

We still have left the possibility of changing M and finding the value of M which gives the largest value of max. (V.A.). By putting the derivative of Eq. (314) with respect to M equal to zero, we find the value of M to give max. max. (V.A.)

M to give max. max. (V.A.)

$$M\omega = \sqrt{r_p R_2} \tag{315}$$

or

$$\frac{M^2\omega^2}{R_2} = r_p \tag{316}$$

The first term of Eq. (316) is that part of the approximate equivalent impedance due to the tuned secondary circuit, referred to the primary winding of the transformer. Since the impedance of the primary winding has been assumed negligible in comparison with $M^2\omega^2/R_2$, we have the condition that the equivalent a-c. resistance of the plate load is equal to the internal resistance of the tube when the value of M is such as to secure the maximum possible voltage amplification. It was shown in Chap. VIII that the condition of matched resistances gives maximum power output. In the case being considered the conditions of maximum voltage amplification and maximum power output are the same since, with a given secondary circuit having a resistance R_2, we obtain maximum voltage across the condenser C_2 when we have maximum current in the secondary circuit, and this condition necessarily means maximum power dissipated in R_2.

To find the value of max. max.(V.A.), substitute Eq. (316) in Eq. (314), obtaining

$$\text{Max. max. (V.A.)} = \frac{u_p}{2\sqrt{r_p R_2}} \cdot \sqrt{\frac{L_2}{C_2}} = \frac{u_p L_2}{2M} \tag{317}$$

Since the resistance of the plate load is equal to r_p, the voltage across the primary winding is $\frac{1}{2}u_p\Delta E_g$ when the secondary circuit is tuned to resonance. The factor L_2/M in Eq. (317) is the equivalent step-up ratio of the transformer, giving the ratio of ΔE_2 to the voltage across the primary winding. If the coupling is very close, $M = \sqrt{L_1 L_2}$, and the equivalent step-up ratio of the transformer is $\sqrt{L_2/L_1}$. For concentrated winding, this is the ratio of the numbers of turns.

As shown by Eq. (317), the factor $u_p/\sqrt{r_p} = \sqrt{u_p s_p}$ is the criterion of merit of a triode for voltage amplification when used in the circuit shown in Fig. 169.

The simple formulas given for the amplifier with a tuned-transformer load were deduced by neglecting the effects of the tube capacitances C_{pg} and C_{pf}. When these capacitances are not greater than from 8 to 10$\mu\mu$f, and when the frequency is not much greater than 1,000 kc., or in other words when $[r_p(C_{pg} + C_{pf})\omega]^2$ is negligible in comparison with unity, the simple theory is sufficiently correct for practical purposes. If $[r_p(C_{pg} + C_{pf})\omega]^2$ is large, the voltage amplification can be calculated by Eq. (285). The complete expression is

$$(\text{V.A.}) = \frac{\Delta \mathbf{E}_2}{\Delta \mathbf{E}_g} = \frac{M(u_p - jA)}{C_2\left[r_p + \mathbf{Z}_1(1 + jA')\right]\left[\mathbf{Z}_2 + \dfrac{M^2w^2(1 + jA')}{r_p + \mathbf{Z}_1(1 + jA')}\right]} \tag{318}$$

where $A' = r_p(C_{pg} + C_{pf})\omega$.

The numerical value of (V.A.) from Eq. (318) is

$$(\text{V.A}) = \frac{M\sqrt{u_p^2 + A^2}}{C_2 Z'\sqrt{\left\{R_2 + \dfrac{M^2\omega^2[r_p + R_1(1 + A'^2)]}{Z'^2}\right\}^2 + \left\{X_2 - \dfrac{M^2\omega^2[X_1(1 + A'^2) - r_p A']}{Z'^2}\right\}^2}} \tag{319}$$

where

$$Z' = \sqrt{(r_p + R_1 - A'X_1)^2 + (X_1 + A'R_1)^2}.$$

With any particular set of data many of the terms in Eq. (319) may be found to be negligible and the expression can be much simplified.

The power amplification for the case of a tuned air-cored transformer is given approximately by Eq. (307), where $(R_b)_{\text{res.}} = M^2\omega^2/R_2$. Actually, the plate load is not a pure resistance and the slight phase angle between ΔI_b and ΔE_p may make an appreciable difference in the calculation of (P.A.).

We shall now examine the selectivity of the amplifier as M is varied. The approximate expression for (V.A.) from Eq. (311) is

$$\text{Approx. (V.A.)}_\omega = \frac{u_p M\omega}{C_2\omega r_p\sqrt{\left(R_2 + \dfrac{M^2\omega^2}{r_p}\right)^2 + \left(L_2\omega - \dfrac{1}{C_2\omega}\right)^2}} \tag{320}$$

Then from Eqs. (314) and (320),

$$\left[\frac{(\text{V.A.})_\omega}{(\text{V.A.})_{\max.}}\right]^2 = \frac{\left(R_2 + \dfrac{M^2\omega^2}{r_p}\right)^2}{\left(R_2 + \dfrac{M^2\omega^2}{r_p}\right)^2 + \left(L_2\omega - \dfrac{1}{C_2\omega}\right)^2} \tag{321}$$

The value of ω to make $[(\text{V.A.})_\omega/(\text{V.A.})_{\text{max.}}]^2 = \frac{1}{2}$ is obtained from Eq. (321) by making

$$L_2\omega - \frac{1}{C_2\omega} = \pm\left(R_2 + \frac{M^2\omega^2}{r_p}\right) \tag{322}$$

Substituting for $1/L_2C_2$ the square of the resonance value of ω, that is, ω_m,

$$\frac{\omega^2 - \omega_m^2}{\omega^2} = \pm\frac{1}{L_2\omega}\left(R_2 + \frac{M^2\omega^2}{r_p}\right) \tag{323}$$

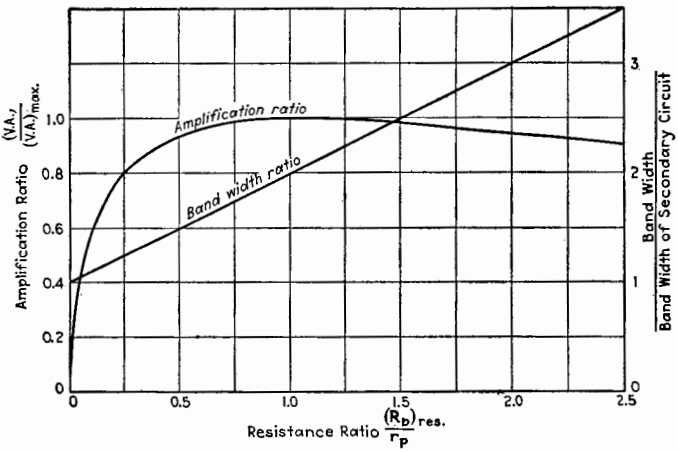


FIG. 170.—Variation of (V.A.) and selectivity with mutual inductance of the tuned transformer.

If ω is not very different from ω_m, we have as a further approximation

$$\frac{2(\omega - \omega_m)}{\omega_m} = \pm\frac{1}{L_2\omega_m}\left(R_2 + \frac{M^2\omega_m^2}{r_p}\right) \tag{324}$$

or

$$\frac{\omega'' - \omega'}{\omega_m} = \frac{R_2}{L_2\omega_m}\left(1 + \frac{M^2\omega_m^2}{R_2 r_p}\right) = \eta_2 + \frac{M}{L_2}\cdot M\omega_m k_p \tag{325}$$

where ω'' and ω' have the significance shown in Fig. 166. Equation (325) gives the fractional frequency band width at 0.707 of the maximum voltage amplification. As the band width for the secondary circuit alone is $R_2/L_2\omega_m$, the effect of the primary

circuit is to increase the band width by an amount determined by the term $M^2\omega_m^2/R_2r_p$ or $(R_b)_{res.}/r_p$, where $(R_b)_{res.}$ is the equivalent resistance of the transformer at resonance.

In Fig. 170 there is a plot of the ratio of the voltage amplification at resonance to the greatest attainable voltage amplification at resonance which occurs when M has the value given by Eq. (315), and also, there is a plot of the ratio of the band width to the band width of the secondary circuit alone, both plotted against $(R_b)_{res.}/r_p$. It was pointed out by Barkhausen[1] in 1920 and later by Hazeltine[2] that, by making $(R_b)_{res.}$ less than r_p, the gain in selectivity counts for more than the slight loss in amplification. Referring to Fig. 170, if $(R_b)_{res.} = \frac{1}{4}r_p$, the voltage amplification is reduced to 0.8 of the maximum attainable, but the band width is reduced to 0.625 of the band width at maximum voltage amplification. This smaller value of $(R_b)_{res.}$ is best attained by reducing M. In general it is better to reduce the number of turns in L_1 rather than to increase the distance between the coils L_1 and L_2.

Typical values for the constants, as taken from a commercial receiver, are given for illustration.

Primary coil	Turns	12
	D-c. resistance	0.66 ohm
	Inductance	11.5μh
	Diameter of coil	1.5 in.
Secondary coil	Turns	92
	D-c. resistance	4.76 ohms
	Inductance	278μh
	Diameter of coil	1.5 in.
	Length of coil	1.12 in.

$u_p = 8$ $\quad M = 29\mu$h
$r_p = 8{,}000$ ohms $\quad C_2 = 18$ to $295\mu\mu$f

Frequency, kilocycles	$M^2\omega^2/\bar{R}_2$ (measured), ohms	$\bar{R}_2$, ohms
600	900	13.3
1,000	870	38.2
1,400	640	100

[1] BARKHAUSEN, *Jahrb. Drahtlos. Tel. Tel.*, **16**, 82 (1920).
[2] HAZELTINE, U. S. patent 1,648,808, filed Feb. 27, 1925.

Using the figures for 1,000 kc., the (V.A.) is 7.5, whereas the maximum (V.A.) attainable with the given secondary circuit is 12.6. The fractional band width is 0.024.

II. CONDITIONS FOR MAXIMUM POWER

In studying the conditions for the development of maximum power in the plate circuit of an amplifier, several cases arise according to the initial conditions stated for the problem.

125. Maximum Power Output When Tube and E_g Are Given.—If the conditions are that the impressed alternating grid voltage ΔE_g is given and fixed, and it is desired to find $\bar{R}_b$ to be used with a given tube so that maximum power is developed in $\bar{R}_b$, then, as shown in Chap. VIII, the condition is

$$\bar{R}_b = r_p \tag{326}$$

The value of the maximum power output for a given triode is

$$\text{Maximum power output} = \frac{u_p^2(\Delta E_g)^2}{4r_p} \tag{327}$$

$$= \tfrac{1}{4}u_p s_p(\Delta E_g)^2 \tag{328}$$

Therefore, $u_p s_p$ is the criterion of merit of a triode when the greatest power output is desired under the condition of a given grid voltage. This criterion is the square of the criterion of merit for voltage amplification of a triode when used with a tuned transformer.

***126. Maximum Undistorted Power Output When Tube and $\bar{E}_B$ Are Given and $\bar{R}_b = \bar{R}_b$.**—Very often the problem is stated in another way. A particular tube is given to be operated at a given battery voltage $\bar{E}_B$. The value of $\bar{R}_b$, which is assumed to be the same as $\bar{R}_b$, is desired to give *maximum undistorted* power output. Figure 171 represents the $i_p - e_p$ diagram for the tube, and $\bar{E}_B$ is the given steady plate-circuit voltage. The resistance line drawn from point $\bar{E}_B$ on the voltage axis makes some angle, as yet undetermined, with the vertical. A horizontal line is drawn dividing the straight portion of the characteristic curves above this line from the curved portions below. Such a line drawn for a current of i' is shown in Fig. 171. The path of operation to give no distortion must be confined to the region of straight characteristics included between the two heavy lines in the diagram, one for $\bar{E}_g = 0$ and the other the horizontal line at i'.

The path is ph. The value of $\tilde{R}_b$ is sought to give maximum power, or to give maximum area to the shaded triangle.

The derivation for finding $\tilde{R}_b$ is as follows: Let the intersection of the projected straight portion of the zero-grid-voltage curve with the voltage axis be e', and the difference $\bar{E}_B - e'$ be denoted by $\bar{E}''$. Then

$$2\overline{Ql} + \overline{jm} = \frac{\bar{E}''}{r_p + \tilde{R}_b}$$

$\overline{Ql}$ is equal to $\sqrt{2}I_p$. The power output is

$$\text{Power output} = \frac{\tilde{R}_b}{8}\left[\frac{\bar{E}''}{r_p + \tilde{R}_b} - i'\right]^2 \tag{329}$$

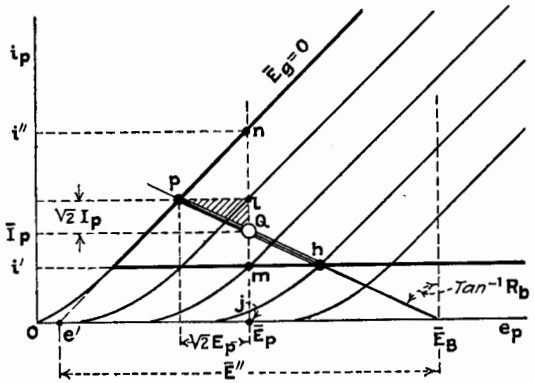


FIG. 171.—Construction for maximum undistorted power output when $\bar{E}_B$ is given.

The value of $\tilde{R}_b$ to give maximum power output can be found by equating to zero the derivative of Eq. (329) with respect to $\tilde{R}_b$. The result is

$$\tilde{R}_b = \sqrt{\left(\frac{\bar{E}''}{2i'}\right)^2 + 2r_p\frac{\bar{E}''}{i'}} - r_p - \frac{\bar{E}''}{2i'} \tag{330}$$

127. Maximum Undistorted Power Output When Tube and $\bar{E}_p$ Are Given.—The case now to be considered is applicable to a plate load having a steady-current resistance either equal to or not equal to its a-c. resistance. If the steady-current resistance is negligible, as is often true, $\bar{E}_p$ and $\bar{E}_B$ are practically the same.

We can use Fig. 171 for this case, the conditions of the problem, however, being different from those of Sec. 126. $\bar{E}_p$ is given, and

a vertical line nj is drawn above $\bar{E}_p$. As before, the path of operation ph for no distortion must be entirely within the area bounded by the curve for $\bar{E}_g = 0$ and the horizontal line at i' below which the characteristic curves have appreciable curvature.

The length mn is fixed by the conditions of the problem. Let $\overline{mn}$ be denoted by $i'' - i'$. Then

$$i'' - i' = 2\sqrt{2}I_p + \frac{\sqrt{2}E_p}{r_p} \tag{331}$$

But $E_p = I_p\tilde{R}_b$, and Eq. (331) becomes

$$I_p = \frac{i'' - i'}{\sqrt{2}\left(2 + \frac{\tilde{R}_b}{r_p}\right)} \tag{332}$$

The alternating power output is

$$\text{Power output} = \frac{(i'' - i')^2\tilde{R}_b}{2\left(2 + \frac{\tilde{R}_b}{r_p}\right)^2} \tag{333}$$

The value of $\tilde{R}_b$ which makes the power output a maximum is

$$\tilde{R}_b = 2r_p \tag{334}$$

Condition (334) is met when $\overline{ml} = \overline{ln}$ in Fig. 171. This condition is shown in Fig. 172. The steady voltage $\bar{E}_B$ is given by the intersection of the resistance line for $\tilde{R}_b$ drawn through Q and the voltage axis.

Substituting Eq. (334) in Eq. (333) gives for the maximum power output

$$\text{Max. power output} = \frac{(i'' - i')^2 r_p}{16} = 2I_p^2 r_p \tag{335}$$

The value of E_g necessary to give maximum power output can be obtained from the expression for I_p,

$$I_p = \frac{u_p E_g}{r_p + \tilde{R}_b} = \frac{u_p E_g}{3r_p}$$

Hence,

$$E_g = \frac{3I_p}{s_p} = \frac{3(i'' - i')}{4\sqrt{2}s_p} \tag{336}$$

Combining Eqs. (335) and (336) gives

$$\text{Max. power} = \tfrac{2}{9}u_p s_p E_g^2 \tag{337}$$

Again we see that $u_p s_p$ is the criterion of merit of a triode when used as an amplifier.

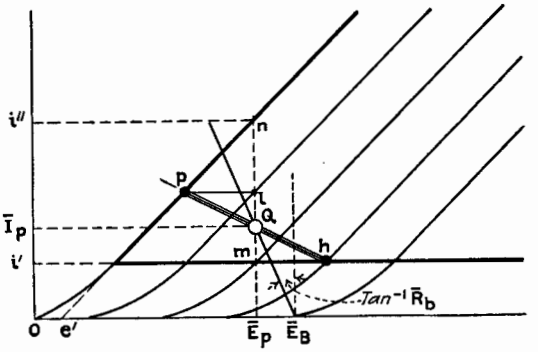


FIG. 172.—Construction for maximum undistorted power output when $\bar{E}_p$ is given.

III. SPECIAL PROBLEMS

128. Use of Vacuum-tube Amplifier with Thermocouple.—Suppose it is desired to know if a vacuum-tube amplifier will assist in detecting small changes of temperature with a thermocouple. Assume that the thermocouple has a low resistance R_0, and that when used directly with a galvanometer the latter has a low resistance R'. It is proposed to connect the thermocouple to the grid of a triode, in the plate circuit of which is a galvanometer of the same type as used directly with the thermocouple, except that it is wound with a large number of turns and has a high resistance R''. If the winding space of the two galvanometers is the same and they are similar in all respects except for the different number of turns, the deflection D in either case is proportional to the ampere-turns. The resistance is proportional to the square of the number of turns. Thus, the ratio of D' and D'', the deflections for the two galvanometers, is

$$\frac{D''}{D'} = \frac{\sqrt{R''}\Delta I''}{\sqrt{R'}\Delta I'} \tag{338}$$

Here, $\Delta I''$ is the change of current through the tube and high resistance galvanometer, and $\Delta I'$ is the change of current through

the thermocouple and low resistance galvanometer. If ΔE_0 is the e.m.f. produced by the thermocouple, $\Delta I' = \dfrac{\Delta E_0}{R_0 + R'}$ and

$$\Delta I'' = \frac{u_p \Delta E_0}{r_p + R''}.$$

Hence,

$$\frac{D''}{D'} = \sqrt{\frac{R''}{R'}} \cdot \frac{u_p(R_0 + R')}{r_p + R''} \tag{339}$$

To make the ratio D''/D' as large as possible, R'' must equal r_p. With this condition Eq. (339) becomes

$$\text{Max. } \frac{D''}{D'} = \frac{u_p(R_0 + R')}{2\sqrt{R' r_p}} \tag{340}$$

We shall assume practical values for the quantities, such as $u_p = 10$, $R_0 = 1$ ohm, $R' = 10$ ohms, $r_p = 20{,}000$ ohms. Hence, $D''/D' = 0.123$, which indicates that a vacuum-tube amplifier is of no help when the input device has low impedance, which is true for a thermocouple.

In Eq. (340), the largest deflection D' is obtained when R' is equal to R_0. With this most favorable value of R', the ratio of deflections is

$$\frac{D''}{D'} = u_p \sqrt{\frac{R_0}{r_p}} \tag{341}$$

Instead of assuming that R_0 is the resistance of a thermocouple, let it be taken as the resistance of any system in which there is a source of potential. Equation (341) gives the ratio of the deflection of a galvanometer of resistance equal to r_p when connected in the plate circuit of a triode to the deflection of the same type galvanometer connected directly to the source of potential. In the latter connection the galvanometer is assumed to have the most favorable resistance, *i.e.*, a resistance equal to the resistance of the source of potential. According to Eq. (341), the triode amplifier helps under the conditions assumed if the source of potential has a resistance R_0 greater than r_p/u_p^2 or $1/u_p s_p$. If we assume r_p is 20,000 ohms and u_p is 10, the triode increases the deflection if R_0 is greater than 200 ohms.

In the problem just considered, if the circuit is interrupted so as to produce an intermittent current, a transformer can be used

to adapt the impedance of the source to that of the vacuum-tube amplifier and considerable amplification can then be obtained.

129. Study of Special Amplifier of Fig. 161.—As a second problem we shall study further the amplifier shown in Fig. 161. Let Δi be the instantaneous current in R_b. Assuming that the resistance of the inductance is negligible, the differential equation for Δi is

$$(R_b + r_p)\frac{d^2(\Delta i)}{dt^2} + \left(\frac{R_b r_p}{L} + \frac{1}{C}\right)\frac{d(\Delta i)}{dt} + \frac{r_p}{LC}\Delta i = u_p\frac{d^2(\Delta e_g)}{dt^2} \quad (342)$$

If Δe_g is a sinusoidal e.m.f. given by

$$\Delta e_g = \sqrt{2}\Delta E_g \sin \omega t$$

the steady-state current is

$$\Delta i = \frac{\sqrt{2}u_p \Delta E_g}{\sqrt{\left(R_b + r_p - \frac{r_p}{LC\omega^2}\right)^2 + \left(\frac{r_p R_b}{L\omega} + \frac{1}{C\omega}\right)^2}} \sin\left(\omega t + \tan^{-1}\frac{\frac{R_b r_p}{L\omega} + \frac{1}{C\omega}}{R_b + r_p - \frac{r_p}{LC\omega^2}}\right) \quad (343)$$

In order that the amplifier may act as a pure resistance amplifier, the following relations must be satisfied:

$$\left.\begin{aligned} &\frac{r_p}{LC\omega^2} \ll (R_b + r_p) \\ &\left[\frac{r_p R_b}{L\omega} + \frac{1}{C\omega}\right]^2 \ll (R_b + r_p)^2 \end{aligned}\right\} \quad (344)$$

The steady-state solution does not constitute the complete solution, for there is a transient current accompanying every change in amplitude and any abrupt change in grid potential. We shall calculate the transient current due to a sudden increment $\overline{\Delta E}_g$ in the value of e_g. This transient current is obtained by solving Eq. (342) with the second member zero. The result is

$$\Delta i = \frac{u_p(\overline{\Delta E}_g)\epsilon^{-\frac{1}{2L}\left[\frac{r_p R_b}{r_p + R_b} + \frac{L}{C(r_p + R_b)}\right]t}}{(r_p + R_b)\Omega'}\left\{\Omega' \cosh \Omega' t - \frac{\frac{r_p R_b}{L} + \frac{1}{C}}{2(R_b + r_p)} \sinh \Omega' t\right\} \quad (345)$$

where

$$\Omega' = \frac{\sqrt{\left(\frac{r_p R_b}{L} + \frac{1}{C}\right)^2 - \frac{4r_p^2}{LC}}}{2(r_p + R_b)}.$$

If $\left(\frac{r_p R_b}{L} + \frac{1}{C}\right)^2 < \frac{4r_p^2}{LC}$, Ω' becomes $j\frac{\sqrt{\frac{4r_p^2}{LC} - \left(\frac{r_p R_b}{L} + \frac{1}{C}\right)^2}}{2(r_p + R_b)} = j\Omega$,

and Eq. (345) takes the oscillatory form

$$\Delta i = \frac{u_p(\overline{\Delta E}_g)\epsilon^{-\frac{1}{2L}\left[\frac{r_p R_b}{r_p + R_b} + \frac{L}{C(r_p + R_b)}\right]t}}{(r_p + R_b)\Omega}\left\{\Omega \cos \Omega t - \frac{\frac{r_p R_b}{L} + \frac{1}{C}}{2(R_b + r_p)} \sin \Omega t\right\} \tag{346}$$

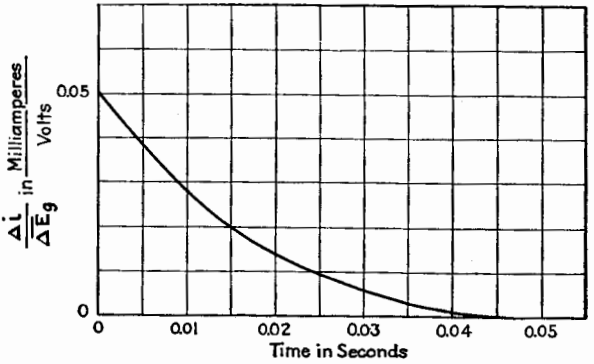


FIG. 173.—Transient current in circuit of Fig. 161.

By equating Ω to zero, we have the critical condition which divides the oscillatory case from the non-oscillatory case. If we solve to determine the value of C for the critical case, we obtain

$$C = \frac{R_b + 2r_p \pm 2\sqrt{r_p(R_b + r_p)}}{\frac{r_p R_b^2}{L}} \tag{347}$$

There are, therefore, two values of C between which the transient current is oscillatory.

The transient current is shown plotted in Fig. 173, for $r_p = R_b = 10{,}000$ ohms, $L = 100$ henries, and $C = 10\mu$f.

If the amplifier is to be used for recording sudden temporary changes in grid potential, the current after deflection should drift

slowly. The fractional rate of change of the current at $t = 0$ is

$$\frac{1}{(\Delta i)_{t=0}} \cdot \left(\frac{d(\Delta i)}{dt}\right)_{t=0} = \frac{\frac{R_b r_p}{L} + \frac{1}{C}}{r_p + R_b} \tag{348}$$

For the constants as given, the fractional change of the current in one one-thousandth of a second is 0.055, showing that a square-topped increase and decrease in grid potential lasting only one one-thousandth of a second would be distorted by drift by over 5 per cent. Increasing L reduces this rate of change of the current but may make the system oscillatory if the condition of Eq. (347) holds.

130. Comparison of Tuned Circuit and Tuned Transformer as Plate Load.—As a third problem we may compare the voltage amplification of an amplifier with a tuned transformer with that of an amplifier having a tuned circuit as plate load, under the condition that the band width be the same for both. Equate the expressions for band width given by Eqs. (306) and (325) for the two forms of plate load. The result is

$$\eta + k_p\sqrt{\frac{L}{C}} = \eta_2 + \frac{M}{L_2} M\omega k_p \tag{349}$$

We may further specify that the coils used in the two cases have the same value of η, and that the same tube is used. Equation (349) becomes

$$\sqrt{\frac{L}{C}} = \frac{M}{L_2} M\omega = \left(\frac{M}{L_2}\right)^2 \sqrt{\frac{L_2}{C_2}} \tag{350}$$

where L and C apply to the tuned circuit and M, L_2, and C_2 apply to the transformer.

We may now express the ratio of the voltage amplification for the two cases from Eqs. (314) and (303), giving

$$\frac{(\text{V.A.})_{\text{trans.}}}{(\text{V.A.})_{\text{simple circuit}}} = \frac{M\left(1 + \frac{r_p RC}{L}\right)}{C_2 r_p\left(R_2 + \frac{M^2\omega^2}{r_p}\right)} \tag{351}$$

By substituting the condition of equal band width from Eq. (350), Eq. (351) reduces to

$$\frac{(\text{V.A.})_{\text{trans.}}}{(\text{V.A.})_{\text{simple circuit}}} = \frac{L_2}{M} \quad \textit{(for equal band width)} \tag{352}$$

Equation (352) shows that for equal band width the amplifier with transformer gives much greater voltage amplification by a factor L_2/M, which is equal to the step-up ratio of the transformer. This demonstrates the advantage of the transformer used in the plate circuit of an amplifier tube and explains why practically all radio-frequency amplifiers are of this type.

CHAPTER XIII

REGENERATION WITH TRIODES

It was shown in Chap. XII that the triode when properly arranged acts as an amplifier of electrical power; *i.e.*, the power output is greater than the power input. When, by suitable electrical circuits, a small fraction of the output power is fed back into the input end of the amplifier, the total input is thereby increased, resulting in an increase in the output. The input is further increased because of this increased output and the power builds up until a greatly increased output results. This process is known as *regeneration* or *retroaction.*

A necessary condition for regeneration is that some power from the output circuits of the amplifier is transferred into the input circuit. In other words, some form of coupling, either resistance coupling, inductive coupling, or capacitive coupling must exist between the output and input circuits of a triode in order that regeneration may occur.

It is very difficult to connect a triode as an amplifier, or for any other use, so that there is no coupling of any kind between the output and input circuits. For this reason, regeneration undoubtedly existed long before it was recognized. E. H. Armstrong[1] is credited with the discovery of the possibilities of regeneration in increasing the effective amplification, and in 1915 he gave several circuits by which regenerative amplification may be obtained and controlled.

The following discussion of regeneration is divided into two parts. In the first part we shall assume that the path of operation of the triode is so small that u_p and k_p remain essentially constant during a cycle of variation, and we shall also assume that the grid draws no conduction current. In the second part there is no limitation placed on the size of the path of operation.

I. REGENERATION FOR SMALL AMPLITUDES

The particular assumptions which characterize the treatment of regeneration given in this part are, first, that all amplitudes are

[1] ARMSTRONG, *Proc. I.R.E.*, **3**, 315 (1915).

so small that all of the tube characteristics such as u_p and k_p are essentially constant and, second, that k_g is zero. The steady-state solutions can then be deduced easily, because the triode system reduces to a simple network having constant elements. Regeneration as caused by several types of couplings will now be discussed.

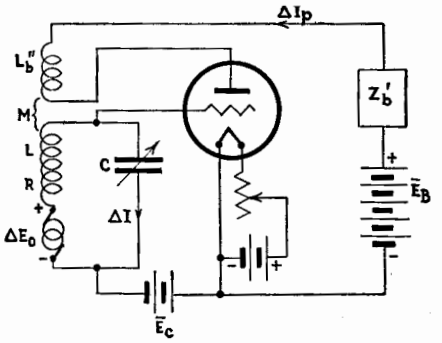


FIG. 174.—Circuits of a regenerative triode.

131. Regeneration with Inductive Coupling. Tuned Grid Circuit.—In this case there is inductive coupling between the plate and grid circuits as shown in Fig. 174. If tube capacitances are neglected, the system reduces to the equivalent network shown in Fig. 175. ΔE_0 is the impressed e.m.f. of frequency $\omega/2\pi$ and is represented in the figures as being introduced in series with the inductance L. The effect is the same if ΔE_0 is an induced e.m.f. in coil L. L_b'' is that portion of the total external plate impedance Z_b which is used for regenerative coupling with the inductance L in the grid circuit. The remainder of Z_b constitutes the output impedance Z_b'. Thus,

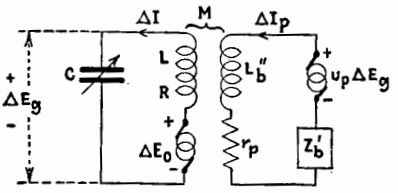


FIG. 175.—Equivalent circuits of the regenerative triode shown in Fig. 174.

$$\mathbf{Z}_b = R_b + jX_b = R_b' + j(X_b' + L_b''\omega).$$

Let $\mathbf{Z}$ stand for $R + j\left(L\omega - \frac{1}{C\omega}\right)$. The equations for the circuits shown in Fig. 175 are

$$\left.\begin{aligned} \mathbf{Z}\Delta\mathbf{I} - jM\omega\Delta\mathbf{I}_p &= \Delta\mathbf{E}_0 \\ -jM\omega\Delta\mathbf{I} + (r_p + \mathbf{Z}_b)\Delta\mathbf{I}_p &= u_p\Delta\mathbf{E}_g \\ \frac{j\Delta\mathbf{I}}{C\omega} &= -\Delta\mathbf{E}_g \end{aligned}\right\} \tag{353}$$

The solution of Eqs. (353) gives

$$\Delta\mathbf{I} = \frac{\Delta\mathbf{E}_0}{\mathbf{Z} - \dfrac{\dfrac{u_pM}{C} - M^2\omega^2}{r_p + \mathbf{Z}_b}} \tag{354}$$

or

$$\Delta\mathbf{I} = \frac{\Delta\mathbf{E}_0}{\left[R - \dfrac{\left(\dfrac{u_pM}{C} - M^2\omega^2\right)(r_p + R_b)}{(r_p + R_b)^2 + X_b^2}\right] + j\left[L\omega - \dfrac{1}{C\omega} + \dfrac{\left(\dfrac{u_pM}{C} - M^2\omega^2\right)X_b}{(r_p + R_b)^2 + X_b^2}\right]} \tag{355}$$

$$\Delta\mathbf{I}_p = \frac{-j(\Delta\mathbf{E}_0)\left(\dfrac{u_p}{C\omega} - M\omega\right)}{[r_p + \mathbf{Z}_b]\left[\mathbf{Z} - \dfrac{\dfrac{u_pM}{C} - M^2\omega^2}{r_p + \mathbf{Z}_b}\right]} \tag{356}$$

Examination of Eq. (355) shows that if the term u_pM/C is greater than $M^2\omega^2$, a condition which nearly always prevails in practice, the effect of regeneration is to subtract a quantity from R and add a quantity to the reactance X. In general, therefore, *regeneration reduces the effective resistance of the oscillatory circuit and increases the positive reactance of the same circuit.*

The coefficient of $(r_p + R_b)$ and this same coefficient of X_b [see Eq. (355)] may be defined as the *coefficient of regeneration.* This quantity is dimensionless and we shall denote it by h. In this particular case the value of h is given by

$$h = \frac{\dfrac{u_pM}{C} - M^2\omega^2}{(r_p + R_b)^2 + X_b^2} \tag{357}$$

The factor h is largely dependent upon M but is also a function of C and of the frequency $\omega/2\pi$ of the impressed e.m.f. The value of h increases to a maximum, then decreases to zero, and

crosses the axis of abscissas to negative values as M is increased. For any given value of M, an increase of frequency decreases h.

Using the symbol h, Eq. (355) becomes

$$\Delta \mathbf{I} = \frac{\Delta \mathbf{E}_0}{\left[R - h(r_p + R_b)\right] + j\left[(L + hL_b)\omega - \frac{1}{C\omega}\right]} \tag{358}$$

When a regenerative circuit like that shown in Fig. 174 is used in practice, the capacitance C is usually varied to make the reactance of the circuit zero for the impressed frequency. The current ΔI then has its maximum value given by

$$\Delta I_{\text{max.}} = \frac{\Delta E_0}{R - h(r_p + R_b)} \tag{359}$$

If h is now increased by increasing M, the current $\Delta I_{\text{max.}}$ and also the current ΔI_p increase. As h is increased, the circuit is thrown slightly out of resonance with the impressed e.m.f. so that C must be readjusted to maintain zero reactance and hence maximum current. According to Eq. (359), if h is made sufficiently large, the denominator becomes zero and the current infinite. Of course, infinite current is never attained in practice because, as the current increases owing to increasing h, the electrical variations exceed the small values assumed as necessary conditions in this treatment, and the theory no longer holds. In other words, the quantities u_p and r_p, assumed constant for small variations, are not constant as the amplitudes of variations increase, but vary over the cycle. The mean values always change in such a direction as to decrease the value of h and prevent the current from approaching the infinite value indicated by the simple theory. Nevertheless, as h is increased and the denominator of Eq. (359) approaches zero, the current $\Delta I_{\text{max.}}$ does increase to large values and the smaller ΔE_0 is the smaller can the denominator of Eq. (359) be made and still conform to the original assumptions of the theory.

The extent to which the effective resistance of an oscillatory circuit can be reduced by regeneration depends greatly upon the constancy of the batteries operating the triode, upon the delicacy possible in the adjustment of M and C, and upon the constancy of temperature of the oscillatory circuit. In one experimental example, the resistance of the coil was normally about 300 ohms, but the effective resistance could be reduced to less than 0.001

ohm by delicately controlled regeneration. That a small variation of temperature has a great effect on the operation of a triode system in which regeneration is pushed to an extreme can be shown by the above example. A 1°C. change in temperature would cause the resistance of the copper coil to vary by about 1.2 ohms. Hence, a thousandth of a degree change would either double the effective resistance of the circuit or reduce it to zero, according as the change of temperature is positive or negative.

Suppose that h, in Eq. (358), approaches the value which makes the equivalent resistance of the circuit equal to zero, and that C is varied so that the reactance is always zero for the impressed frequency. If ΔE_0 were constant, the current ΔI would increase. Suppose, however, that, as the equivalent resistance approaches zero, the value of ΔE_0 is also made to approach zero in such a way that the current remains finite. When the equivalent resistance becomes zero, the current can be finite even if ΔE_0 becomes zero. The system then oscillates by itself at a frequency which makes the equivalent reactance of the oscillatory circuit zero.

The frequency of self-oscillation, when oscillation just begins, is obtained by equating to zero the reactance term of Eq. (358) and at the same time inserting the value of h obtained by equating to zero the resistance term of Eq. (358). The result is

$$\omega^2 = \frac{1}{LC\left(1 + \dfrac{R}{r_p + R_b} \cdot \dfrac{L_b}{L}\right)} = \frac{1}{LC\left(1 + \dfrac{\eta}{\eta_b}\right)} \tag{360}$$

where $\eta = \dfrac{R}{L\omega}$ and $\eta_b = \dfrac{r_p + R_b}{L_b\omega}$. The fraction $\dfrac{\eta}{\eta_b}$ is usually very small so that the frequency of *self-oscillation* is only very slightly *less* than the natural frequency of *forced oscillation* given by $\omega_0^2 = 1/LC$. If the amplitude of self-oscillation builds up so that the original hypothesis of small amplitude is violated, the frequency of self-oscillation is not accurately given by Eq. (360). It is apparent that there is a gradual and smooth transition from regeneration to oscillation.

The very great increase in ΔI and hence in ΔI_p can be better understood by making a plot of ΔI as $M\omega$ is varied for a

particular case. Assume that u_p is 10, r_p is 20,000 ohms, R is 5 ohms, ω is 10^7 radians per second, C is 0.0002μf, L_b is 10^{-5} henry, and R_b is negligible in comparison with r_p. The two values of $M\omega$ which make the equivalent resistance of the oscillatory circuit zero are found by placing the denominator of Eq. (359) equal to zero.

Hence,

$$R = h(r_p + R_b) = \frac{\frac{u_p M}{C} - M^2\omega^2}{(r_p + R_b)^2 + L_b^2\omega^2} \cdot (r_p + R_b) \quad (361)$$

Solving for $M\omega$,

$$M\omega = \frac{u_p}{2C\omega} \pm \sqrt{\left(\frac{u_p}{2C\omega}\right)^2 - \frac{R}{r_p + R_b}\left[(r_p + R_b)^2 + L_b^2\omega^2\right]} \quad (362)$$

The numerical values of $M\omega$ for the constants of the circuit as given are 20 ohms and 4,980 ohms. It is impossible to obtain the latter value because the value of $M\omega$ corresponding to perfect coupling between L_b and L is only 224 ohms. In this example the value of $M^2\omega^2$ is negligible in comparison with u_pM/C, so that the critical value of $M\omega$ can be obtained from Eq. (361) by neglecting the $M^2\omega^2$ term. Thus

$$M\omega = \frac{RC\omega}{u_p} \cdot \frac{(r_p + R_b)^2 + L_b^2\omega^2}{r_p + R_b} \quad (363)$$

Figure 176 gives a plot of h and of $\Delta I/\Delta E_0$ against $M\omega$ for the constants as assumed. If $M\omega$ is set at 19.5 ohms, the ratio $\Delta I/\Delta E_0$ is forty times as great as with no regeneration. If the impressed voltage ΔE_0 is 2.5 microvolts and $M\omega$ is set at 19.5 ohms the grid voltage is about 0.01 volt. As this value is surely sufficiently small to be within the limitations of the theory, the full forty times amplification due to regeneration can be obtained. If the impressed voltage is ten or one hundred times 2.5 microvolts, the electrical variations are too great, the theory does not apply exactly, and full amplification is not obtained.

In the example given, the change in the equivalent inductive reactance due to regeneration is very slight. With no regeneration, the reactance is 500 ohms. Full regeneration increases this reactance by only 0.025 ohm.

A second example is given to illustrate the case when $M^2\omega^2$ is not negligible, and when there are consequently two possible value of $M\omega$ which give zero equivalent resistance. Let the con-

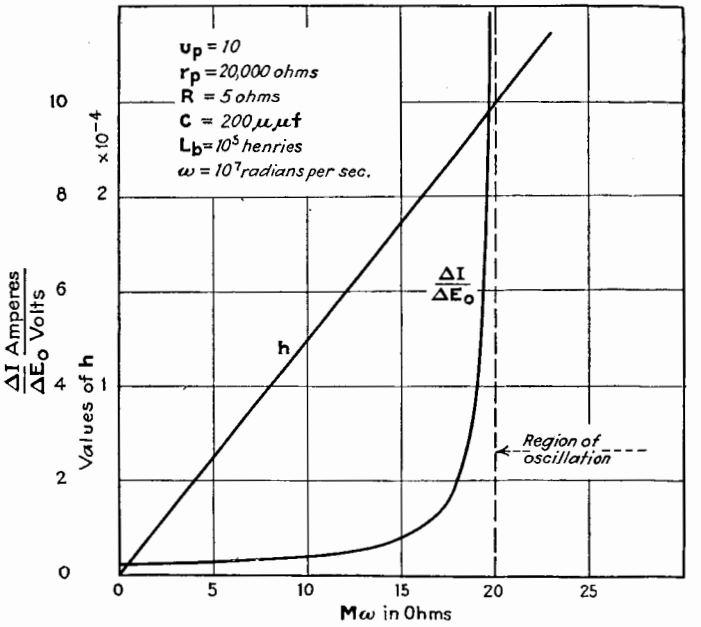


FIG. 176.—Variation of resonance current with feed-back coupling.

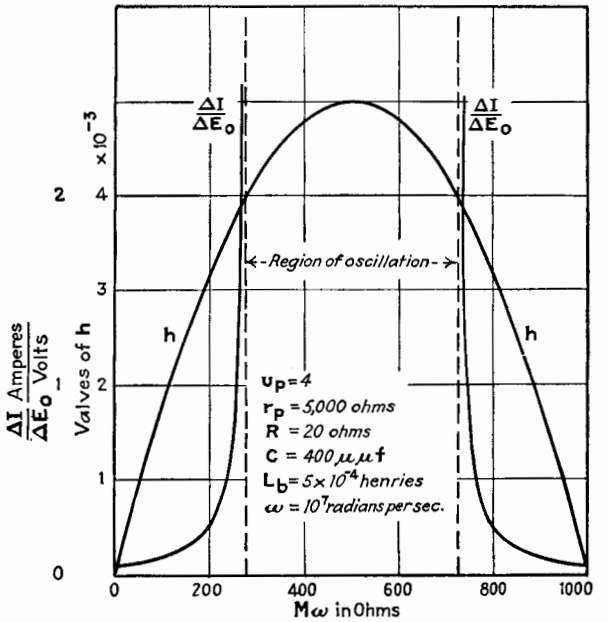


FIG. 177.—Variation of resonance current with feed-back coupling.

stants of the system be: $u_p = 4$, $r_p = 5{,}000$ ohms, $R = 20$ ohms, $\omega = 10^7$ radians per second, $C = 0.0004\mu$f, $L_b = 5 \times 10^{-4}$ henry, R_b is negligible. Equation (262) must be used to obtain the values of $M\omega$ which give zero equivalent resistance. These values of $M\omega$ are 276 ohms and 724 ohms. They are both possible because the value of $M\omega$ for perfect coupling is 1,120 ohms. A plot of h and $\Delta I/\Delta E_0$ both against $M\omega$ for this case is shown in Fig. 177.

The essential points of the theory just given have been checked experimentally by Jolliffe and Rodman.[2]

An interesting condition may exist when a regenerative triode is used. Suppose that h is adjusted to be near the point of oscillation for a certain frequency for which the reactance is zero. If the frequency of the impressed e.m.f. is reduced, h is thereby increased, as shown by Eq. (357). The equivalent resistance may become negative. We then have a circuit with capacitive reactance and an equivalent negative resistance. The circuit, instead of absorbing power, delivers power to ΔE_0.

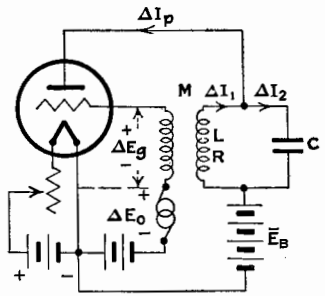


FIG. 178.—Circuits of a regenerative triode with a tuned plate load.

132. Regeneration with Inductive Coupling. Tuned Plate Circuit.—A second type of regenerative circuit is illustrated in Fig. 178. The equations for this case are

$$\left.\begin{aligned} (R + jL\omega)\Delta\mathbf{I}_1 - \frac{j\Delta\mathbf{I}_2}{C\omega} &= 0 \\ (R + jL\omega)\Delta\mathbf{I}_1 + r_p\Delta\mathbf{I}_p &= u_p\Delta\mathbf{E}_g \\ \Delta\mathbf{E}_g - jM\omega\Delta\mathbf{I}_1 &= \Delta\mathbf{E}_0 \\ \Delta\mathbf{I}_1 &= \Delta\mathbf{I}_p + \Delta\mathbf{I}_2 \end{aligned}\right\} \tag{364}$$

The solution of these equations is

$$\Delta\mathbf{I}_1 = \frac{-ju_p\Delta\mathbf{E}_0/r_pC\omega}{\left[R + \frac{L}{Cr_p} - \frac{u_pM}{Cr_p}\right] + j\left[L\omega - \frac{1}{C\omega}\left(1 + \frac{R}{r_p}\right)\right]} \tag{365}$$

If the e.m.f. ΔE_0 is introduced in coil L or in series with L, instead of in the grid circuit, the current $\Delta\mathbf{I}_1$ is

[2] JOLLIFFE and RODMAN, *Sci. Paper Bur. Standards* 487, April, 1924.

$$\Delta \mathbf{I}_1 = \frac{\Delta \mathbf{E}_0}{\left[R + \frac{L}{Cr_p} - \frac{u_p M}{Cr_p}\right] + j\left[L\omega - \frac{1}{C\omega}\left(1 + \frac{R}{r_p}\right)\right]} \tag{366}$$

In both Eqs. (365) and (366), the denominator represents the equivalent impedance of the oscillatory circuit. The resistance of the oscillatory circuit is increased by the term L/Cr_p, because the internal plate circuit of resistance r_p is in parallel with this oscillatory circuit. The coupling M has the effect of reducing the equivalent resistance of the circuit by an amount u_pM/Cr_p, and it is important to note that this term and therefore the resistance-reducing effect are *independent of frequency*. The coefficient of regeneration in this case is

$$h = \frac{u_p M}{Cr_p^2} \tag{367}$$

The equivalent resistance of the circuit can be written $R + \frac{L}{Cr_p} - hr_p$. Regeneration does not affect the reactance of the oscillatory circuit, so that an adjustment of M does not affect the tuning as it did in the previous case.

Equations (365) and (366) give the current in the inductance of the oscillatory circuit when the e.m.f. ΔE_0 is introduced in the grid circuit or directly in the oscillatory circuit, and it is natural to expect that a larger current would result in the first case, because the amplifying effect of the tube is used. However, this is not always true, as is easily seen if the two equations are compared. The introduction of ΔE_0 in the grid circuit gives a larger current only if $u_p/r_pC\omega > 1$, *i.e.*, if $s_p/C\omega > 1$. We may assume that for a receiving tube s_p is of the order of 500 micromhos. If C is 0.0001 μf and ω is 10^{-7} radian per second, $s_p/C\omega = 0.5$ and a larger current would be obtained by introducing the e.m.f. ΔE_0 directly into the oscillatory circuit. The triode would then serve merely as a resistance-reducing device. In making the comparison of the two ways of introducing the e.m.f. it should be remembered that no current is drawn from the e.m.f. ΔE_0 in the first case, shown in Fig. 178, whereas current and power are demanded from the source ΔE_0 if it is introduced into the oscillatory circuit.

If the equivalent resistance of the circuit shown in Fig. 178 is reduced to zero, self-oscillation starts at a frequency which makes

the equivalent reactance of the circuit equal to zero. Therefore, referring to either Eq. (365) or Eq. (366), the frequency at which oscillation starts is given by the relation

$$\omega^2 = \frac{1 + \dfrac{R}{r_p}}{LC} = \omega_0^2\left(1 + \frac{R}{r_p}\right) \tag{368}$$

where ω_0 is the natural angular velocity of the circuit, or $1/\sqrt{LC}$. If the self-oscillations build up to an amplitude which goes beyond the limitations of the theory to small values, Eq. (368) is more and more departed from.

In all of the cases of regeneration just discussed, the effect of the capacitance C_{pg} between grid and plate of the triode has been neglected. This capacitance provides an active coupling path between the grid and plate circuits especially at high frequencies and hence is always a contributing cause of regeneration. The effect of C_{pg} in comparison with the inductive coupling is small, provided the ratio C_{pg}/C is small. The effect of C_{pg} in producing regeneration is also small, provided L_b is small when the oscillatory circuit is connected to the grid and provided L_c is small when the oscillatory circuit is connected to the plate.

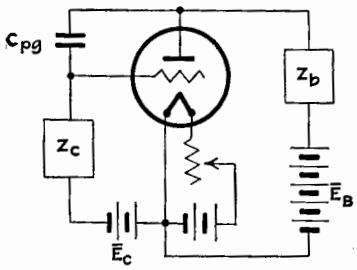


FIG. 179.—Circuits of a triode, with capacitance coupling between the grid and plate circuits.

133. Regeneration with Capacitive Coupling.—The coupling between the output or plate circuit and the input or grid circuit may be entirely capacitive, as shown in Fig. 179. The capacitance may be only the capacitance of the tube and connections but it may include the capacitance of a physical condenser connected between grid and plate. We shall denote by C_{pg} the *total* capacitance between grid and plate terminals of the triode.

In Chap. XI the equivalent input and output admittances were calculated, and it was shown that under certain conditions the input conductance g_g may be negative, in which case power is delivered to circuit Z_c by the connected system, and regeneration takes place. This regeneration is not very effective in increasing the potential across Z_c unless the current in Z_c is determined largely by the resistance of the circuit. If the reactance of Z_c

can be reduced to zero by a tuning condenser, strong regeneration can take place and the limit of regeneration when oscillation begins is determined by the equations

$$\left.\begin{aligned} g_g + G_c &= 0 \\ b_g + B_c &= 0 \end{aligned}\right\} \tag{369}$$

where $\mathbf{Z}_c = \dfrac{1}{G_c - jB_c}$, and g_g and b_g are given by Eqs. (278) and (279) of Chap. XI.

In a similar way, regeneration may take place when the oscillatory circuit is in the plate circuit and when the output conductance g_p of the triode is negative. Oscillation begins if

$$\left.\begin{aligned} g_p + G_b &= 0 \\ b_p + B_b &= 0 \end{aligned}\right\} \tag{370}$$

where $\mathbf{Z}_b = \dfrac{1}{G_b - jB_b}$, and g_p and b_p are given by Eqs. (280) and (281) of Chap. XI.

We shall now consider a few special cases of capacitive coupling.

***134. Regeneration with Capacitive Coupling. Tuned Grid Circuit.**—If the oscillatory circuit is connected to the grid of the triode, the curves of Fig. 153, page 278, show that g_g is negative only if X_b is positive. Let the plate load have an equivalent inductance L_b and resistance R_b. Figure 180 illustrates this case, which was described by Armstrong[3] in 1915. The plate inductance was a variometer, and Armstrong gave a curve, reproduced in Fig. 181, which shows the increase in signal strength as the inductance L_b is varied.

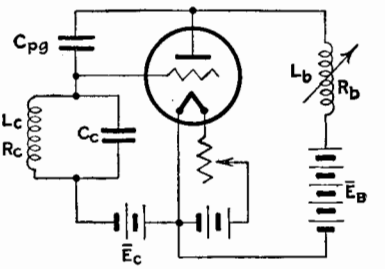


FIG. 180.—Circuits of a regenerative triode having a variable inductance as plate load.

We can treat this case theoretically by substituting in Eq. (369) the value of g_g from Eq. (278). Neglecting R_c^2 in comparison with $L_c^2\omega^2$, and R_b^2 in comparison with $L_b^2\omega^2$, the condition for maximum regeneration bordering on self-oscillation is

[3] ARMSTRONG, *Proc. I.R.E.*, **3**, 220 (1915).

$$[G_c] = G_c + g_g = \frac{R_c}{L_c^2\omega^2}$$

$$+ C_{pg}^2\omega^2 r_p \cdot \frac{1 + \frac{R_b r_p}{L_b^2\omega^2} + u_p\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{L_b C_{pg}\omega^2}\right)}{\left(1 + \frac{R_b r_p}{L_b^2\omega^2}\right)^2 + C_{pg}^2\omega^2 r_p^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{L_b C_{pg}\omega^2}\right)^2} = 0 \quad (371)$$

If the last term in the numerator of Eq. (371) is not sufficient to reduce the equivalent conductance $[G_c]$ to zero, regeneration still exists if g_g is less than zero.

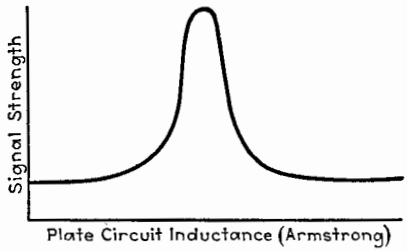


FIG. 181.—Variation of signal strength with plate-circuit inductance for the regenerative triode of Fig. 180. (*Armstrong.*)

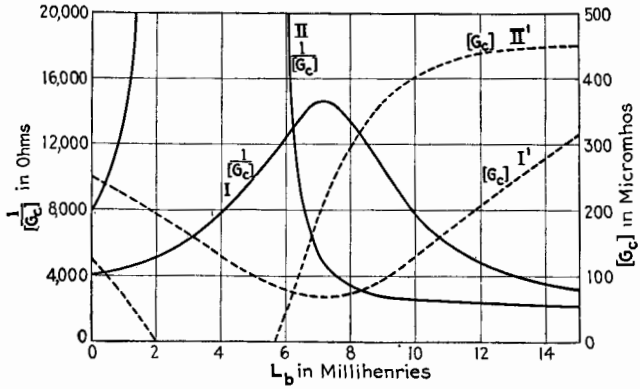


FIG. 182.—Variation of the equivalent conductance of the grid circuit of a regenerative triode with plate-circuit inductance.

If the reciprocal of the equivalent conductance $[G_c]$ is plotted against L_b for values of the constants which do not make $[G_c] = 0$, a curve similar to that of Fig. 181 is obtained. As an example $1/[G_c]$ is plotted in curve I of Fig. 182 for the following values of the constants: $u_p = 8$, $r_p = 16{,}000$ ohms, $\omega = 10^6$ radians per sec., $C_{pg} = 50\mu\mu f$, $C_{pf} = 20\mu\mu f$, $R_c = 40$ ohms, and $L_c = 400$

microhenries. With these values of the constants, the value of $A = C_{pg}\omega r_p$ is 0.8. Since the current in the oscillatory circuit is proportional to $1/[G_c]$, the curve I shows that maximum regeneration occurs for a value of L_b equal to 7.1 millihenries, and regeneration gives an increase of 3.64 times in the current. The dotted curve I′ is a plot of $[G_c]$ for the example just given. If the angular velocity of the impressed e.m.f. is $\sqrt{2}\cdot 10^6$ radians per second, the equivalent conductance $[G_c]$ is shown in the dotted curve II′ and becomes zero at two values of L_b. Between these two values of L_b, the system oscillates. The shape of the graph of $1/[G_c]$ is shown in curve II.

Referring to Eq. (371), the term $R_b r_p/L_b^2\omega^2$ is generally small in comparison with unity and, by neglecting it, an approximate expression is obtained which can be analyzed more easily. If we equate to zero the derivative with respect to L_b of the last term of Eq. (371), we find the value of $L_b\omega$ which gives a minimum $[G_c]$ and hence a maximum regeneration. The result is expressed in the following equation:

$$\frac{1}{L_b\omega} = C_{pf}\omega + C_{pg}\omega\left(1 + \frac{1}{u_p} + \frac{1}{u_p}\sqrt{1 + \frac{u_p^2}{A^2}}\right) \tag{372}$$

Substituting this value of $L_b\omega$ in Eq. (371) gives as the minimum value of $[G_c]$

$$[G_c]_{\text{min.}} = \frac{R_c}{L_c^2\omega^2} - \frac{s_p u_p}{2\left(1 + \sqrt{1 + \frac{u_p^2}{A^2}}\right)} \tag{373}$$

As the plate inductance L_b is varied to adjust for maximum regeneration, the equivalent input susceptance changes and necessitates a resetting of the capacitance C_c to maintain resonance with the impressed e.m.f. ΔE_0. The effective addition to C_c due to the triode and its plate load is given by Eq. (279), Chap. XI, and is

$$\Delta C_c = C_{gf} + C_{pg}\left[1 + \frac{u_p\left(1 + \frac{R_b r_p}{L_b^2\omega^2}\right) - A^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{C_{pg}L_b\omega^2}\right)}{\left(1 + \frac{R_b r_p}{L_b^2\omega^2}\right)^2 + A^2\left(1 + \frac{C_{pf}}{C_{pg}} - \frac{1}{C_{pg}L_b\omega^2}\right)^2}\right] \tag{374}$$

where $A = C_{pg}\omega r_p$, R_c^2 has been neglected in comparison with $L_c^2\omega^2$, and R_b^2 has been neglected in comparison with $L_b^2\omega^2$. We

may neglect the term $R_b r_p / L_b^2 \omega^2$ in comparison with unity, as in Eq. (371), and substitute in Eq. (374) that value of $L_b\omega$ from Eq. (372) which gives maximum regeneration. We can then express the fractional change in C_c required when L_b is varied from zero to the value for maximum regeneration. The result is

$$\text{Fractional change in } C_c = \frac{u_p C_{pg}}{2\left(\frac{1}{L_c\omega^2} - C_{pg} - C_{pf}\right)} \tag{375}$$

For the values of the constants which apply to curve I of Fig. 182, $\Delta C_c / C_c$ is 0.0823. By Eq. (374) the total added capacitance due to the triode at maximum regeneration is

$$\Delta C_c = C_{gf} + C_{pg}\left(1 + \frac{u_p}{2}\right) \tag{376}$$

If conditions are such that for a certain range of L_b the system oscillates, maximum regeneration takes place for a value of L_b very slightly less than the smaller value of L_b at which oscillation starts, or very slightly more than the larger value of L_b at which oscillation stops. The two values of $L_b\omega$ at which oscillation begins can be calculated from Eq. (371) and are

$$\frac{1}{L_b\omega} = (C_{pf} + C_{pg})\omega + \frac{s_p C_{pg} L_c^2 \omega^3}{2R_c}\left[1 \pm \sqrt{1 - \frac{4R_c}{u_p^2 C_{pg}^2 L_c^2 \omega^4}\left(\frac{R_c}{L_c^2\omega^2} + C_{pg}^2\omega^2 r_p\right)}\right] \tag{377}$$

As before, the term $R_b r_p / L_b^2 \omega^2$ is neglected in comparison with unity. The capacitance added to the condenser C_c by the triode which is set at either border line of oscillation is obtained by substituting Eq. (377) in Eq. (374) and is

$$\Delta C_c = C_{gf} + C_{pg}\left[1 - u_p \frac{\frac{2R_c}{L_c^2 C_{pg}^2 \omega^4 r_p} + 1 \pm B}{2 - \frac{u_p^2 L_c^2 \omega^2}{R_c r_p}(1 \pm B)}\right] \tag{378}$$

where B stands for the radical of Eq. (377). The negative sign corresponds to the smaller value of L_b. Filling in the values for the constants corresponding to curve II of Fig. 182, the value of ΔC_c for the smaller value of L_b is $C_{gf} + 1.24\ C_{pg}$. This is much

less than the added capacitance, given by Eq. (376), for the system adjusted to give maximum regeneration without oscillation. Thus we see that if regeneration of this type is to be utilized, it is better to choose the circuits so that the system is adjusted to the border line between oscillation and non-oscillation. The added capacitance at the larger value of L_b is much less than that for the smaller L_b and is $C_{gf} + 1.034\, C_{pg}$ for the values of the constants for curve II of Fig. 182. The fractional change in C_c due to varying L_b is very small in this case.

The coil L_b always has distributed capacitance which should be added to C_{pf} in Eqs. (371) and (374). Similarly, the capacitance C_{gf} should be considered as a part of C_c. The discussion applies if a condenser is connected across the plate inductance L_b. Regeneration in the grid circuit is possible only when the reactance of the plate circuit is positive or, in other words, when the plate circuit is resonant to a *higher* frequency than that of the impressed e.m.f.

***135. Regeneration with Capacitive Coupling. Tuned Plate Circuit.**—When the plate circuit is tuned to the impressed frequency, regeneration takes place in that circuit if the output conductance g_p is negative, where g_p is given by Eq. (280), page 279, Chap. XI. Extreme regeneration in the plate circuit, to the point of self-oscillation, is expressed by Eq. (370). We can analyze this case in a manner similar to that used in the preceding section. Neglecting, as before, the squares of the resistances in comparison with the squares of the inductive reactances, the equivalent conductance of the plate load is, from Eq. (280),

$$[G_b] = \frac{R_b}{L_b^2\omega^2} + \frac{1}{r_p} + C_{pg}^2\omega^2 r_p \cdot \frac{\dfrac{R_c r_p}{L_c^2\omega^2} + u_p\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{L_c C_{pg}\omega^2}\right)}{\left(\dfrac{R_c r_p}{L_c^2\omega^2}\right)^2 + C_{pg}^2\omega^2 r_p^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{L_c C_{pg}\omega^2}\right)^2} \qquad (379)$$

In Eq. (379), any capacitance in parallel with the grid coil L_c is considered a part of C_{gf}, and the capacitance in parallel with L_b is assumed tuned to give maximum current in the oscillatory circuit.

The effective capacitance added to C_b by the triode system is given by Eq. (281), page 279, Chap. XI, and is

$$\Delta C_b =$$

$$C_{pf} + C_{pg}\left[1 + \frac{\dfrac{u_p R_c r_p}{L_c^2\omega^2} - C_{pg}^2\omega^2 r_p^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{L_c C_{pg}\omega^2}\right)}{\left(\dfrac{R_c r_p}{L_c^2\omega^2}\right)^2 + C_{pg}^2\omega^2 r_p^2\left(1 + \dfrac{C_{gf}}{C_{pg}} - \dfrac{1}{L_c C_{pg}\omega^2}\right)^2}\right] \tag{380}$$

The conditions for maximum regeneration, and the boundary between oscillation and non-oscillation, can be found in a manner similar to that used in Sec. 134. The analysis is not given here because in practice this particular case is of less importance than the preceding one.

Equations (379) and (380) also hold if the coil L_c is shunted by a capacitance (treated as C_{gf}). Regeneration in the plate circuit is possible only when the susceptance of the grid circuit is positive, that is to say, when the natural frequency of the grid circuit is *higher* than that of the impressed e.m.f. ΔE_0 to which the plate circuit is tuned.

136. The "Ultra-audion" Regenerative Circuit.—The circuit invented by de Forest and known as the "ultra-audion" circuit is

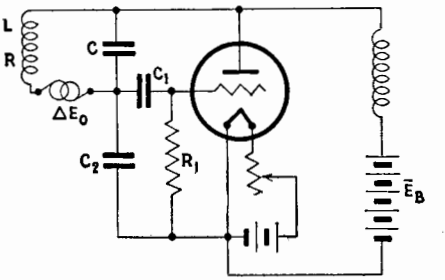


FIG. 183.—The "ultra-audion" circuits.

another form of regenerative circuit. The connections are shown in Fig. 183. Regeneration is varied by changing condenser C_2. Capacitance C_1 is only for the purpose of preventing the steady plate voltage from acting upon the grid. This was one of the first regenerative circuits used but is seldom found in present-day practice. The theoretical treatment of this circuit is left to the reader.

137. Regeneration with Resistance Coupling and with Negative-resistance Devices.—There is no known practical method of obtaining regeneration with a single high-vacuum triode using *resistance coupling*. It is possible, however, to obtain regenera-

tion with resistance coupling if more than one triode is used. Since a discussion of combinations of tubes forming multistage amplifiers is to be given in a later chapter, the discussion of resistance regeneration using two tubes will be presented at that time.

Before leaving this part of the discussion of regeneration, mention should be made of the regenerative effect produced by connecting across an oscillatory circuit a triode having a negative k_g or a negative k_p. A negative k_g is possible if the triode contains gas, as has been pointed out in Chap. X, or if secondary emission takes place at the grid, so that the grid-current curves plotted against grid voltage have a negative slope. A negative k_p is obtained when considerable secondary electron emission exists at the plate of a triode, as in the device known as the *dynatron*. The dynatron will be discussed in Chap. XXIII.

II. REGENERATION FOR LARGE AMPLITUDES

Part I is limited in its application to problems in which the electrical variations are very small. This limitation was imposed because the triode, which forms a part of the circuit, has a curved characteristic, and its internal resistance r_p and voltage ratio u_p vary over the characteristic curve. If only sufficiently small variations are considered, a very fair approximation to a true solution is obtained if these tube parameters are considered constant. With this limitation, the e-p-c. theorem can be applied and the system is reducible to a simple-circuit network possessing constant elements. The smaller the electrical variations, the more justifiable is the theory and the more accurate the results. No definite limiting magnitude of electrical variations can be stated below which the theory of Part I is applicable, and above which a new form of treatment must be used. As the electrical variations are increased, the error in the results of Part I increases. The error for any given magnitude of electrical variation depends upon the point on the characteristic surface of the triode about which the triode is operating. The greater the curvature of the path of operation, the greater is the error for a given amplitude of electrical variation, or the weaker must be the electrical variation if the error is to be small. On the other hand, if the path of operation falls upon a nearly plane portion of the characteristic surface, the electrical variations may be comparatively

large and still the solution of the problem as given in Part I is sufficiently close for practical purposes.

The magnification of the grid voltage of a regenerative system, when great regenerative effect is used, is usually so great that even if the impressed voltage is very small the theory of the first part is inapplicable. The prediction from the simple theory that infinite current should result at the limit of regeneration is untrue, because the simple theory never holds at the limit of regeneration unless the impressed signal voltage is vanishingly small.

138. Theory of the Method.—When the electrical variations are large, an approximate idea of the regenerative effect can be had from the following graphical treatment of the problem. The

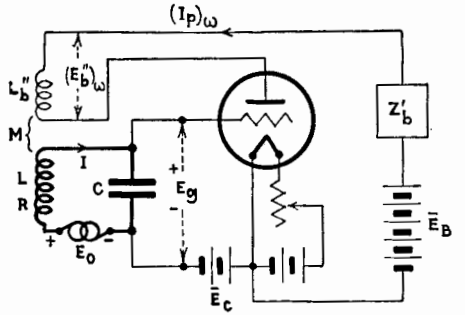


FIG. 184.—Circuits of a triode with regenerative coupling between the grid and plate circuits.

graphical analysis is based upon certain experimental data on the characteristic curves of the triode when subjected to large electrical variations.

Figure 184 shows the circuit to which the treatment applies. Although magnetic coupling is shown, the treatment may be extended to other forms of coupling. The constants of the circuit are represented on the diagram and the electrical quantities are also indicated. The total impedance Z_b in the plate circuit comprises the coupling coil L_b'' and some sort of load Z_b'.

The principle of the method may be stated as follows: The heavy lines of Fig. 184 represent an oscillatory circuit to which is attached a system, shown in light lines, which supplies energy to the oscillatory circuit through the mutual inductance M. The energy impulses are controlled by the voltage E_g and are so directed as to aid the impressed e.m.f. E_0 to produce a current I in the oscillatory circuit. The boosting e.m.f. induced in L from

L_b'' is denoted by $(E_0')_\omega$. The object of this treatment is first to find $(E_0')_\omega$ in terms of E_g, and next to find the relation between I and E_0.

Since the electrical variations are large and the characteristics of the triode are curved, the plate current I_p is nonsinusoidal, although, because of the low resistance of the oscillatory circuit, the current I and the voltage E_g may be assumed to be practically

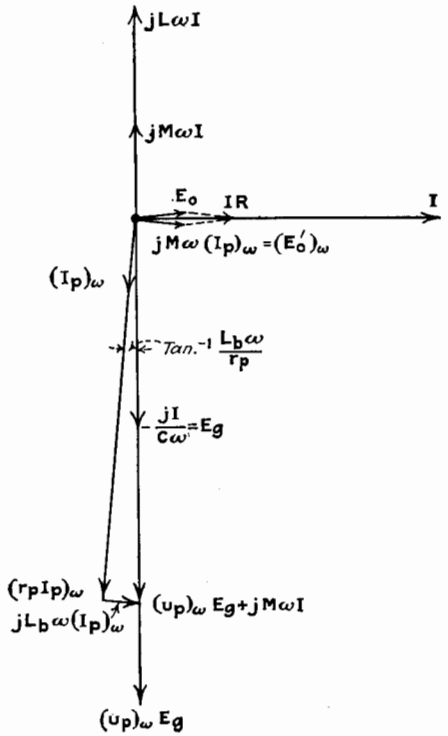


FIG. 185.—Vector diagram for the regenerative triode shown in Fig. 184.

sinusoidal. Hence, the voltage $(E_0')_\omega$ is induced by the fundamental component of I_p, denoted by $(I_p)_\omega$. We can now set up the equations for the system of Fig. 184 which hold for the fundamental components. They are as follows:

$$\left.\begin{aligned} R\mathbf{I} + j\left(L\omega - \frac{1}{C\omega}\right)\mathbf{I} &= \mathbf{E}_0 + jM\omega(\mathbf{I}_p)_\omega \\ (r_p\mathbf{I}_p)_\omega + jL_b\omega(\mathbf{I}_p)_\omega &= (u_p)_\omega\mathbf{E}_g + jM\omega\mathbf{I} \\ \mathbf{E}_g &= -\frac{j\mathbf{I}}{C\omega} \end{aligned}\right\} \qquad (381)$$

In Eq. (381), $(r_p)_\omega$ and $(u_p)_\omega$ denote quantities analogous to the usual r_p and u_p but evaluated for large variations and for fundamental components of the plate current.

The equations are represented in vector form in Fig. 185 in order to give a picture of the phase relation of the various quantities. As appears from the vector diagram, the boosting e.m.f. $(E_0')_\omega$ is nearly in phase with E_0 and differs in phase from it only because of reactance in the plate circuit. Assume that the total reactance X_b is small in comparison with r_p. Hence, it can be assumed without much error that the total effective e.m.f. producing I is numerically equal to the sum of E_0 and $(E_0')_\omega$.

For any given mutual inductance M, the voltage $(E_0')_\omega$ is given by the relation

$$(\mathbf{E}_0')_\omega = jM\omega(I_p)_\omega = jM\omega f(E_g) \tag{382}$$

Therefore, $(E_0')_\omega$ is proportional to $(I_p)_\omega$ and to $M\omega$. The law connecting $(E_0')_\omega$ and E_g can be obtained only by experiment. This law is the important and necessary relation which describes the regenerative action of the triode and its circuits for the particular plate load Z_b', coupling coil L_b'', and inductive reactance $M\omega$. This law connecting $(I_p)_\omega$ and E_g is to some extent a function of the frequency, because of reactance in the plate circuit of the triode. Frequently the system shown in Fig. 184 is used as a detector, in which case Z_b' is practically zero and the only reactance is the tickler coil L_b''.

In Fig. 186, the full-line curve marked $(E_0')_\omega$ is assumed as giving the relation connecting $(E_0')_\omega$ and E_g for a certain value of $M\omega$ denoted by $\mathfrak{M}\omega$. Since $(E_0')_\omega$ is directly proportional to $M\omega$, the ordinates of the curve of Fig. 186 are increased or decreased in proportion to the increase or decrease in $M\omega$, as shown by the dotted curves in Fig. 186.

If the mutual inductive reactance $M\omega$ were zero, the resonance current in the oscillatory circuit would be E_0/R, and the relation between the voltage E_0 and the voltage across the condenser, *i.e.*, the grid voltage, would be

$$\mathbf{E}_0 = jC\omega R\mathbf{E}_g \tag{383}$$

A plot of E_o *vs.* E_g from Eq. (383) is the straight line of Fig. 186, which is drawn making with the horizontal an angle whose tangent is $C\omega R$.

In Fig. 186, the ordinate to the straight line corresponding to any voltage E_g represents the total voltage necessary to produce a current which, when flowing through the condenser C, gives rise to E_g. The ordinate to the curve of $(E_0')_\omega$ represents the regenerative voltage introduced into the oscillatory circuit. The difference between these ordinates is the impressed voltage E_0 required to produce E_g with the particular amount of regeneration corresponding to the particular value of $M\omega$. A plot of E_g against E_0, shown in Fig. 187, is derived from Fig. 186

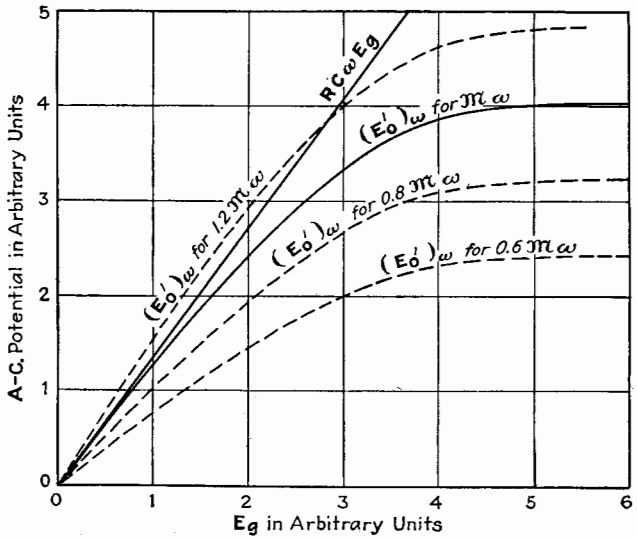


FIG. 186.—Curves of $C\omega RE_g$ and $(E_0')_\omega$ *vs.* E_g for large-signal regeneration

and shows the way in which E_g, which is proportional to I, varies with the impressed voltage. The plot of $(E_0')_\omega$ against the impressed voltage E_0 shows the way in which $(I_p)_\omega$ (see Eq. (382)) varies with the impressed voltage E_0. Such curves are shown in Fig. 187 as dotted lines.

In Fig. 186, if $M\omega$ has the particular value $\mathfrak{M}\omega$ which makes the curve for $(E_0')_\omega$ tangent to the straight line $C\omega RE_g$ at the origin, the regeneration is then adjusted to the point of oscillation for small amplitudes. The value of $\mathfrak{M}\omega$ is in this case correctly given by Eq. (362) or (363) of Part I. The theory of small electrical variations applies to a limited region near the origin of Fig. 186. If $M\omega$ is now reduced to 0.8 $\mathfrak{M}\omega$ and to 0.6$\mathfrak{M}\omega$,

the curves so marked in Figs. 186 and 187 give the corresponding results. On the other hand, if $M\omega$ is increased to $1.2\mathfrak{M}\omega$, the $(E_0')_\omega$ curve and the straight line cross. The point of intersection gives the value of E_g maintained by the continuous self-oscillation of the system. The curves of Fig. 187 marked $M\omega = 1.2\mathfrak{M}\omega$ show the way in which the oscillatory current and the plate current vary as the impressed voltage in phase with $(E_0')_\omega$ is increased.

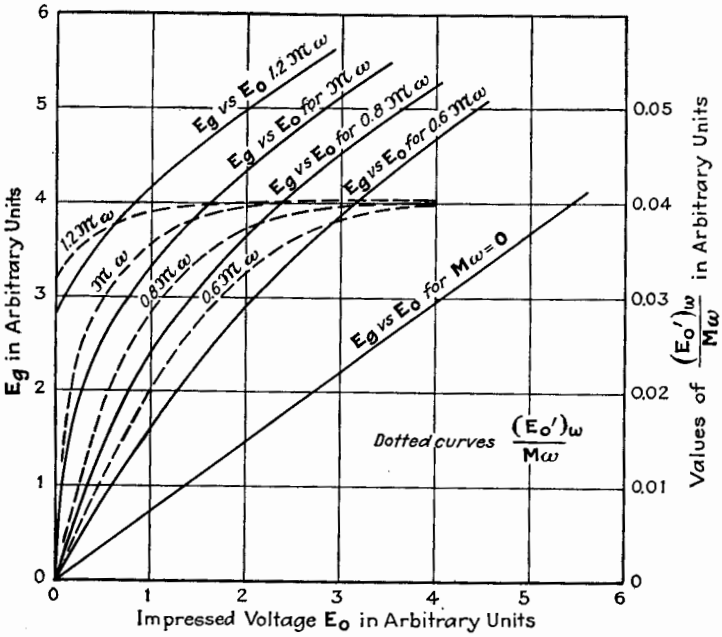


FIG. 187.—Curves of E_g and $\frac{(E_0')_\omega}{M\omega}$ *vs.* E_0 for large-signal regeneration.

The shapes of the curves for $(E_0')_\omega$, Fig. 186, depend greatly upon the conditions of use of the triode. For great regeneration the most desirable curve of E_g against E_0 is one having a long straight portion extending from the origin upward and close to the axis of ordinates. This curve is derived from a curve of $(E_0')_\omega$ *vs.* E_g, Fig. 186, which is long and straight and which is nearly coincident with the circuit line $C\omega RE_g$. If the graph of E_g against E_0 is curved, distortion necessarily results, as explained in Chap. XII. If the polarizing voltages are so chosen that the

quiescent point is on a plane portion of the characteristic surface, and if E_g remains always negative, the graph of $(E_0')_\omega$ is practically straight over a much larger extent than if the quiescent point is situated upon a curved portion of the surface.

In all of this discussion it is assumed that the grid is polarized negatively to such an extent that no grid current flows. If grid current does flow, the curve of $(E_0')_\omega$ against E_g is unchanged, but the circuit line is no longer straight. The conductance of the grid-to-filament path has the effect of increasing the resist-

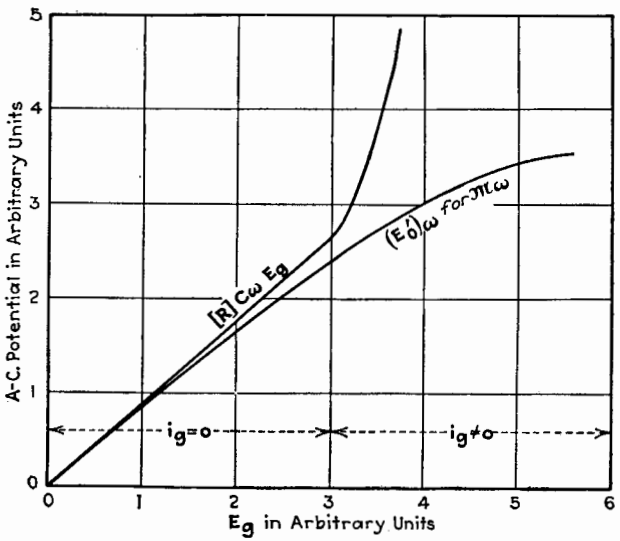


FIG. 188.—The circuit line when grid current flows.

ance R of the circuit, making the total equivalent resistance $[R]$ a function of the amplitude of E_g, and causing the circuit line to curve upward, as shown in Fig. 188. This curvature of the circuit line accentuates the flattening out of the resultant curves of E_g against E_0 and of $(E_0')_\omega/M\omega$ against E_0.

The way in which the system goes into oscillation as M is varied gives a good idea of the shape of the curve of $(E_0')_\omega$ against E_g. Referring to Fig. 186, the intersection point of the circuit line and the curve of $(E_0')_\omega$ gives the value of E_g sustained by the self-oscillation of the system. As M is increased above the critical value $\mathfrak{M}$, the intersection point travels up the circuit line, and it is then possible to plot the values of E_g corresponding

to oscillation against M or against $M\omega$. If the graph of $(E_0')_\omega$ against E_g is nearly straight, the oscillatory current, and hence E_g, increase very rapidly as M is increased. But if the characteristic graph of $(E_0')_\omega$ against E_g is curved, the increase of E_g due to oscillation is slow with respect to M.

The increase in E_g as M is varied can be easily tested experimentally and gives a rough method of determining the best polarizing potentials to obtain large regenerative effect. Curves A, B, C, and D of Fig. 189 are experimentally determined curves

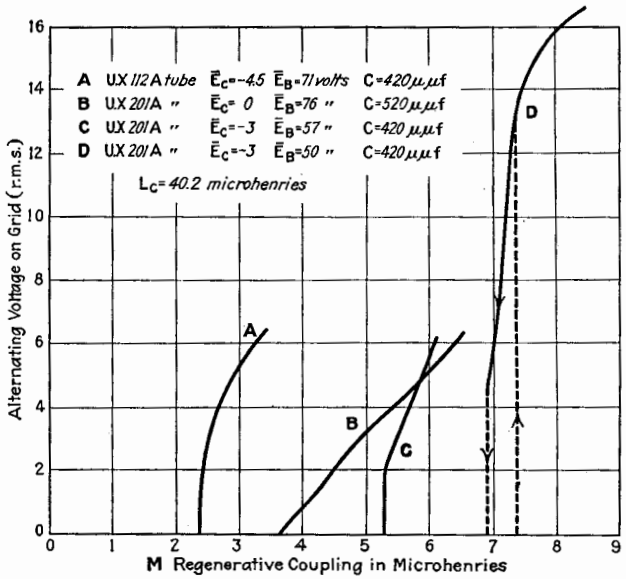


FIG. 189.—Variation of the voltage across the oscillatory circuit of Fig. 184 as the regenerative coupling is varied.

of the onset of oscillation under the various conditions stated on the figure. Curves A and C show a rapid increase in oscillatory current and hence indicate conditions for large regenerative effect. The bend in curve C is due to the current flow to the grid, which begins at a peak voltage of about 3 volts. The slowness of the increase in voltage of curve B is due to the increasing current taken by the grid.

If the grid is polarized so negatively that the plate current is nearly zero, the plate-to-filament resistance is very large. As the amplitude of E_g increases, the effective resistance of the plate-to-filament path decreases. This gives rise to curves of $(E_0')_\omega$

against E_g similar to those shown in Fig. 190. A certain circuit line $RC\omega E_g$ is assumed. As $M\omega$ is increased, the ordinates of the curve of $(E_0')_\omega$ are increased proportionally to the increase in $M\omega$. When $M\omega$ reaches the value $M_3\omega$, the $(E_0')_\omega$ curve just touches the circuit line. No oscillation starts and there is only slight regeneration for small signals. The mutual inductance must be increased to the value $\mathfrak{M}\omega$, in order to make the two curves tangent at the origin. For this critical value $\mathfrak{M}\omega$, oscillations start. As the oscillations build up, the regenerative voltage $(E_0')_\omega$ is greater than the voltage required to maintain the corresponding value of E_g, the oscillations are at first unstable

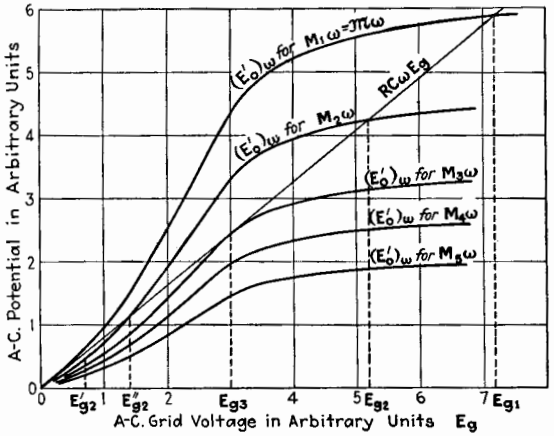


FIG. 190.—Curves of $(E_0')_\omega$ *vs.* E_g for large-signal regeneration.

and build up until the value of E_g reaches the second intersection point E_{g1}. As $M\omega$ is further increased, the oscillations increase slowly and E_g rises in a manner similar to that shown by curve D in Fig. 189. If $M\omega$ now is decreased, the oscillations persist until $M\omega$ reaches the value $M_3\omega$ and the grid voltage has decreased to E_{g3}. On further decrease in $M\omega$, the oscillations abruptly cease, as indicated by the dotted line and arrow pointing downward in Fig. 189.

The regenerative effect corresponding to a curve of the type shown in Fig. 190 is unsatisfactory for practical purposes. If $M\omega$ is set at some value $M_2\omega$ slightly less than the critical value, $\mathfrak{M}\omega$, regeneration is slight for voltages which produce grid voltages less than E_{g2}''. Any impressed voltage greater than that corresponding to E_{g2}' causes the system to break into

oscillation which persists until the mutual reactance $M\omega$ is reduced to a value less than $M_3\omega$. If the mutual reactance has

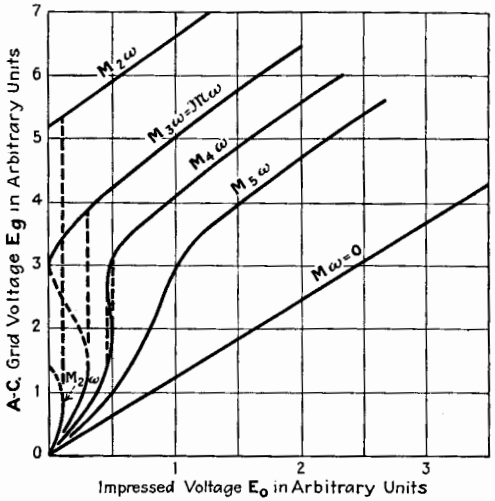


FIG. 191.—Curves of E_g *vs.* E_0 for the case of Fig. 190.

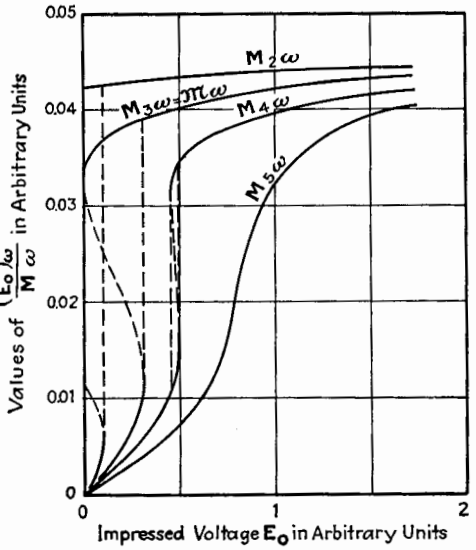


FIG. 192.—Curves of $\frac{(E_0')_\omega}{M\omega}$ *vs.* E_0 for the case of Fig. 190.

a value less than $M_3\omega$, the system cannot oscillate. The graph of regeneration is curved and discontinuities exist. The curves

of E_g and $(E_0')_\omega/M\omega$ against E_0, corresponding to the mutual inductances $M_2\omega$, $M_3\omega$, $M_4\omega$, and $M_5\omega$ of Fig. 190, are plotted in Figs. 191 and 192.

The results of the analysis just given may be briefly summarized as follows: A discontinuity in the curve of E_g against E_0 results if the circuit curve $RC\omega E_g$ and the curve of $(E_0')_\omega$ converge as E_g increases. If the two curves are parallel over a region, regeneration of weak voltages may be small, but, for voltages of a certain magnitude, the magnification is very great, as shown by the curve marked $M_5\omega$ of Figs. 191 and 192, derived from the corresponding curve of Fig. 190. This type of performance would give considerable distortion but, if a modulated radio carrier wave is being amplified, and the variations in amplitude due to modulation are small and fall within the range of voltages corresponding to the nearly vertical straight portion, distortion of the modulation might not be serious.

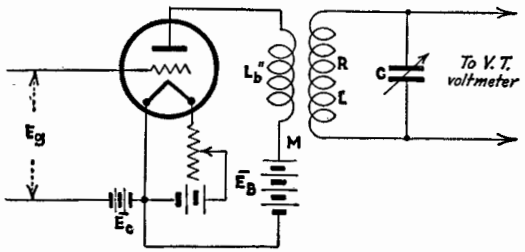


FIG. 193.—Circuits for the first method of obtaining experimentally the regeneration characteristics of a triode.

139. Experimental Determination of Characteristics of Triode for Regeneration with Large Amplitudes.—The curves of Fig. 190 are not experimentally determined but are curves assumed for the purpose of illustrating the discussion just given. We shall now consider two methods of obtaining these curves experimentally.

The first method gives the curves illustrated in Fig. 190 at the operating frequency. Although somewhat less convenient than the second method to be described, it is more accurate since it includes the effect of the tube capacitances. The circuit arrangement for this first method is shown in Fig. 193. Since only the fundamental component of $(I_p)_\omega$ is effective in producing $(E_0')_\omega$, the method adopted to determine E_0 must exclude the harmonics. The oscillatory circuit of Fig. 184, consisting

of R, L, and C, is coupled to the plate coil L_b'' but disconnected from the grid of the active triode, as shown in Fig. 193. The voltage across the oscillatory circuit, tuned to resonance, is

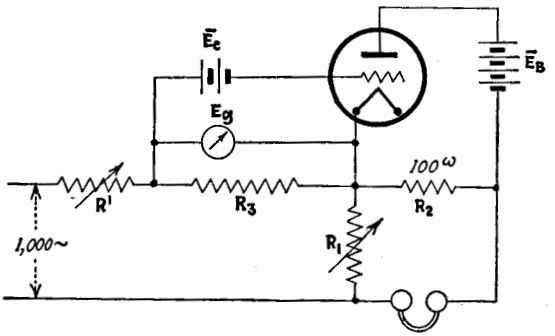


FIG. 194.—Circuits for the second method of obtaining experimentally the regeneration characteristics of a triode.

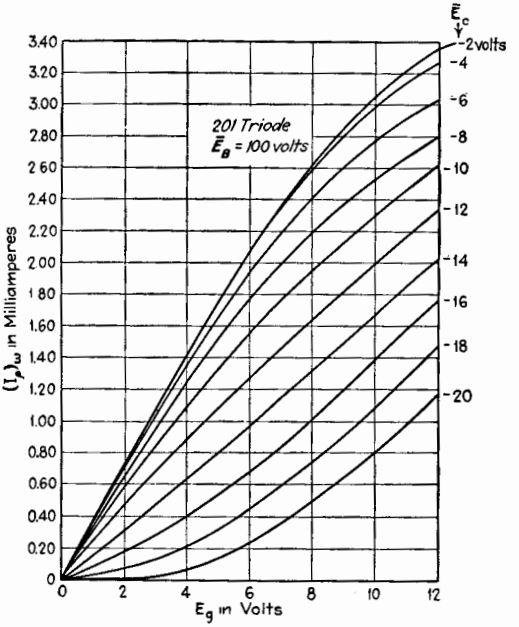


FIG. 195.—Regeneration characteristics of a triode obtained by the second method shown in Fig. 194.

measured by a vacuum-tube voltmeter as the impressed radio-frequency grid voltage E_g is varied. Knowing the values of the constants of the circuits, $(E_0')_\omega$ can be readily calculated.

The effect of the grid conduction upon the circuit line in Fig. 188 can be obtained by measuring the voltage across the oscillatory circuit connected as shown in Fig. 184 with $M = 0$. Knowing the value of the impressed voltage, the curve can readily be calculated.

The second method for determining experimentally the regeneration characteristics of a triode is illustrated in Fig. 194. Resistance R_2 is chosen about equal to the normal resistance load of the plate circuit. The resistance $M^2\omega^2/R$ introduced through M is usually very small, perhaps of the order of 10 ohms. R_3 can have any convenient value, such as 100 or 500 ohms. R_1 is variable. The grid voltage due to the 1,000-cycle current flowing through R_3 is measured by the voltmeter marked E_g in the figure. This voltage can be varied by resistance R'. The resistance R_1 is adjusted to give balance for the fundamental. The ear serves very well as a selective detector to determine when the fundamental component balances, even though unbalanced harmonic currents are present. At balance, the component of the plate current of fundamental frequency is

$$(I_p)_\omega = \frac{E_g R_1}{R_3 R_2}$$

Figure 195 shows a set of curves obtained by the second method for a particular triode at a plate voltage of 100 volts. Other conditions would give curves of different shapes.

CHAPTER XIV

REGENERATION IN COUPLED CIRCUITS WITH SMALL SIGNALS[1]

In Chap. XIII it was shown that in effect a regenerative tube reduces the resistance of a single circuit with which it is associated. Frequently, in practice, a regenerative tube is associated with one of two coupled circuits. The purpose of the present chapter is to extend the treatment of regeneration to two magnetically coupled circuits and to show that in this case also regeneration, in effect, reduces the resistance of the circuit with which it is directly associated. The analysis includes a study of the properties of coupled circuits when the effective resistance of one of them is reduced and becomes negative.

Since the treatment with regeneration must follow the same line as that used for coupled circuits without regeneration, a brief review will be given of the theory of forced oscillations in magnetically coupled circuits.

I. THEORY OF TWO MAGNETICALLY COUPLED CIRCUITS WITHOUT REGENERATION

140. Derivation of the Currents in Coupled Circuits.—Let the two circuits have the constants shown in Fig. 196. The impressed electromotive force E_0 has a frequency $\omega_0/2\pi$. The equations describing the conditions of steady-state forced oscillation of the system are

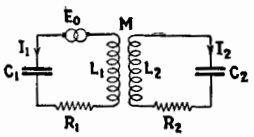


FIG. 196.—Magnetically coupled circuits.

$$\left.\begin{aligned} \mathbf{Z}_1\mathbf{I}_1 + jM\omega_0\mathbf{I}_2 &= \mathbf{E}_0 \\ jM\omega_0\mathbf{I}_1 + \mathbf{Z}_2\mathbf{I}_2 &= 0 \end{aligned}\right\} \tag{384}$$

where

$$\mathbf{Z}_1 = R_1 + j\left(L_1\omega_0 - \frac{1}{C_1\omega_0}\right) = R_1 + jX_1$$

[1] The theory in this chapter was published by the author in greater detail in *Proc. I.R.E.*, **12**, 299 (1924).

and

$$\mathbf{Z}_2 = R_2 + j\left(L_2\omega_0 - \frac{1}{C_2\omega_0}\right) = R_2 + jX_2$$

Solving these equations, the expressions for $\mathbf{I}_1$ and $\mathbf{I}_2$ are

$$\left.\begin{aligned}\mathbf{I}_1 &= \frac{\mathbf{E}_0}{\mathbf{Z}_1 + \dfrac{M^2\omega_0^2}{\mathbf{Z}_2}} = \frac{\mathbf{E}_0}{\mathbf{Z}_{12}} \\ &= \frac{\mathbf{E}_0}{\left[R_1 + \dfrac{M^2\omega_0^2 R_2}{Z_2^2}\right] + j\left[X_1 - \dfrac{M^2\omega_0^2 X_2}{Z_2^2}\right]} \\ &= \frac{\mathbf{E}_0}{R_{12} + jX_{12}}\end{aligned}\right\} \tag{385}$$

$$\left.\begin{aligned}\mathbf{I}_2 &= \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_1\mathbf{Z}_2 + M^2\omega_0^2} \\ &= \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_2\left(\mathbf{Z}_1 + \dfrac{M^2\omega_0^2}{\mathbf{Z}_2}\right)} \\ &= \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_2\mathbf{Z}_{12}}\end{aligned}\right\} \tag{386}$$

The impedance Z_{12} is the equivalent impedance of the whole system, as viewed from the first or primary circuit, and is analytically defined by Eq. (385).

It is frequently more useful to express results in terms of ratios. This can be done easily with Eqs. (385) and (386), giving the following equivalent expressions:

$$\mathbf{I}_1 = \frac{\mathbf{E}_0/L_1\omega_0}{\left[\eta_1 + \dfrac{\tau^2\eta_2}{\eta_2^2 + \left(1 - \dfrac{\omega_2^2}{\omega_0^2}\right)^2}\right] + j\left[1 - \dfrac{\omega_1^2}{\omega_0^2} - \dfrac{\tau^2\left(1 - \dfrac{\omega_2^2}{\omega_0^2}\right)}{\eta_2^2 + \left(1 - \dfrac{\omega_2^2}{\omega_0^2}\right)^2}\right]} \tag{385-R}$$

$$\mathbf{I}_2 = \frac{-j\tau\mathbf{E}_0/\omega_0\sqrt{L_1L_2}}{\left[\eta_1 + j\left(1 - \dfrac{\omega_1^2}{\omega_0^2}\right)\right]\left[\eta_2 + j\left(1 - \dfrac{\omega_2^2}{\omega_0^2}\right)\right] + \tau^2} \tag{386-R}$$

where

$$\eta_1 = \frac{R_1}{L_1\omega_0};\quad \eta_2 = \frac{R_2}{L_2\omega_0};\quad \tau = \frac{M}{\sqrt{L_1L_2}};\quad \omega_1^2 = \frac{1}{L_1C_1};\quad \omega_2^2 = \frac{1}{L_2C_2}.$$

Equations (385) and (386) give the primary and secondary currents for any adjustments of the primary circuit, of the secondary circuit, and of their coupling. If the coupling between the two circuits is fixed, X_1 and X_2 are the independent variables and the current as dependent variable can be plotted vertically against the variables X_1 and X_2 (or against any other quantities upon which X_1 and X_2 depend) as the two horizontal coordinates. Thus there is formed a curved surface bounding a space model. Such curved surfaces are shown in Plate IV for the secondary current I_2.

Since, in most practical cases, the value of the secondary current is of greater interest than the value of the primary current, the shape of the representative surface for I_2 will be examined in detail. The expression (386) may be written

$$\left.\begin{aligned}\mathbf{I}_2 &= \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_1\left(\mathbf{Z}_2 + \dfrac{M^2\omega_0^2}{\mathbf{Z}_1}\right)} = \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_1\mathbf{Z}_{21}} \\ &= \frac{-jM\omega_0\mathbf{E}_0}{\mathbf{Z}_1\left[\left(R_2 + \dfrac{M^2\omega_0^2R_1}{R_1^2 + X_1^2}\right) + j\left(X_2 - \dfrac{M^2\omega_0^2X_1}{R_1^2 + X_1^2}\right)\right]}\end{aligned}\right\} \quad (387)$$

where $\mathbf{Z}_{21}$ is the impedance of the system as viewed from the secondary circuit. Equation (387) can be written in terms of coefficients as follows:

$$\mathbf{I}_2 = \frac{-j\tau\mathbf{E}_0/\omega_0\sqrt{L_1L_2}}{\left[\eta_1 + j\beta_1\right]\left[\left(\eta_2 + \dfrac{\tau^2\eta_1}{\eta_1^2 + \beta_1^2}\right) + j\left(\beta_2 - \dfrac{\tau^2\beta_1}{\eta_1^2 + \beta_1^2}\right)\right]} \quad (387\text{-}R)$$

where

$$\beta_1 = 1 - \frac{\omega_1^2}{\omega_0^2} = 1 - \frac{\lambda_0^2}{\lambda_1^2}, \text{ and } \beta_2 = 1 - \frac{\omega_2^2}{\omega_0^2} = 1 - \frac{\lambda_0^2}{\lambda_2^2}.$$

141. Conditions for and Value of Maximum Secondary Current.—Since X_2 occurs only in the last term in the denominator of the second form of Eq. (387), the value of X_2 to give a maximum secondary current for any value of X_1 is

Max. 1-2 line. Secondary reactance to give max. I_2

$$\left.\begin{aligned} X_2 &= \frac{M^2\omega_0^2X_1}{R_1^2 + X_1^2} \quad \text{or} \quad (388) \\ \beta_2 &= \frac{\tau^2\beta_1}{\eta_1^2 + \beta_1^2} \quad (388\text{-}R)\end{aligned}\right\}$$

The relation (388) gives a curve on the horizontal plane, the $X_1 - X_2$ plane of the space model, above which may be plotted the maximum values of I_2 obtained by first setting X_1 and then adjusting X_2. This locus of maximum secondary current is called the *max.* 1-2 *line*, and its shape is independent of the value of the resistance of the secondary circuit. If the value of X_2 given by Eq. (388) is substituted in Eq. (387), then, as X_1 is varied, X_2 always has the correct value to give a maximum I_2, and we follow along the ridge of the space model shown in Plate IV. The maximum value of I_2 is

Max. I_2 over max. 1-2 line

$$\begin{cases} \text{Max. } \mathbf{I}_2 = \dfrac{-jM\omega_0\mathbf{E}_0}{(R_1 + jX_1)\left(R_2 + \dfrac{M^2\omega_0^2 R_1}{R_1^2 + X_1^2}\right)} & \text{or} \quad (389) \\ \text{Max. } \mathbf{I}_2 = \dfrac{-j\tau\mathbf{E}_0/\omega_0\sqrt{L_1L_2}}{(\eta_1 + j\beta_1)\left(\eta_2 + \dfrac{\tau^2\eta_1}{\eta_1^2 + \beta_1^2}\right)} & (389\text{-}R) \end{cases}$$

142. Conditions for and Value of Max. Max. Secondary Current.—Some value of X_1 gives the largest value of I_2 attainable (called the max. max. I_2) represented by the highest peak of the ridge on the space model. If Eq. (389) is differentiated to find the value of X_1 which gives this max. max. I_2, the following roots result:

Conditions for max. max. I_2 or min. max. I_2

$$\begin{cases} X_1 = 0 & (390\text{-}1) \\ Z_1^2 = \dfrac{M^2\omega_0^2 R_1}{R_2} & (390\text{-}2) \end{cases}$$

Gives imaginary X_1

$$Z_1^2 = -\frac{M^2\omega_0^2 R_1}{R_2} \qquad (390\text{-}3)$$

Root (390-3), obtained by equating to zero the denominator of Eq. (389), gives an imaginary value of X_1. Roots (390-1) and (390-2) give the same value of max. max. I_2 if

Critical coupling

$$\begin{cases} M^2\omega_0^2 = R_1R_2 & \text{or} \quad (391) \\ \tau^2 = \eta_1\eta_2 & (391\text{-}R) \end{cases}$$

This relation divides the problem into two parts, and the coupling which satisfies Eq. (391) is known as *critical coupling*. For couplings less than critical, root (390-2) gives imaginary values of X_1, but root (390-1) gives the max. max. secondary current, whose value is

Max. max. I_2 when $\tau^2 < \eta_1\eta_2$

$$\text{Max. max. } \mathbf{I}_2 = \frac{-jM\omega_0\mathbf{E}_0}{R_1R_2 + M^2\omega_0^2} \quad \text{or} \quad (392)$$

$$\text{Max. max. } \mathbf{I}_2 = \frac{-j\tau\mathbf{E}_0/\omega_0\sqrt{L_1L_2}}{\eta_1\eta_2 + \tau^2} \quad (392\text{-}R)$$

For couplings greater than critical coupling, root (390-1) gives a min. max. I_2, and root (390-2) gives the max. max. I_2.

Max. max. I_2 when $\tau^2 > \eta_1\eta_2$

$$\text{Max. max. } \mathbf{I}_2 = \frac{-jM\omega_0\mathbf{E}_0}{2R_2\left(R_1 \pm j\sqrt{\frac{M^2\omega_0^2R_1}{R_2} - R_1^2}\right)} \quad \text{or} \quad (393)$$

$$\text{Max. max. } \mathbf{I}_2 = \frac{-j\tau\mathbf{E}_0/\omega_0\sqrt{L_1L_2}}{2\eta_2\left(\eta_1 \pm j\sqrt{\frac{\tau^2\eta_1}{\eta_2} - \eta_1^2}\right)} \quad (393\text{-}R)$$

The numerical values for max. max. I_2 from Eqs. (393) and (393-R) are

Numerical value of max. max. I_2 when $\tau^2 > \eta_1\eta_2$

$$\text{Max. max. } I_2 = \frac{E_0}{2\sqrt{R_1R_2}} \quad \text{or} \quad (394)$$

$$\text{Max. max. } I_2 = \frac{E_0/\omega_0\sqrt{L_1L_2}}{2\sqrt{\eta_1\eta_2}} \quad (394\text{-}R)$$

Equation (394) shows that max. max. I_2 is the same for all values of τ greater than the critical value.

143. Study of the Space Model for Secondary Current.—In constructing a space model for the secondary current, the ratio of the numerical value of the secondary current to the numerical value of the max. max. secondary current is convenient. This ratio, derived from Eqs. (386) and (394), is applicable to all cases above critical coupling and including the critical case itself.

$\tau^2 > \eta_1\eta_2$

$$\frac{I_2}{I_{2mm}} = \frac{2M\omega_0\sqrt{R_1R_2}}{\sqrt{Z_1^2Z_2^2 + M^4\omega_0^4 + 2M^2\omega_0^2(R_1R_2 - X_1X_2)}} \quad (395)$$

$$\frac{I_2}{I_{2mm}} = \frac{2\tau\sqrt{\eta_1\eta_2}}{\sqrt{(\eta_1\eta_2 - \beta_1\beta_2 + \tau^2)^2 + (\eta_1\beta_2 + \eta_2\beta_1)^2}} \quad (395\text{-}R)$$

The corresponding expressions for this ratio for couplings less than critical are obtained from Eqs. (386) and (392) and are

$\tau^2 < \eta_1\eta_2$

$$\frac{I_2}{I_{2mm}} = \frac{R_1R_2 + M^2\omega_0^2}{\sqrt{Z_1^2Z_2^2 + M^4\omega_0^4 + 2M^2\omega_0^2(R_1R_2 - X_1X_2)}} \quad (396)$$

$$\frac{I_2}{I_{2mm}} = \frac{\eta_1\eta_2 + \tau^2}{\sqrt{(\eta_1\eta_2 - \beta_1\beta_2 + \tau^2)^2 + (\eta_1\beta_2 + \eta_2\beta_1)^2}} \quad (396\text{-}R)$$

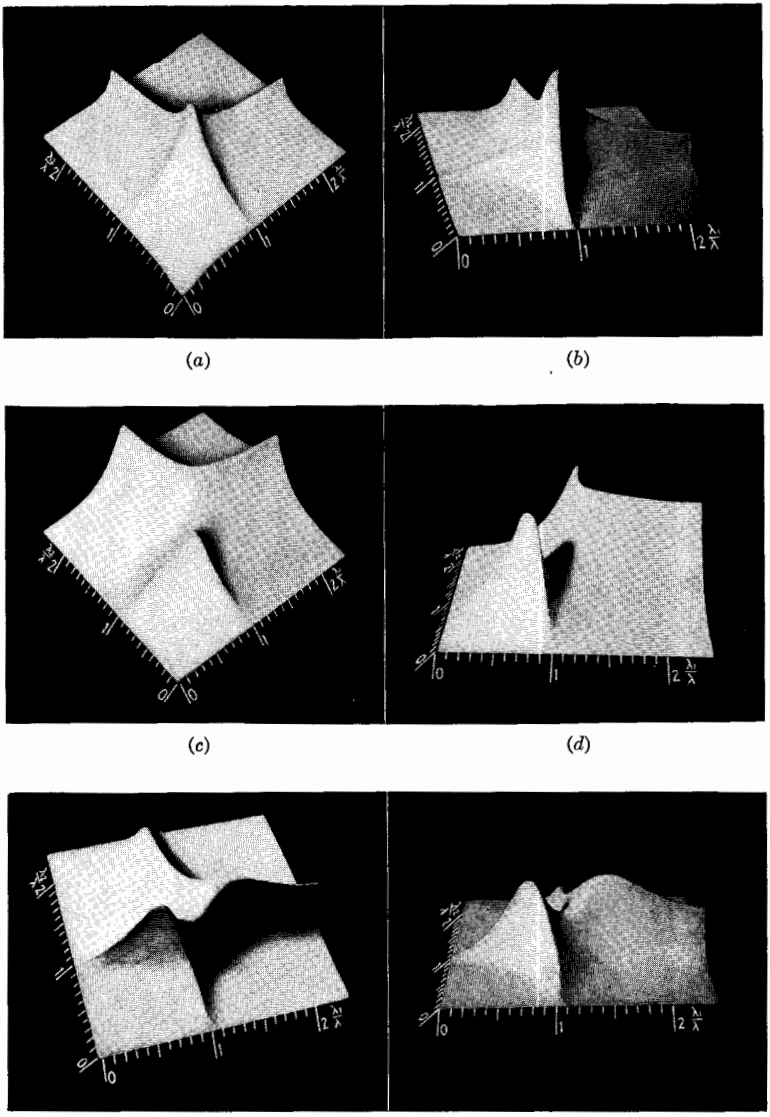


PLATE IV.—Space models of secondary current of two magnetically coupled circuits. (*a*) and (*b*) critical coupling; (*c*) and (*d*) coupling greater than critical. $\eta_1 = \eta_2$. (*e*) and (*f*) coupling greater than critical. $\eta_1 > \eta_2$.

(*Facing page* 348)

In plotting the ratio given in Eqs. (395) and (396), or any of the other relations, the independent variables β_1 and β_2 may be used instead of the corresponding quantities X_1 and X_2. On the other hand, a better physical picture usually results if the quantities $\omega_0/\omega_1 = \lambda_1/\lambda_0$ and $\omega_0/\omega_2 = \lambda_2/\lambda_0$ are used as coordinates. The surfaces shown in Plate IV were calculated from Eq. (395) and are constructed with λ_1/λ_0 and λ_2/λ_0 as independent variables.

If the coupling between two circuits is less than critical coupling as defined by Eq. (391), the space model for the secondary-current

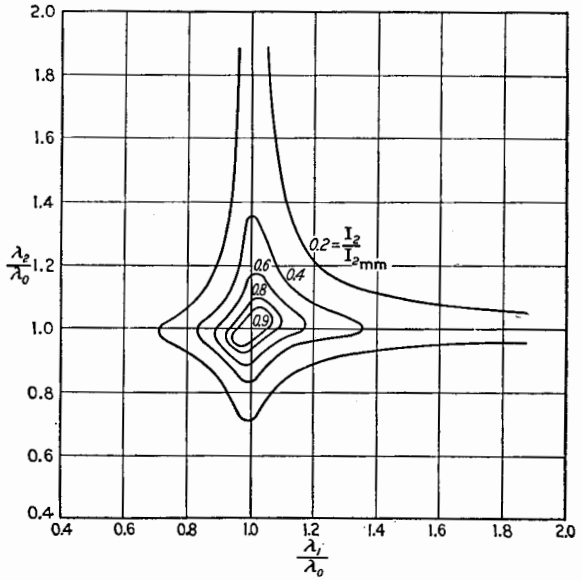


FIG. 197.—Contour lines of constant secondary current for critical coupling.

ratio given by Eq. (396) has a single maximum at values of the independent variables

$$\beta_1 = 0 \text{ or } X_1 = 0 \text{ or } \frac{\lambda_1}{\lambda_0} = 1$$

and

$$\beta_2 = 0 \text{ or } X_2 = 0 \text{ or } \frac{\lambda_2}{\lambda_0} = 1$$

The height of this maximum is given by Eq. (396).

At critical coupling, the space model has still a single peak, the height of which is unity. Plate IV, *a* and *b*, shows a calculated model for critical coupling. The peak has a peculiar shape best shown by contour lines given in Fig. 197.

If the coupling is greater than critical, the space model has in general two ridges, each having a highest point of value unity. Two models calculated from specific values of the coefficients are shown in Plate IV, *c* and *d*, and *e* and *f* (*c* and *d* are two views of the same model and similarly *e* and *f* are two views of another model). As pointed out, if the secondary circuit is adjusted for

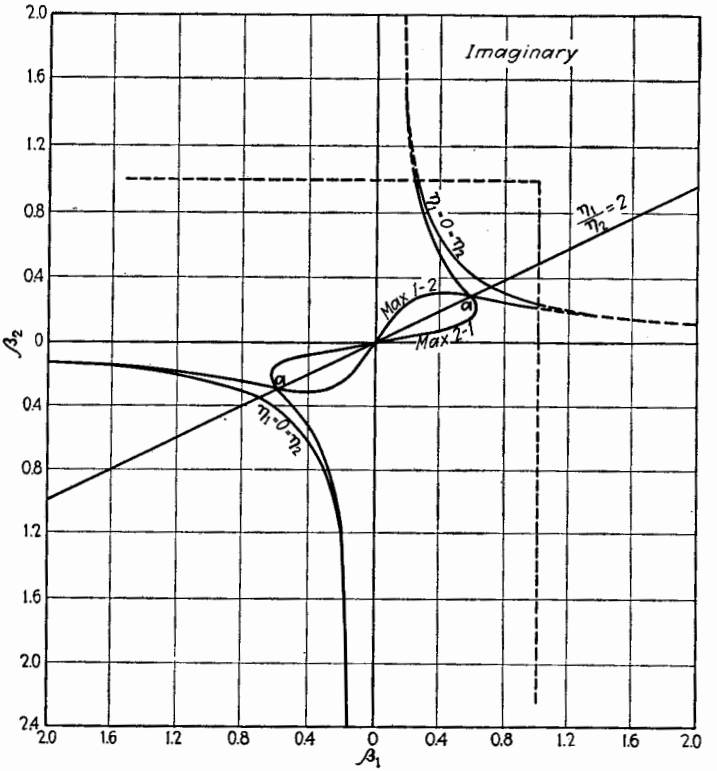


FIG. 198.—Curves of β_1 *vs.* β_2 for max. I_2. $\eta_2 = 0.4$. $\eta_2 = 0.2$. $\tau = 0.5$.

maximum current for successive adjustments of the primary circuit, a locus of points on the horizontal plane is found by Eq. (388). Such a locus is shown by the curve marked max. 1-2 (meaning primary reactance given a value and then secondary reactance adjusted) in Fig. 198, where the coordinates are β_1 and β_2, and in Fig. 199, where the same curve is plotted against λ_1/λ_0 and λ_2/λ_0. It is to be noted that all curves plotted to β_1 and β_2 are symmetrical about the origin, but points for β_1 and β_2

greater than +1 give imaginary values of λ_1/λ_0 and λ_2/λ_0. The max. 1-2 line (see Fig. 199) has a maximum and a minimum value of λ_2/λ_0, and passes through the point ($\lambda_1/\lambda_0 = 1$, $\lambda_2/\lambda_0 = 1$). The maximum and minimum values of λ_2/λ_0 have their corresponding maximum and minimum values of

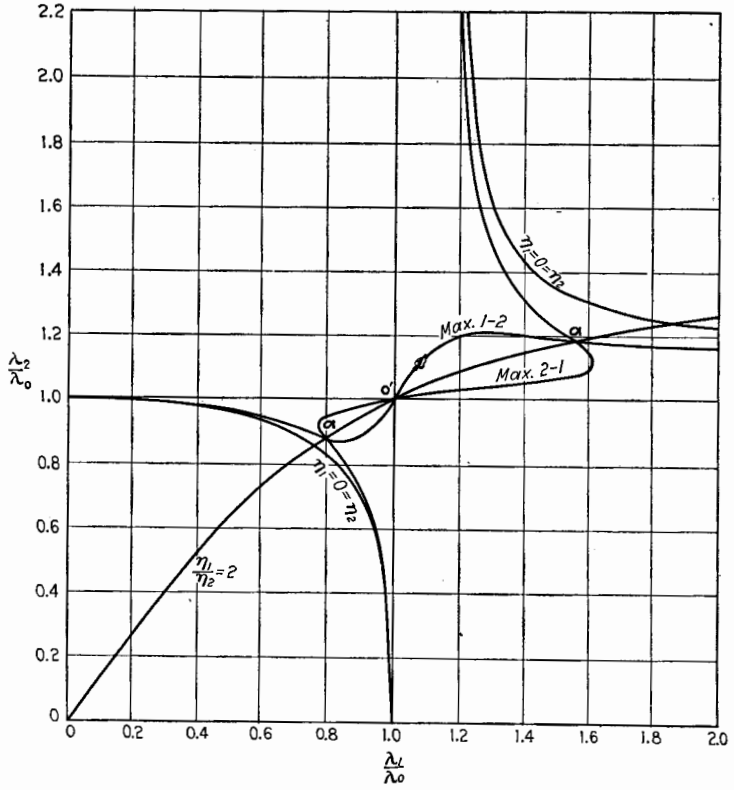


FIG. 199.—Curves of $\frac{\lambda_2}{\lambda_0}$ *vs.* $\frac{\lambda_1}{\lambda_0}$ for max. I_2. $\eta_1 = 0.4$. $\eta_2 = 0.2$. $\tau = 0.5$.

β_2 in Fig. 199. From Eq. (388-*R*), the value of β_1 which makes β_2 equal to unity is

$$\beta_1 = \frac{\tau^2 \pm \sqrt{\tau^4 - 4\eta_1^2}}{2} \left(\text{for } \frac{\lambda_2}{\lambda_0} = \infty\right) \tag{397}$$

From Eq. (397), the maximum of the max. 1-2 line does not extend to $\lambda_2/\lambda_0 = \infty$ if $\tau^2 < 2\eta_1$. The curve just reaches infinity if $\tau^2 = 2\eta_1$. The curve approaches infinity asymptotic to the two

vertical lines $\beta_1 = \dfrac{\tau^2 \pm \sqrt{\tau^4 - 4\eta_1^2}}{2}$. The minimum of the max. 1-2 line never reaches the horizontal axis of $\lambda_2/\lambda_0 = 0$ except when $\eta_1 = 0$.

It can be shown that if the reverse order of adjusting the circuits for a maximum is adopted, *i.e.*, if the primary circuit is adjusted for a maximum secondary current for every adjustment of the secondary circuit, a new curve is obtained, given by the equation

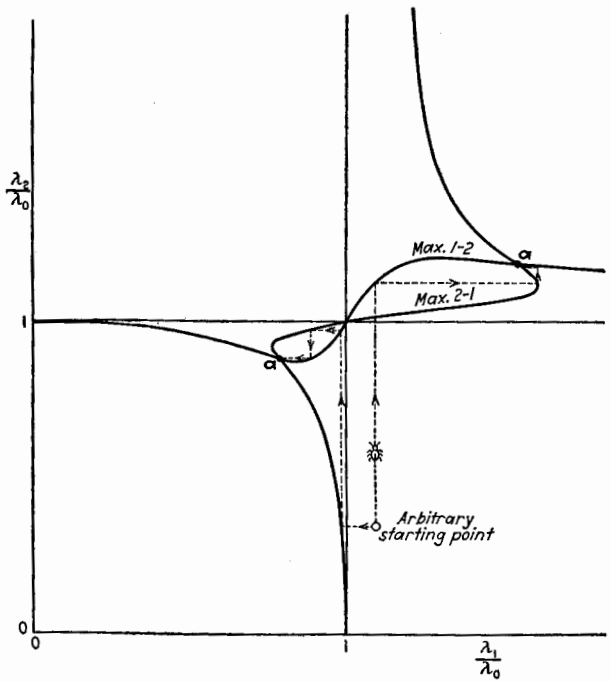


FIG. 200.—Method of finding the positions of max. max. I_2.

Max. 2-1 line. Primary reactance to give max. I_2

$$\begin{cases} X_1 = \dfrac{M^2\omega_0^2 X_2}{R_2^2 + X_2^2} & \text{or} \quad (398) \\ \beta_1 = \dfrac{\tau^2\beta_2}{\eta_2^2 + \beta_2^2} & (398\text{-}R) \end{cases}$$

The locus given by Eq. (398) is shown by the curve max. 2-1 in Figs. 198 and 199. The intersections of the max. 1-2 and max. 2-1 lines, shown at *a* and *a* in Figs. 198 and 199, give the positions

of the max. max. secondary current, the λ_1/λ_0 coordinates of which are given by Eq. (390).

That the intersections of the max. 1-2 and max. 2-1 lines give the coordinates of the max. max. I_2 can be shown as follows: Suppose that an intelligent bug starts from any point on the space model and moves only in directions parallel to the coordinate axes. When the bug moves parallel to the λ_2/λ_0 axis, it moves until it arrives at the highest point of the surface, which is

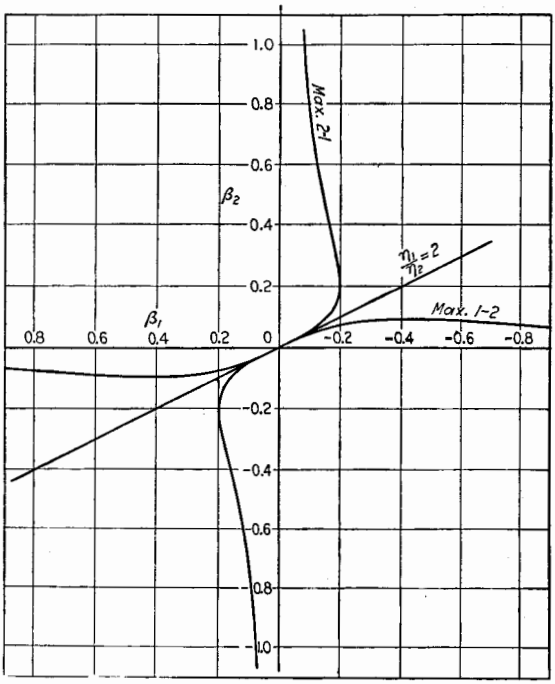


FIG. 201.—Curves of β_1 *vs.* β_2 for max. I_2. Case of critical coupling. $\eta_1 = 0.4$. $\eta_2 = 0.2$. $\tau^2 = 0.08$.

over the max. 1-2 line. When the bug moves parallel to the λ_1/λ_0 axis, it moves until it arrives at the highest point of the surface which lies above the max. 2-1 line. Two possible courses of the bug are shown in Fig. 200. One course leads to one max. max. point and the other course leads to the other max. max. point. The same process can be applied to the case of critical coupling, shown in Fig. 201.

If $\eta_1 = \eta_2 = 0$, curves max. 1-2 and max. 2-1 resolve into the curves so marked in Figs. 198 and 199. At critical coupling, the

intersection points a a of Fig. 198 merge into a single point at the origin as in Fig. 201. If $\eta_1 = \eta_2$, the max. 1-2 and max. 2-1 lines are similar in shape and symmetrical about the 45-deg. line, but otherwise the two lines are dissimilar, as is shown in the figures.

The reason the two curves max. 1-2 and max. 2-1 differ according to the order of adjustment becomes apparent if the shape of

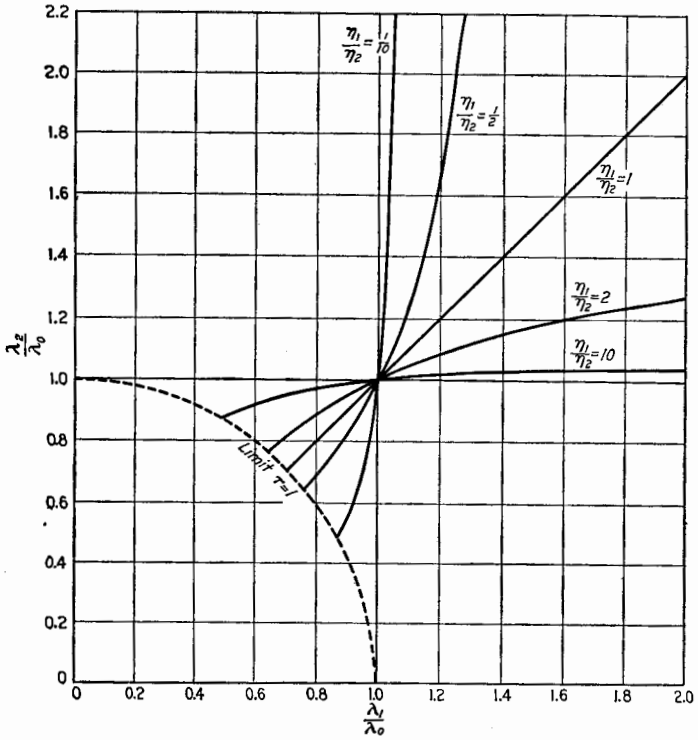


FIG. 202.—Loci of max. max. I_2 as M is varied.

the saddle between the two peaks (see Plate IV, c, d, e, and f) is considered. Suppose a card with its plane vertical and its lower edge horizontal and parallel to the λ_2/λ_0 axis rests on the surface at one point. If the card is now moved in the λ_1/λ_0 direction, the point of tangency is always directly over the max. 1-2 line. The max. 2-1 line can be traced by the card placed parallel to the λ_1/λ_0 axis and moved in the direction of the λ_2/λ_0 axis.

For any particular coupling, the highest peaks have definite coordinates. If the coupling is changed, the two peaks move along a curve obtained by eliminating the coupling from Eqs. (388) and (390-2), giving

Locus of max. max. I_2 as M is varied

$$\begin{cases} \dfrac{X_1}{X_2} = \dfrac{R_1}{R_2} & (399) \\ \dfrac{\beta_1}{\beta_2} = \dfrac{\eta_1}{\eta_2} & (399\text{-}R) \end{cases}$$

Equation (399) is plotted in Figs. 198, 199, and 201 for the particular value $\eta_1/\eta_2 = 2$, and for other values of η_1/η_2 in Fig. 202. The dotted line in Fig. 202 gives the max. I_2 line for $r = 1$ and hence is the limit of the loci of the peaks.

It can be shown that, for all values of coupling, the ridge along the saddle between the two peaks is directly above the curve of Eq. (399) for the appropriate value of η_1/η_2.

Attention is now called to the fact that if η_1 is equal to η_2, the model is symmetrical with respect to the coordinate axes, and the peaks lie over the 45-deg. line, *i.e.*, over the line $\beta_1 = \beta_2$ or $\lambda_1/\lambda_0 = \lambda_2/\lambda_0$ (see Plate IV, *c* and *d*). If η_1 is not equal to η_2, the peaks are skewed around as shown in Plate IV, *e* and *f*. In practice, the primary circuit may be the antenna circuit and the secondary circuit of the present discussion may be the secondary circuit of a coupled-circuit receiver. On account of the ground and radiation resistance, the value of η for the antenna circuit is many times that for the secondary circuit in almost all such practical cases. This is especially true when regeneration is used in the secondary circuit. It is evident that if η_1 is greater than η_2 the space model of the practical receiver is very much distorted, even more so than is represented by the model in Plate IV, *e* and *f*.

144. Sections through the Space Model for Secondary Current. Examine, by means of these space models, the selectivity and tuning relations of a coupled-circuit receiver. If the circuits are adjusted to any point on the $\dfrac{\lambda_1}{\lambda_0} - \dfrac{\lambda_2}{\lambda_0}$ plane, and the natural wave length of the primary circuit is varied while a signal of constant wave length λ_0 is being received, the variation in secondary current is given by the cross section of the model along a line through the given point parallel to the λ_1/λ_0 axis. Simi-

larly, a cross section parallel to the λ_2/λ_0 axis gives the variation of current when the secondary circuit reactance is varied.

If the two circuits are fixed and the incoming wave length is changed, the resulting variation of secondary current can be obtained from Eq. (395) or Eq. (396) by varying λ_0, remembering that X_1 and X_2 or β_1 and β_2 are functions of λ_0. Since λ_0 appears symmetrically in the coordinates of the space model, the variation in I_2/I_{2mm} is given by allowing λ_1/λ_0 and λ_2/λ_0 to vary proportion-

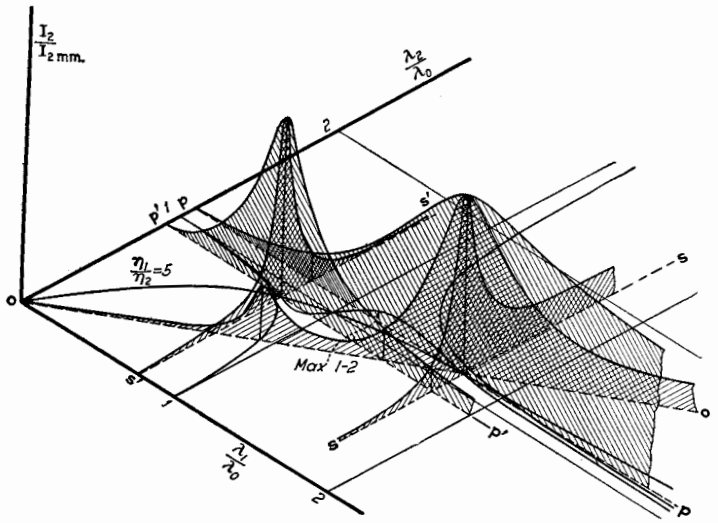


FIG. 203.—Sections of the space model for I_2. Coupling greater than critical. $\eta_1 = 5\eta_2$.

ately, *i.e.*, by taking a cross section through the space model along a radial line through a point determined by the settings of λ_1 and λ_2. In using this method of cross-sectioning, it must be remembered that, since the variation in the coordinates λ_1/λ_0 and λ_2/λ_0 is inversely proportional to λ_0, the shape of the curve is somewhat distorted from what it would be if plotted in the usual way against λ_0 directly. Moving along the radial line toward the origin means an *increase* in λ_0. If the distance moved is very small, the distortion due to this effect is small. However, the abscissa scale for this section is directly proportional to ω_0.

Examine now a specific case, when $\eta_1 = 5\eta_2$. The factor of proportionality 5 is much less than is often met with in practice. Because of distributed capacitance of the coils, a stronger signal

is often obtained when the coupling is above critical and the adjustments are made for the long wave-length maximum. This gives sufficient reason for considering the case shown in Plate IV, *e* and *f*, as a practical case. It is evident that the primary circuit is adjusted to a much longer wave length than the resonant value for the incoming wave, and that the apparent tuning in the primary circuit is very dull, as is shown by the section *pp*, Fig. 203, parallel to the λ_1/λ_0 axis of the space model. The tuning in the secondary circuit is sharp, as is indicated by the section *ss*. The selectivity against interference from other stations is relatively low, as is indicated by the diagonal section along the

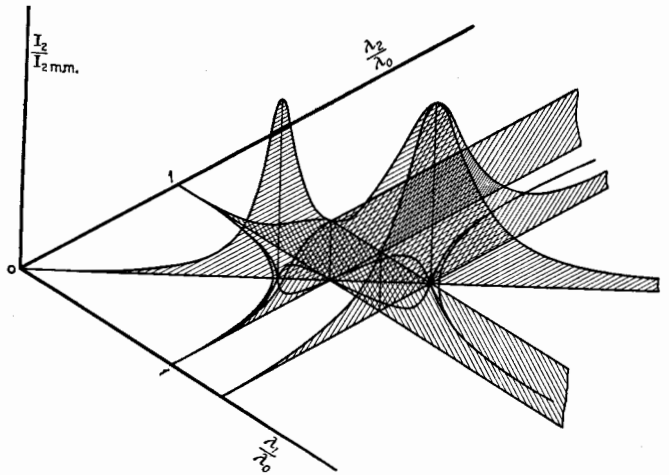


FIG. 204.—Sections of space model for I_2. Coupling greater than critical. $\eta_1 = \eta_2$.

line *oo*. The corresponding sections $p'p'$ and $s's'$ through the other peak of the space model are also shown in Fig. 203.

Figure 204 is a sectional drawing of the special case when $\eta_1 = \eta_2$. As already pointed out, the space model under these conditions is symmetrical (see *c* and *d* of Plate IV), and the two max. max. current peaks lie over the 45-deg. line on the $\frac{\lambda_1}{\lambda_0} - \frac{\lambda_2}{\lambda_0}$ plane.

The preceding discussion indicates the manner in which the space model can be used to study the equivalent resonance curve and the selectivity and tuning relations under various conditions of the circuits. Enough has been given to show that the ratio of

the values of η for the two circuits is important, as well as their absolute values.

Referring now to the case of critical coupling, there is only one main maximum. But if the circuits are not in resonance, there

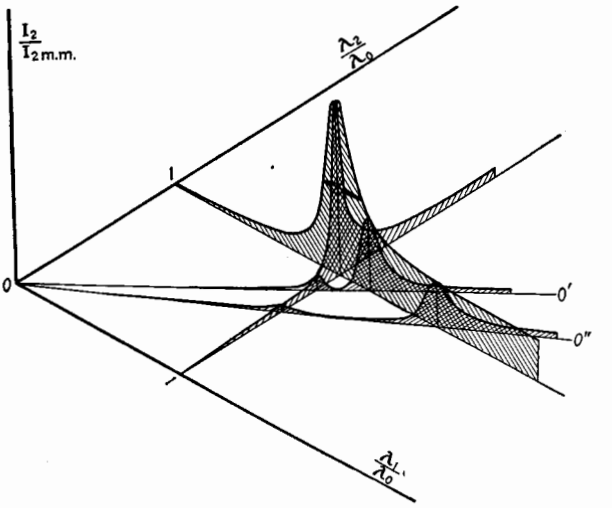


FIG. 205.—Sections of the space model for I_2. Critical coupling. $\eta_1 > \eta_2$.

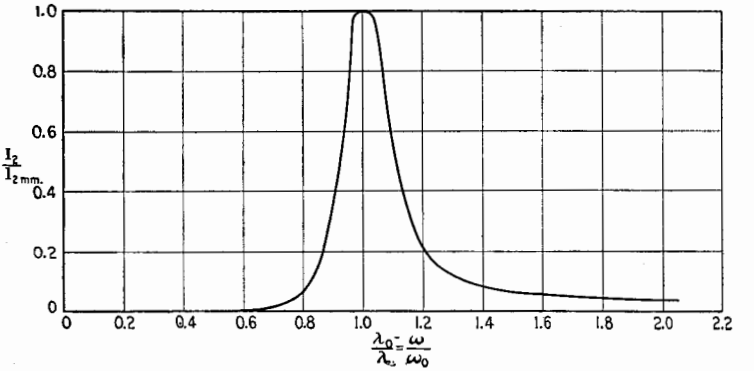


FIG. 206.—Section of the space model for I_2 along the 45-deg. line. Critical coupling. $\eta_1 = \eta_2 = 0.10$. $\tau = 0.10$.

are two values of the incoming wave length which give maxima. This is shown by the sections oo' and oo'' in Fig. 205. This figure shows cross sections of the space model for critical coupling when $\eta_1 > \eta_2$. It is interesting to note that the equivalent resonance curve at critical coupling, when the incoming wave

length is varied, is broad at the top, as shown in Fig. 206, whereas the curve obtained for a fixed λ_0 passing through the point ($\lambda_1/\lambda_0 = 1$, $\lambda_2/\lambda_0 = 1$) is sharper as either λ_1 or λ_2 is varied (see Fig. 197). This broadness of resonance shown by the flat top of the curve is of advantage when receiving a radio-telephone signal having a broad spectrum.

The reason for the broadness of the critical-coupling resonance curve as λ_0 is varied is easily seen when it is remembered that, for all couplings above zero, coupled circuits have two free periods of oscillation and there are two peaks of secondary current when the frequency of the impressed e.m.f. is varied. For couplings above critical, the two current peaks can be resolved, but at and below critical coupling the peaks are so close that they merge and cannot be resolved.

145. Conditions for and Value of Maximum Secondary Voltage.—The conditions for and value of maximum secondary current have been deduced. Often a triode amplifier is connected across the secondary condenser, and the conditions for and value of the maximum voltage across the secondary condenser are of more interest than the conditions for and value of maximum current.

The value of the voltage E_{c2} (see Fig. 196) across the condenser in the secondary circuit is

$$\mathbf{E}_{c2} = \frac{-M\omega_0\mathbf{E}_0}{C_2\omega_0[\mathbf{Z}_1\mathbf{Z}_2 + M^2\omega_0^2]} \tag{400}$$

$$= \frac{-M\omega_0\mathbf{E}_0}{C_2\omega_0\mathbf{Z}_2\left[\left(R_1 + \frac{M^2\omega_0^2R_2}{Z_2^2}\right) + j\left(X_1 - \frac{M^2\omega_0^2X_2}{Z_2^2}\right)\right]} \tag{401}$$

In Eq. (401) the primary capacitance C_1 occurs only in X_1, so that the value of C_1 to give a maximum secondary voltage E_{c2} is given by equating to zero the second parenthesis of the denominator.

Max. 2-1 line. Primary reactance to give max. E_{c2}

$$\begin{cases} X_1 = \dfrac{M^2\omega_0^2X_2}{Z_2^2} \text{ or} & (402) \\ \beta_1 = \dfrac{\tau^2\beta_2}{\eta_2^2 + \beta_2^2} & (402\text{-}R) \end{cases}$$

This equation is the same as Eq. (398) which gives the locus on the $\frac{\lambda_1}{\lambda_0} - \frac{\lambda_2}{\lambda_0}$ plane above which the secondary current is a

maximum. Hence, maximum secondary voltage, as well as maximum secondary current, occurs over the max. 2-1 line when the order of adjustment of the circuits is secondary reactance and then primary reactance.

The value of max. E_{c2} over the max. 2-1 line is

Max. E_{c2} over max. 2-1 line

$$\left.\begin{aligned} \text{Max. } \mathbf{E}_{c2} &= -\frac{M\mathbf{E}_0}{C_2\mathbf{Z}_2\left(R_1 + \dfrac{M^2\omega_0^2 R_2}{Z_2^2}\right)} \quad \text{or} \qquad (403) \\ \text{Max. } \mathbf{E}_{c2} &= -\frac{\tau\mathbf{E}_0\sqrt{L_2/L_1}\cdot\omega_2^2/\omega_0^2}{(\eta_2 + j\beta_2)\left(\eta_1 + \dfrac{\tau^2\eta_2}{\eta_2^2 + \beta_2^2}\right)} \qquad (403\text{-}R) \end{aligned}\right\}$$

Since in Eq. (400) C_2 occurs outside the bracket as well as in $\mathbf{Z}_2$, the locus of max. E_{c2}, when the reverse order of setting the two circuits is used, does *not* coincide with the max. 1-2 line. The locus of the maximum secondary voltage, corresponding to the max. 1-2 line for secondary current, can be obtained by differentiating Eq. (400) with respect to C_2 or X_2.

146. Locus of Max. Max. Secondary Voltage.—The value of X_2 to give max. max. E_{c2} can be found by differentiating Eq. (403) with respect to X_2. Equation (403) can be rewritten as follows:

$$\text{Max. } E_{c2} = \frac{M\omega_0 E_0(X_2 - L_2\omega_0)}{\sqrt{R_2^2 + X_2^2}\left[R_1 + \dfrac{M^2\omega_0^2 R_2}{R_2^2 + X_2^2}\right]} \qquad (404)$$

Equating to zero tne derivative of Eq. (404) with respect to X_2,

$$\left.\begin{aligned} &R_1\left(1 - \frac{X_2(X_2 - L_2\omega_0)}{Z_2^2}\right) \\ &\qquad + \frac{M^2\omega_0^2 R_2}{Z_2^2}\left(1 + \frac{X_2(X_2 - L_2\omega_0)}{Z_2^2}\right) = 0 \\ &\text{or} \\ &\frac{M^2\omega_0^2 R_2}{Z_2^2} = R_1\frac{1}{1 - \dfrac{2(R_2^2 + X_2^2)}{R_2^2 + X_2 L_2\omega_0}} \end{aligned}\right\} \qquad (405)$$

Equation (405) is a cubic in X_2 and difficult of solution. However, we can obtain the result desired, *i.e.*, the locus of max. max. E_{c2}, by substituting Eq. (402) in Eq. (405). Equation (405) gives the value of X_2 for max. max. E_{c2}. This value of X_2 must also be located on the max. 2-1 line. By eliminating

$M^2\omega_0^2$ from Eqs. (402) and (405), the resulting equation in X_1 and X_2 is the locus of max. max. E_{c2} as M is varied. The result of this elimination is

Locus of max. max. E_{c2} as M is varied

$$\left.\begin{aligned} X_1 &= X_2 \cdot \frac{R_1}{R_2} \cdot \frac{-(R_2^2 + X_2 L_2 \omega_0)}{R_2^2 - X_2 L_2 \omega_0 + 2X_2^2} \quad \text{or} \qquad (406) \\ \beta_1 &= \beta_2 \cdot \frac{\eta_1}{\eta_2} \cdot \frac{-(\eta_2^2 + \beta_2)}{\eta_2^2 - \beta_2 + 2\beta_2^2} \qquad (406\text{-}R) \end{aligned}\right\}$$

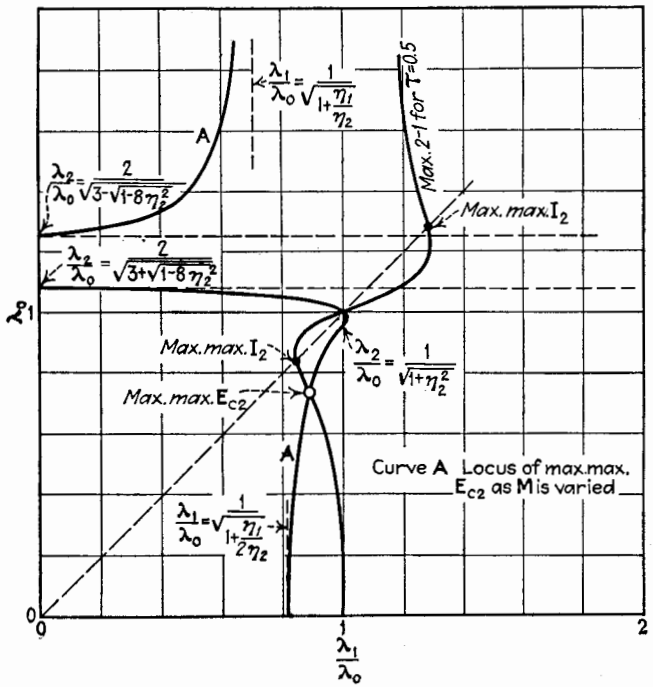


FIG. 207.—Locus of max. max. E_{c2} as M is varied, and conditions for max. max. E_{c2} for $\tau = 0.5$. $\eta_1^2 = \eta_2^2 = 0.10$.

Equation (406) differs from Eq. (399) for the locus of max. max. I_2 only in the presence of the last fraction in Eq. (406).

Examine on the $\frac{\lambda_2}{\lambda_0} - \frac{\lambda_1}{\lambda_0}$ plane the shape of the locus given by Eq. (406-R). When β_2 is zero, β_1 is also zero. Hence, the locus passes through the point $\lambda_1/\lambda_0 = 1$, $\lambda_2/\lambda_0 = 1$. From the numerator, β_1 is zero when $\beta_2 = -\eta_2^2$. That is, the locus cuts across the vertical $\lambda_1/\lambda_0 = 1$ line at a second point whose ordinate

is $\lambda_2/\lambda_0 = 1/\sqrt{1 + \eta_2^2}$, as shown in Fig. 207. The denominator of Eq. (406-*R*) is zero when $\beta_2 = \frac{1 \pm \sqrt{1 - 8\eta_2^2}}{4}$. β_1 is then infinite or λ_1/λ_0 is zero. Transforming these values of β_2 into wave-length ratios, it is seen that the locus of max. max. E_{c2} intersects the vertical coordinate axis when λ_2/λ_0 has the two values $\frac{2}{\sqrt{3 \pm \sqrt{1 - 8\eta_2^2}}}$, Fig. 207. The values of λ_1/λ_0 between these two values of λ_2/λ_0 are either greater than unity or imaginary. For the constants assumed in plotting Fig. 207, the values of λ_1/λ_0 are imaginary between these two values of λ_2/λ_0. The locus intersects the horizontal coordinate axis when β_2 is infinite. This occurs when $\beta_1 = -\frac{\eta_1}{2\eta_2}$ or $\frac{\lambda_1}{\lambda_0} = \frac{1}{\sqrt{1 + \frac{\eta_1}{2\eta_2}}}$. Finally, β_1 is equal to $-\eta_1/\eta_2$ when β_2 is equal to unity, or, in terms of wave lengths, $\frac{\lambda_2}{\lambda_0}$ is infinite when $\frac{\lambda_1}{\lambda_0} = \frac{1}{\sqrt{1 + \frac{\eta_1}{\eta_2}}}$, as indicated by the position of the infinite asymptote in Fig. 207. This branch of the locus which approaches the infinite asymptote has no physical significance.

Figure 207 is drawn for the rather large value of η_1^2 and η_2^2 of 0.1. Since the values of η_1 and η_2 are equal, the locus of max. max. I_2 is a 45-deg. line through the point (1, 1). The positions of the max. max. I_2 are given by the intersections of this locus and the max. 2-1 line, as shown in Fig. 207. Similarly, the max. max. E_{c2} occurs at the intersection of the locus for max. max. E_{c2} and the max. 2-1 line. Except for the intersection at the point (1, 1), there is only *one* point of intersection in the particular case shown. Referring to Eq. (405), since X_2 is not a factor, the point (1, 1) is not a point of maximum or minimum E_{c2}. Equation (405) has three roots for X_2, two of which may mark the positions of max. max. E_{c2} and the third the position of a min. max. E_{c2}. If two of the roots are imaginary, there is but *one maximum*, as in Fig. 207.

Figure 208 shows the locus of max. max. E_{c2} for $\eta_1^2 = \eta_2^2 = 0.05$, and the max. 2-1 line for $\tau = 0.6$. In this case, all three roots of Eq. (405) are real, two giving the positions of max. max. E_{c2}, as indicated in the figure. The other root gives the position of min. max. E_{c2}.

If η_2 is made smaller than the value assumed in the two charts just described, the two branches of the locus of max. max. E_{c2} approach, as in Fig. 209, drawn for $\eta_1^2 = \eta_2^2 = 0.001$. The locus, for all practical purposes, consists of the line $\lambda_2/\lambda_0 = 1$ and a curved line whose equation can be obtained from Eq. (406-*R*) by neglecting η_2^2. This simplified approximate equation is

Approx. equation of locus of max. max. E_{c2} when $\eta_2^2 << \beta_2^2$

$$\begin{cases} X_1 = \dfrac{R_1}{R_2} \cdot \dfrac{L_2\omega_0}{L_2\omega_0 - 2X_2} & \text{or} \quad (407) \\ \beta_1 = \dfrac{\eta_1}{\eta_2} \cdot \dfrac{\beta_2}{1 - 2\beta_2} & (407\text{-}R) \end{cases}$$

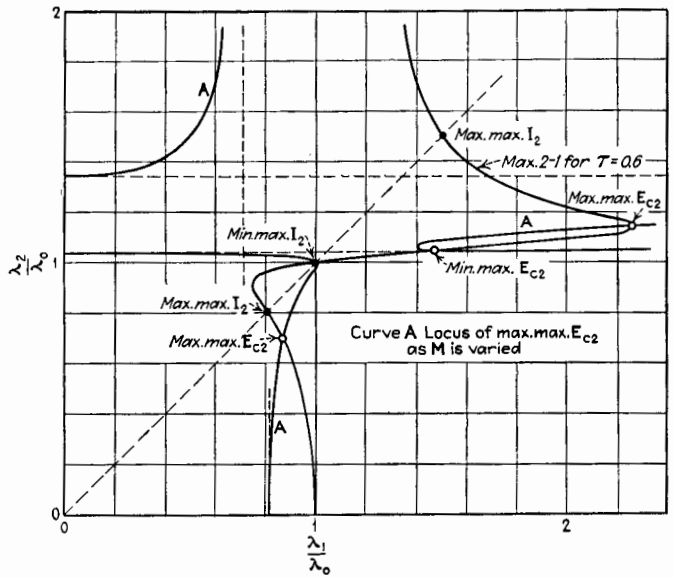


FIG. 208.—Locus of max. max. E_{c2} as M is varied and conditions for max. max. E_{c2} when $\tau = 0.6$. $\eta_1^2 = \eta_2^2 = 0.05$.

This equation is valid for small values of η_2^2, provided β_1 and β_2 are not close to zero.

147. Value of Max. Max. Secondary Voltage.—The value of max. max. E_{c2} for values of coupling greater than critical coupling can be found by substituting Eq. (405) in Eq. (404), giving

$$\text{Max. max. } E_{c2} = \frac{E_0 L_2\omega_0}{2\sqrt{R_1R_2}}\left(1 + \frac{\eta_2^2}{\beta_2}\right)\sqrt{1 - \frac{2(\eta_2^2 + \beta_2^2)}{\eta_2^2 + \beta_2}} \quad (408)$$

If η_2^2 is negligible compared with β_2, this equation reduces to the approximate form

$$\text{Approx. max. max. } E_{c2} = \frac{E_0 L_2 \omega_0}{2\sqrt{R_1 R_2}} \sqrt{2\frac{\lambda}{\lambda_2^2} - 1} \qquad (409)$$

Because of the radical, max. max. E_{c2} is greater at the point for which λ_2/λ_0 is less than unity. This would be expected, because at this point the secondary condenser is smaller than at the other point of intersection for max. max. E_{c2}.

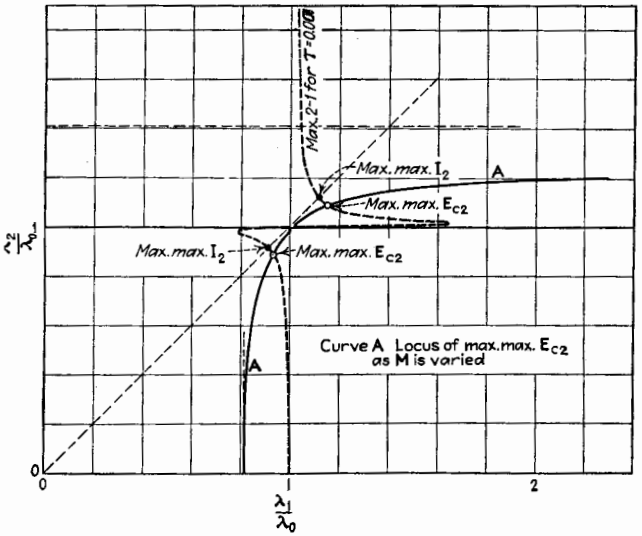


FIG. 209.—Locus of max. max. E_{c2} as M is varied and conditions for max. max. E_{c2} when $\tau = 0.001$. $\eta_1^2 = \eta_2^2 = 0.001$.

For values of τ less than for critical coupling, as defined by Eq. (391), there is one position of max. max. E_{c2}. This single max. max. E_{c2} occurs when λ_1/λ_0 and λ_2/λ_0 are each less than unity. At critical coupling, the max. 2-1 line and the locus of max. max. E_{c2}, as given by Eq. (407), have the same slope at the point ($\lambda_1/\lambda_0 = 1$, $\lambda_2/\lambda_0 = 1$) and do not cross. There is an intersection point at a value of λ_2/λ_0 less than unity.

II. THEORY OF MAGNETICALLY COUPLED CIRCUITS WITH REGENERATION IN THE SECONDARY CIRCUIT

148. Derivation of the Currents in Coupled Circuits with Regeneration.—With the theory of coupled circuits before us,

consider how this theory is affected when a regenerative tube is associated with the secondary circuit. The connections are shown in Fig. 210, and the equations for these circuits are

$$\left.\begin{aligned} \mathbf{Z}_1\Delta\mathbf{I}_1 + jM\omega_0\Delta\mathbf{I}_2 &= \Delta\mathbf{E}_0 \\ jM\omega_0\Delta\mathbf{I}_1 + \mathbf{Z}_2\Delta\mathbf{I}_2 - jM_b\omega_0\Delta\mathbf{I}_p &= 0 \\ j\left(\frac{u_p}{C_2\omega_0} - M_b\omega_0\right)\Delta\mathbf{I}_2 + \mathbf{z}_p\Delta\mathbf{I}_p &= 0 \end{aligned}\right\} \tag{410}$$

where

$$\mathbf{Z}_1 = R_1 + j\left(L_1\omega_0 - \frac{1}{C_1\omega_0}\right) = R_1 + jX_1$$

$$\mathbf{Z}_2 = R_2 + j\left(\mathrm{L}_2\omega_0 - \frac{1}{C_2\omega_0}\right) = R_2 + jX_2$$

and

$$\mathbf{z}_p = r_p + jL_b\omega_0.$$

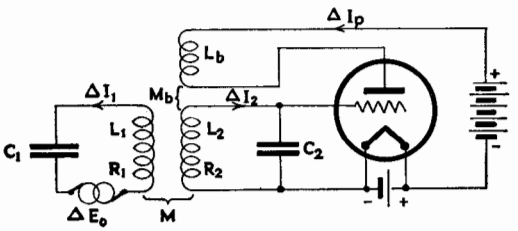


FIG. 210.—Magnetically coupled circuits with regeneration in the secondary circuit.

As before, $\omega_0/2\pi$ is the frequency of the e.m.f. ΔE_0 impressed in the primary circuit. All currents and voltages in Part II are limited to small values in order that u_p and r_p may be assumed constant. Small values are indicated by the delta sign.

The solution of Eq. (410) is

$$\Delta\mathbf{I}_1 = \frac{(\Delta\mathbf{E}_0)\left(\mathbf{Z}_2 - \dfrac{\dfrac{u_pM_b}{C_2} - M_b^2\omega_0^2}{\mathbf{z}_p}\right)}{\mathbf{Z}_1\left(\mathbf{Z}_2 - \dfrac{\dfrac{u_pM_b}{C_2} - M_b^2\omega_0^2}{\mathbf{z}_p}\right) + M^2\omega_0^2} \tag{411}$$

$$\Delta\mathbf{I}_2 = \frac{-jM\omega_0\Delta\mathbf{E}_0}{\mathbf{Z}_1\left(\mathbf{Z}_2 - \dfrac{\dfrac{u_pM_b}{C_2} - M_b^2\omega_0^2}{\mathbf{z}_p}\right) + M^2\omega_0^2} \tag{412}$$

A comparison of Eqs. (411) and (412) with Eqs. (385) and (386) shows that the mutual inductance M_b, which permits regeneration, affects *only* the impedance $\mathbf{Z}_2$ of the circuit to which the regenerative tube is directly coupled.

The equivalent impedance of the secondary circuit, therefore, is

$$[\mathbf{z}_2] = \mathbf{Z}_2 - \frac{\frac{u_p M_b}{C_2} - M_b^2\omega_0^2}{\mathbf{z}_p} \tag{413}$$

$$= \left[R_2 - \frac{\frac{u_p M_b}{C_2} - M_b^2\omega_0^2}{z_p^2} \cdot r_p\right] + j\left[X_2 + \frac{\frac{u_p M_b}{C_2} - M_b^2\omega_0^2}{z_p^2} \cdot L_b\omega_0\right] \tag{414}$$

$$= (R_2 - hr_p) + j(X_2 + hL_b\omega_0) \tag{415}$$

$$= [r_2] + j[x_2] \tag{416}$$

The dimensionless quantity h, known as the *coefficient of regeneration*, has been introduced in Eq. (415) simply as an abbreviation, where

$$h = \frac{\frac{u_p M_b}{C_2} - M_b^2\omega_0^2}{r_p^2 + L_b^2\omega_0^2} \tag{417}$$

Using this new symbol, regeneration in effect changes the resistance of the secondary circuit from R_2 to $[r_2]$, where

$$[r_2] = R_2 - hr_p \tag{418}$$

and changes the reactance of the secondary circuit from X_2 to $[x_2]$, where

$$[x_2] = X_2 + hL_b\omega_0 \tag{419}$$

In accordance with Eq. (419), the change in the reactance due to regeneration may be considered to be due to a change in the equivalent inductance of the circuit from L_2 to $[l_2]$, where $[l_2] = L_2 + hL_b$. Corresponding to the equivalent inductance $[l_2]$, there is an equivalent angular velocity $[\omega_2]$, wave length $[\lambda_2]$, and couplings $[\tau]$ and $[\tau_b]$. As with regeneration in a single circuit, the change in reactance is very small in comparison with the change in resistance. Hence, the equivalent values $[l_2]$, $[\omega_2]$, $[\lambda_2]$, $[\tau]$, and $[\tau_b]$ are in general practically the same as L_2, ω_2, λ_2, τ, and τ_b.

Using the equivalent values, Eq. (412) can be written in much simpler form as follows:

$$\Delta \mathbf{I}_2 = \frac{-jM\omega_0 \Delta \mathbf{E}_0}{\mathbf{Z}_1[\mathbf{z}_2] + M^2\omega_0^2} = \frac{-jM\omega_0 \Delta \mathbf{E}_0}{[\mathbf{z}_2]\left(\mathbf{Z}_1 + \dfrac{M^2\omega_0^2}{[\mathbf{z}_2]}\right)} \tag{420}$$

Equation (420) corresponds to Eq. (386) which is for no regeneration. It must be remembered that the two cases, with and without regeneration, differ in that R_2 is always positive in Eq. (386), whereas $[r_2]$ in Eq. (420) can have any value either positive or negative. The study of coupled circuits, one of which is regenerated, resolves itself into the study of coupled circuits one of which can have a negative resistance. Unfortunately, $[r_2]$ is not usually constant but, being dependent upon h, varies if any of the factors of Eq. (417) change. We may safely assume that u_p, r_p, and L_b are constant if the electrical amplitudes are small, and that M_b may be held constant. The factor C_2 varies if the secondary circuit is tuned by varying its capacitance, and, since ω_0 is also a factor of Eq. (417), the value of h, and hence $[r_2]$, is a function of whatever frequency exists in the circuits.

We have shown in Chap. XIII that in most practical cases the terms $M_b^2\omega_0^2$ and $L_b^2\omega_0^2$ in Eq. (417) are small in comparison with the other terms. In order to simplify the following analysis, it is assumed that these terms are so small that they can be safely neglected. An approximate coefficient of regeneration will be used, defined by the relation

$$h' = \frac{u_p M_b}{C_2 r_p^2} \tag{421}$$

We must realize that there are two possible conditions.

The *first condition* deals with a constant regenerative effect, obtained by holding all factors of Eq. (421) constant. Since C_2 must be kept constant, the secondary circuit must be tuned by a variable inductance having no mutual inductance with L_1 or with L_b, or by a variable series capacitance not connected between the grid and filament of the regenerative tube. If h' is constant, $[r_2]$ is constant. The treatment for this condition is the same as for ordinary coupled circuits when extended to include a negative resistance in the secondary circuit and hence to allow for the possibility of self-oscillation.

The *second condition* simulates more nearly the practical case in which C_2 is the variable element used to tune the secondary circuit. For this condition, $[r_2]$ is variable in a way determined by Eqs. (418) and (421), giving

$$[r_2] = R_2 - \frac{u_p M_b}{C_2 r_p} \tag{422}$$

where C_2 is the variable factor which also determines $[\lambda_2]$, the natural wave length of the secondary circuit.

Condition I. Constant Regenerative Effect

The treatment for this condition follows closely that for coupled circuits with positive resistances. Since regeneration, which is here assumed constant, alters slightly the effective inductance of the secondary circuit, as shown by Eq. (419), all quantities relating to the secondary circuit will be expressed in equivalent values.

The secondary current is given by

$$\mathbf{\Delta I}_2 = \frac{-jM\omega_0 \mathbf{\Delta E}_0}{\mathbf{Z}_1\left[\left([r_2] + \frac{M^2\omega_0^2 R_1}{Z_1^2}\right) + j\left([x_2] - \frac{M^2\omega_0^2 X_1}{Z_1^2}\right)\right]} \tag{423}$$

This equation corresponds to Eq. (387) for the case of positive resistances and may be expressed in coefficients similar to Eq. (387-R). This second form will not be given here.

149. Conditions for and Value of Max. Secondary Current with Constant Regenerative Effect.—Referring to Eq. (423) and following the same procedure as before, if X_1 has some fixed value and $[x_2]$ is varied until a maximum for ΔI_2 is obtained, the value of $[x_2]$ is

$$\left.\begin{matrix}\textit{Max. 1-2 line.} \\ [x_2]\ \textit{to give max. } \Delta I_2\end{matrix}\right\} [x_2] = \frac{M^2\omega_0^2 X_1}{R_1^2 + X_1^2} \tag{424}$$

Equation (424) corresponds to Eq. (388) and is independent of the value of $[r_2]$, whether negative or positive.

The value of max. ΔI_2 is

$$\text{Max. } \mathbf{\Delta I}_2 = \frac{-jM\omega_0 \mathbf{\Delta E}_0}{(R_1 + jX_1)\left([r_2] + \frac{M^2\omega_0^2 R_1}{R_1^2 + X_1^2}\right)} \tag{425}$$

This expression gives a finite value of ΔI_2 even if $[r_2]$ is negative but is less in absolute value than $M^2\omega_0^2 R_1/Z_1^2$.

150. Conditions for Oscillation at Frequency $\omega_0/2\pi$ with Constant Regenerative Effect.—The conditions under which ΔI_2 has its greatest value can be found, as was done in the simple theory of coupled circuits, by differentiating Eq. (425). The roots thus found for X_1 are similar to Eq. (390) in Part I.

Conditions for max. max. ΔI_2 or min. max. ΔI_2

$$\begin{cases} X_1 = 0 & (426\text{-}1) \\ Z_1^2 = \dfrac{M^2\omega_0^2 R_1}{[r_2]} & (426\text{-}2) \\ Z_1^2 = -\dfrac{M^2\omega_0^2 R_1}{[r_2]} & (426\text{-}3) \end{cases}$$

Expression (426-2), which in Part I gave the value of X_1 for the max. max. ΔI_2, gives an imaginary value of X_1 when $[r_2]$ is negative. Therefore root (426-3) must then be used.

Examining more closely the conditions given in Eq. (426), it is evident that root (426-3) gives in general two values of X_1 for both of which the denominator of Eq. (425) is zero. For these two values of X_1, the current ΔI_2 (and also ΔI_1) becomes infinite or oscillation begins at frequency $\omega_0/2\pi$. Values of X_1, with corresponding values of $[x_2]$ according to Eq. (424), give finite values of ΔI_2 between these two points. Since the only adjustments under consideration lie on the max. 1-2 line shown in Figs. 198 and 199, this line included between the boundary points of oscillation at frequency $\omega_0/2\pi$ is the locus of max. ΔI_2, just as in Part I when the secondary resistance was positive. It should be noted that the shape of the max. 1-2 line is independent of whether $[r_2]$ is positive or negative. Also, the *same* two values of X_1 satisfy Eq. (390-2) that satisfy Eq. (426-3) if the negative $[r_2]$ has the same numerical value that the positive R_2 had in Part I.

As was demonstrated in Part I, the max. 2-1 line (adjustment of primary circuit to give max. ΔI_2 for various settings of the secondary circuit) passes through the points of max. max. ΔI_2. In this case of negative resistance, these points have been shown to be the boundary points between oscillation and non-oscillation. The portion of this max. 2-1 line between these two points is the locus of max. ΔI_2 of frequency $\omega_0/2\pi$, where the order of adjustment of the two circuits is first secondary reactance then primary reactance.

For points on the max. 1-2 line outside the region between the boundary points of oscillation, $[r_2] + \frac{M^2\omega_0^2 R_1}{Z_1^2}$ (see Eq. (423)) is negative and $[x_2] - \frac{M^2\omega_0^2 X_1}{Z_1^2} = 0$. This condition cannot exist. If the first expression were negative, a larger amount of power would be supplied to the system than would be dissipated, an obvious impossibility. What actually happens at first is an increase in magnitude of the oscillation. Then the vacuum tube operates over such a large region of its characteristic surface that the average values of u_p and r_p are altered. This alteration in both u_p and r_p is in a direction to make $[r_2] + \frac{M^2\omega_0^2 R_1}{Z_1^2}$ less negative. Equilibrium is established when this expression becomes zero. Harmonics are introduced, but as a first approximation in the qualitative discussion they may be neglected. The variations in u_p and r_p alter the value of $[x_2]$ also, but as a first approximation, Eq. (424) may still be considered true and the oscillation at frequency $\omega_0/2\pi$ will take place over the portions of the max. 1-2 line outside the boundary points.

If Eqs. (424) and (426-3), both of which hold for the two points of boundary between oscillation and non-oscillation, are solved simultaneously, there results

Locus of points where oscillation begins at frequency $\frac{\omega_0}{2\pi}$ as M is varied

$$\begin{cases} \dfrac{X_1}{[x_2]} = -\dfrac{R_1}{[r_2]} \text{ or} & (427) \\ \dfrac{\beta_1}{[\beta_2]} = -\dfrac{\eta_1}{[\eta_2]} & (427\text{-}R) \end{cases}$$

Equation (427) is evidently the same as Eq. (399) in Part I; *i.e.*, it is the equation of the straight lines in Figs. 198 and 201 and of the curved lines in Figs. 199 and 202. The boundary points given by Eq. (426-3) are found to be the points of intersection of the max. 1-2 line, the max. 2-1 line, and the appropriate radial line, as in Part I. In Part I these points of intersection were the points of max. max. ΔI_2. The current in the present case increases toward infinity instead of to a finite max. max. value.

If $M\omega_0$ has the value given by Eq. (428)

Critical coupling

$$\begin{cases} M^2\omega_0^2 = -R_1[r_2] \text{ or} & (428) \\ [\tau]^2 = -\eta_1[\eta_2] & (428\text{-}R) \end{cases}$$

a relation which gave critical coupling in Part I, the boundary

points come together at $X_1 = 0$, and $[x_2] = 0$. Relation (428) marks a sort of critical coupling for negative resistance in one circuit. The region of no oscillation at frequency $\omega_0/2\pi$ shrinks to a point and an infinite value of ΔI_2 at frequency $\omega_0/2\pi$ occurs only at this point. This should be the condition for maximum signal strength with greatest selectivity. Analysis shows that under this condition the max. 1-2 line and the locus of Eq. (427) are tangent to each other at the origin, as shown in Fig. 201.

We may summarize as follows: When two circuits are coupled, and one, say the secondary circuit, has regeneration so that its effective resistance is negative but less in absolute value than $M^2\omega_0^2 R_1/Z_1^2$, the locus called the max. 1-2 line, which marks the maximum value of ΔI_2 in simple coupled circuits, also marks the maximum value of current of frequency $\omega_0/2\pi$ when the secondary resistance is negative. As this locus is traversed, there are two values of X_1 which mark a range outside of which the system oscillates at frequency $\omega_0/2\pi$. Between the two values of X_1, all points on the max. 1-2 line give max. ΔI_2, except when the tube starts oscillating at some other frequency, a possibility which will be explained later.

Referring to Fig. 199, which may be used when $[r_2]$ has a negative value, oscillation at frequency $\omega_0/2\pi$ takes place if the adjustments correspond to points on the locus max. 1-2 outside the region between the intersections of the max. 1-2 line and the max. 2-1 line. Over the portions of the max. 1-2 line and the max. 2-1 line between these points of intersection, the current ΔI_2 is a maximum when the adjustments are made in the proper order. Maximum regeneration at frequency $\omega_0/2\pi$ occurs at the points of intersection of the max. 1-2 and max. 2-1 lines, *i.e.*, at the points a a.

If M is varied, the points a a travel along the locus

$$\frac{\eta_1}{[\eta_2]} = \pm\frac{\beta_1}{[\beta_2]}.$$

If M is fixed and $[r_2]$ is varied, the points a a travel along the max. 1-2 line which is unchanged in shape. If $[r_2]$ is positive, the points a a represent points of max. max. ΔI_2. As $[r_2]$ approaches zero from a positive value, the points a a recede from the origin O' along the max. 1-2 line until, when $[r_2] = 0$, the points are at $\lambda_1/\lambda_0 = 0$ and $\lambda_1/\lambda_0 = \infty$. As $[r_2]$ becomes negative, the points a a approach O' along the same max. 1-2 line. They now

mark the points where oscillation begins at frequency $\omega_0/2\pi$, or the points of maximum regeneration. At these points ΔI_2 is theoretically infinite.

151. Conditions for Oscillation at Any Frequency $\omega/2\pi$ with Constant Regenerative Effect.—Until now, there has been considered only the possibility of self-oscillation of the system at the same frequency as that of the impressed e.m.f. ΔE_0. Under certain conditions the system can maintain self-oscillation at some frequency other than $\omega_0/2\pi$, whether or not the e.m.f. ΔE_0 is impressed.

The whole system oscillates at a frequency $\omega/2\pi$ if the equivalent impedance of the whole system, as viewed from the circuit to which the regenerative tube is attached (the secondary circuit in this case), is zero for this frequency. Expressed analytically, the condition for oscillation is

$$\text{Condition for oscillation at frequency } \frac{\omega}{2\pi} \left\{ [\mathbf{z}_{21}] = [\mathbf{z}_2] + \frac{M^2\omega^2}{(\mathbf{Z}_1)_\omega} = 0 \right. \tag{429}$$

The total impedance is zero only when both the resistance and reactance of $[\mathbf{z}_{21}]$ are zero. Therefore, expanding Eq. (429), the conditions for oscillation are

$$\text{Conditions for oscillation at frequency } \frac{\omega}{2\pi} \begin{cases} [r_2] + \dfrac{M^2\omega^2 R_1}{R_1^2 + (X_1)_\omega^2} = 0 = [r_{21}] & (430) \\ [x_2] - \dfrac{M^2\omega^2 (X_1)_\omega}{R_1^2 + (X_1)_\omega^2} = 0 = [x_{21}] & (431) \end{cases}$$

Equations (430) and (431) may be expressed in ratio form as follows:

$$\text{Conditions for oscillation at frequency } \frac{\omega}{2\pi} \begin{cases} [\eta_2] + \dfrac{[\tau]^2\eta_1}{\eta_1^2 + \left(1 - \dfrac{\omega_1^2}{\omega^2}\right)^2} = 0 & (430\text{-}R) \\ 1 - \dfrac{[\omega_2]^2}{\omega^2} - \dfrac{[\tau]^2\left(1 - \dfrac{\omega_1^2}{\omega^2}\right)}{\eta_1^2 + \left(1 - \dfrac{\omega_1^2}{\omega^2}\right)^2} = 0 & (431\text{-}R) \end{cases}$$

In investigating the conditions for oscillation, it is immaterial whether or not the e.m.f. ΔE_0 is impressed. The frequency of ΔE_0 will be used as a reference frequency so that quantities can be plotted as before in terms of the ratios λ_1/λ_0 and $[\lambda_2]/\lambda_0$. For convenience the following abbreviations will be used:

$$\frac{\lambda_1}{\lambda_0} = \theta_1, \frac{\lambda_2}{\lambda_0} = \theta_2, \text{ and } \frac{\lambda}{\lambda_0} = \theta.$$

In Eqs. (430-*R*) and (431-*R*), the ratios $\eta_1 = R_1/L_1\omega$ and $[\eta_2] = [r_2]/[l_2]\omega$ are functions of ω rather than of ω_0 as in all previous equations. Although ω_0 is constant, ω is not. With most coils the resistance changes with frequency, η being nearly constant, so the quantities η_1 and $[\eta_2]$ will first be assumed constant. The treatment for constant η is presented as Case *a* under Condition I; Case *b*, under Condition I, in which the resistances are assumed constant rather than η, will be presented later.

152. Oscillation-boundary Curve. Case *a*, Constant η.— The boundary curve marking the border between oscillation and non-oscillation on the $\lambda_1/\lambda_0 - [\lambda_2]/\lambda_0$ plane can be found by eliminating ω from Eqs. (430-*R*) and (431-*R*). The equation of the *boundary curve* thus attained is

$$\left(r_1^2 + 1 + \tau^2\frac{\eta_1}{\eta_2}\right)\theta_1^4 + 2(\eta_1\eta_2 - 1 + \tau^2)\theta_1^2\theta_2^2 + \left(\eta_2^2 + 1 + \tau^2\frac{\eta_2}{\eta_1}\right)\theta_2^4 = 0$$

or

$$\theta_1^2 = \theta_2^2\frac{1 - \tau^2 - \eta_1\eta_2 \pm (\eta_1 + \eta_2)\sqrt{-\left(1 + \frac{\tau^2}{\eta_1\eta_2}\right)}}{1 + \tau^2\frac{\eta_1}{\eta_2} + \eta_1^2} \qquad (432\text{-}R)$$

where τ, η_2, and θ_2 are equivalent values, the brackets being omitted for simplicity.

Equation (432-*R*), when plotted on the $\theta_1 - [\theta_2]$ plane for a chosen set of values of η_1, $[\eta_2]$, and $[\tau]$, gives two straight lines radiating from the origin. When $\eta_1 = -[\eta_2]$ or when $[\tau]^2 = -\eta_1[\eta_2]$, Eq. (432-*R*) reduces to $\theta_1 = [\theta_2]$, a single line through O'. In Fig. 211, the boundary lines are plotted for $[\tau] = 0.5$, $\eta_1 = 0.4$, and $[\eta_2] = -0.2$. In the region *between* these boundary lines, the equivalent resistance $[r_{21}]$ of the system is positive for *all* frequencies and the system does not oscillate. For each point *on* the boundary lines, both the resistance $[r_{21}]$ and the reactance $[x_{21}]$ of the system, as viewed from the secondary circuit, vanish for some frequency which is different for each point, and the system oscillates at the frequency which makes $[r_{21}]$ and $[x_{21}]$ zero. For each point *outside* the region between the boundary lines, the resistance $[r_{21}]$ is negative for some

frequencies, one of which makes the reactance $[x_{21}]$ zero. Oscillations at this frequency build up until equilibrium is established, which occurs when the variation in tube coefficients reduces the effective resistance to zero. The region *outside* the two lines is then a region of oscillation.

The max. 1-2 line for $\theta = 1$ is also shown in Fig. 211. At all points of this max. 1-2 line the reactance $[x_{21}]$ is zero for the frequency $\omega_0/2\pi$. At the intersection points a a of the max. 1-2 line

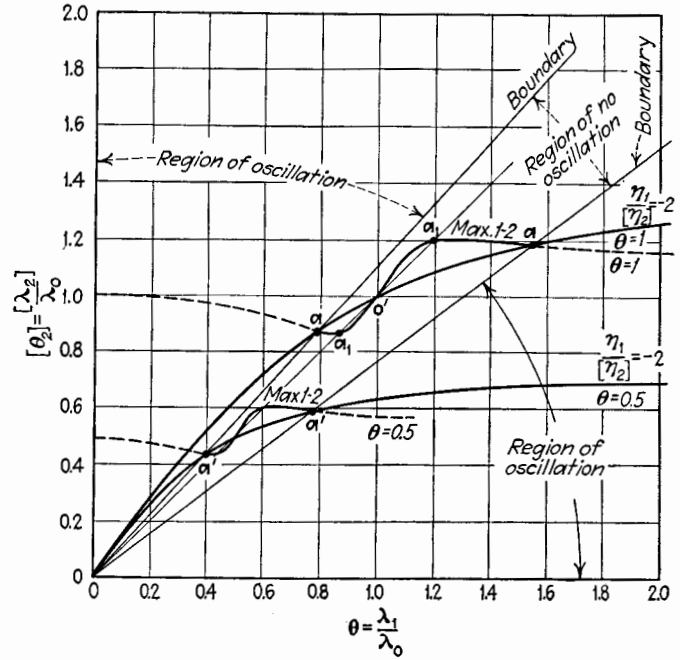


FIG. 211.—Oscillation boundary lines for constant regenerative effect. Condition of constant η. $[r] = 0.5$. $\eta_1 = 0.4$. $[\eta_2] = -0.2$.

and the boundary lines, the resistance $[r_{21}]$ is also zero and the system oscillates at a frequency $\omega_0/2\pi$, corresponding to a wave length λ_0. We may draw other max. 1-2 lines for other values of θ, as shown for $\theta = 0.5$, for example, in Fig. 211. Along this curve the reactance $[x_{21}]$ is zero for a frequency $2\omega_0/2\pi$, corresponding to a wave length of $0.5\lambda_0$, and at points $a'a'$ the resistance $[r_{21}]$ is also zero for this frequency, so that the system oscillates at these points at the frequency $\omega_0/2\pi\theta$. The system oscillates at all points except those in the region between the two

radiating lines. The frequency or wave length of oscillation for any point is determined by the value of θ for the max. 1-2 line which passes through the point in question.

As $[\eta_2]$ is varied, the points a a through which the boundary lines pass move along the max. 1-2 line. As $[\eta_2]$ is made more negative, the tendency to oscillate is greater and the points a a move toward O' so that the region of no oscillation decreases in area.

As $[\tau]$ is varied, the points a a move along the $\eta_1/[\eta_2]$ line, Eq. (427-R), the max. 1-2 line changing shape so that points a a are the intersection points of the three curves.

The max. 1-2 line has a slope of unity at the point O' when $\eta_1 = [\tau]$, and a greater slope when $\eta_1 < [\tau]$.

Consider now the two cases when $\eta_1 > [\tau]$ and when $\eta_1 < [\tau]$. Suppose, that $\eta_1 > [\tau]$. As $[\eta_2]$ is made more negative, points a a approach O' and the region of no oscillation collapses at critical coupling to a single 45-deg. line through O'. If, however, $\eta_1 < [\tau]$, the boundary lines become coincident at two values of $[\eta_2]$; first, when $[\eta_2] = -\eta_1$, for which condition points a a arrive at points a_1 a_1 on the 45-deg. line through O', Fig. 211; second, when $[\eta_2]$ satisfies the critical-coupling relation and points a a come together at O'. Referring to Fig. 211, the two boundary lines, which have just come together when the intersection points a a arrive at position a_1 a_1, pass each other as the intersection points move from a_1 a_1 toward O'. In this last range there is no region of no oscillation, but the two regions of oscillation overlap in a small diverging strip from O. As the intersection points travel along the $\theta = 1$ line from a_1 a_1 toward O, the overlapping region of oscillation first increases in area and then decreases in area until it vanishes at critical coupling. The conditions for this overlapping of areas of oscillation are, therefore, that $[\tau]$ must be greater than η_1 and also greater than critical coupling, and that $[\eta_2]$ must lie between $-\eta_1$ and $-[\tau]^2/\eta_1$.

153. Values of the Equivalent Resistance and Reactance.—We shall study further the system under the conditions of Fig. 211. As already explained, points a a are points of maximum regeneration for signals of frequency $\omega_0/2\pi$. Suppose that the circuits are set to correspond to the lower a-point, having coordinates $\theta_1 = 0.794$ and $\theta_2 = 0.880$. Examine first, at various frequencies in the neighborhood of $\omega_0/2\pi$, the equivalent resistance $[r_{21}]$ and the equivalent reactance $[x_{21}]$ of the system as

viewed from the secondary circuit. The expression for $[r_{21}]$ is

$$\frac{[r_{21}]}{[l_2]\omega} = [\eta_2] + \frac{[\tau]^2\eta_1}{\eta_1^2 + \left(1 - \frac{\theta^2}{\theta_1^2}\right)^2} \qquad (433\text{-}R)$$

and for $[x_{21}]$ is

$$\frac{[x_{21}]}{[l_2]\omega} = 1 - \frac{\theta^2}{[\theta_2]^2} - \frac{[\tau]^2\left(1 - \frac{\theta^2}{\theta_1^2}\right)}{\eta_1^2 + \left(1 - \frac{\theta^2}{\theta_1^2}\right)^2} \qquad (434\text{-}R)$$

These are plotted against $\theta = \lambda/\lambda_0$ in Fig. 212. The curves show that only at the point $\theta = 1$, *i.e.*, for a wave length equal to λ_0,

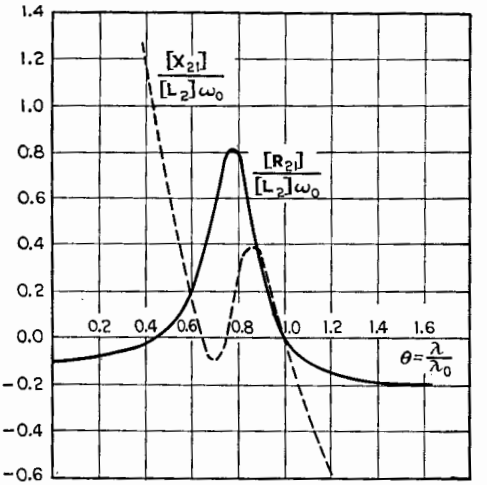


FIG. 212.—Resistance and reactance of coupled circuits as viewed from the secondary circuit. Conditions: $\theta_1 = 0.794$; $[\theta_2] = 0.880$; $\eta_1 = 0.4$; $[\eta_2] = -0.2$; $[\tau] = 0.5$.

are both $[r_{21}]$ and $[x_{21}]$ equal to zero, at which point the system just oscillates.

We may also plot the equivalent resistance $[r_{12}]$ and reactance $[x_{12}]$ of the system, viewed from the primary circuit, as θ is varied, when the system is set for the same point a. The results are shown in Fig. 213. These curves give the resistance and reactance of the system offered to any signal of wave length $\theta\lambda_0$ impressed in the primary circuit. When the impressed wave length is λ_0, the resistance and reactance are zero and maximum

regeneration occurs. It is interesting to note that, if θ is between 0.73 and 1.0 in the example given, the system offers a *negative* resistance. The physical meaning of this is that the system supplies energy instead of absorbing energy at these wave lengths, but this occurs *only* when a signal is impressed. If the primary circuit is the antenna circuit of a coupled-circuit receiver, if the circuits are set to receive a signal of wave length λ_0, and if another signal of wave length between $0.73\lambda_0$ and λ_0 is impressed on the antenna, the system reradiates power of the impressed wave length in proportion to the strength of the impressed signal. The phase of the emitted power is determined by the values of

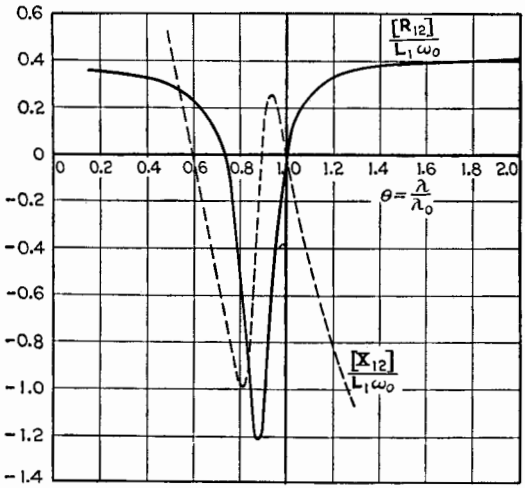


FIG. 213.—Resistance and reactance of coupled circuits as viewed from the primary circuit. Conditions: $\theta_1 = 0.794$; $[\theta_2] = 0.880$; $\eta_1 = 0.4$; $[\eta_2] = -0.2$; $[\tau] = 0.5$.

$[r_{12}]$ and $[x_{12}]$. In some regions around the antenna the direct and reradiated waves reinforce each other, while in other regions they partially neutralize each other.

154. Changes of Frequency of Oscillation of Secondary Circuit Caused by Tuning Primary Circuit for Case *a*. Drag Loop. Detection of Resonance.—It is of interest to examine the frequency of response of the system as θ_1 of the primary circuit is varied, the secondary circuit being set for $[\theta_2] = 1$. From Eq. (434-*R*) equated to zero, we may plot θ against θ_1, when $[\theta_2]$ is given the value unity. As the equation is a cubic in θ^2, it is necessary to solve for θ_1^2, in terms of θ^2. The result is

$$\frac{1}{\theta_1^2} = \frac{1}{\theta^2}\left(1 - \frac{[\tau^2] \pm \sqrt{[\tau]^4 - 4\eta_1^2(1-\theta^2)^2}}{2(1-\theta^2)}\right) \qquad (435\text{-}R)$$

Equation (435-R) is shown in Fig. 214 for $\eta_1 = 0.4$ and for four values of $[\tau]$.

Although the full-line curves of Fig. 214 give the values of θ *vs.* θ_1 for which the reactance of the system is zero as viewed from the secondary circuit, they alone do not tell the limits of the regions of oscillation. These limits are given by Eq. (433-R) equated to zero. The same results may be obtained more easily

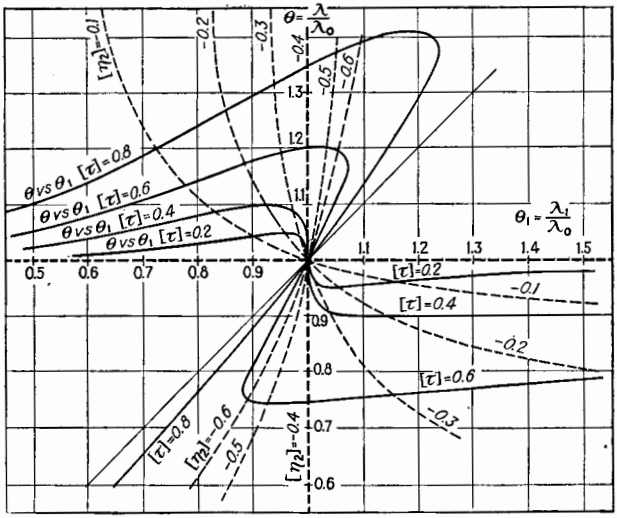


FIG. 214.—Variation of the wave length of oscillation as the primary circuit is varied. $\eta_1 = 0.4$.

by combining Eqs. (433-R) and (434-R). Eliminating $[\tau]^2$ from these equations, both equated to zero, gives

$$\theta^2 = \frac{\eta_1 + [\eta_2]}{\dfrac{\eta_1}{[\theta_2]^2} + \dfrac{[\eta_2]}{\theta_1^2}} \qquad (436\text{-}R)$$

This is shown plotted in the curves marked $\eta_1/[\eta_2] = -2$, Fig. 211, and for $[\theta_2] = 1$ in the dotted curves of Fig. 214, for various values of $[\eta_2]$. These curves give the loci of the intersections of the curves of Eqs. (433-R) and (434-R), both equated to zero, as $[\tau]$ varies. Hence, they give the loci of the limits of oscillation.

The interpretation of Fig. 214 is as follows: Suppose that we are making accurate comparisons of frequency or wave lengths of two circuits, such as two wavemeters P_1 and P_2. We shall assume that the secondary circuit of Fig. 210, is an oscillating reference circuit used in this accurate comparison of the two wavemeters. This secondary circuit S, when unaffected by either wavemeter, is oscillating at wave length $\lambda_0(\theta = 1)$. The common procedure in the comparison of the circuits P_1 and P_2 is to bring P_1 into proximity with S, and tune P_1 until some indication in S shows that P_1 is in resonance with S. The same process is applied to P_2. The two settings of P_1 and P_2 are then taken to be the settings corresponding to wave length λ_0. The indicator to detect resonance of P_1 or P_2 with S may be an ammeter in the plate or grid circuit of the triode attached to S, a telephone receiver in the plate circuit of the triode, or a second oscillating circuit arranged to beat with S so that any change in the frequency of S can be detected by a change in an audible beat note. This last method is by far the most accurate. In applying Fig. 214, it should be realized that the value of η_1 assumed in plotting the figure is excessively large and that certain limitations, as explained, are placed on the theory. But, qualitatively, the results of the analysis are of practical value.

Suppose that P_1 is coupled to S with a coefficient of coupling $[\tau] = 0.2$ and that $[\eta_2]$ is set at -0.1. As θ_1 is varied (the natural period of P_1 is varied) from 0.5 upward, the frequency of oscillation decreases because of the increase of θ along the full-line curve marked $[\tau] = 0.2$. At $\theta_1 = 0.9$, the oscillation stops. It begins again at a higher frequency than $\omega_0/2\pi$ when θ_1 is 1.145. The two points where oscillation stops may be indicated by clicks in the telephone receivers or by abrupt changes in the ammeter reading. Often, the mid-point between the two values of θ_1 corresponding to these two points is assumed as the point of resonance. Such an assumption is in error, as is shown by the fact that the point $\theta_1 = 1$ is not midway between these two critical points. Suppose now that the same discussion is repeated with $[\eta_2]$ set at -0.35. Since the dotted curve for $[\eta_2] = -0.35$ does not intersect the full-line curve for $[\tau] = 0.2$, Fig. 214, except at the point $\theta_1 = 1$ and $\theta = 1$, the system does not stop oscillating. The frequency of oscillation goes through the changes indicated by the curve of Fig. 214. The beat note used as indicator varies but returns at $\theta_1 = 1$ to the beat note

unaffected by the secondary circuit. This method of beat-note indication gives a very accurate determination of resonance; the steeper the full-line curve at the point (1, 1) the more accurate the determination. The steepness can be increased by increasing $[\tau]$ up to the value η_1, at which value the reactance curve is vertical as is shown by the curve marked $[\tau] = 0.4$ in Fig. 214. But with such a value of $[\tau]$, $[\eta_2]$ must be more negative than $-\eta_1$ in order that the system may oscillate throughout the range of θ_1. If the ammeter method of detection is used, the most rapid change in ammeter reading is obtained when $[\tau]$ is equal to or slightly less than η_1, and $[\eta_2]$ is about equal to $-\eta_1$, for then the system almost stops oscillating at $\theta_1 = 1$.

If $[\tau]$ is increased to 0.6 and $-[\eta_2]$ is equal to η_1, (*i.e.*, 0.4 in this example), oscillation persists along the curve $\tau = 0.6$, Fig. 214, as θ_1 is increased to unity. At this point the oscillation frequency suddenly changes from the value corresponding to $\theta = 1.21$ to the value corresponding to $\theta = 0.742$, and thereafter, as θ_1 is increased, follows the lower portion of the full-line curve marked $[\tau] = 0.6$. As θ_1 is varied back and forth through unity, this abrupt change in frequency occurs each time at $\theta_1 = 1$ and a telephone receiver indicates by a click this point of abrupt change in frequency. These conditions are best for the telephone method of detecting resonance.

Suppose that $[\tau]$ is 0.6 as before, but $[\eta_2]$ is set at -0.5. The upper full-line curve is followed as θ_1 is increased until the intersection point with the dotted curve is reached at $\theta_1 = 1.033$. Here, the frequency changes abruptly to that corresponding to the lower portion of the curve at this value of θ_1. As θ_1 is still further increased, the frequency follows the lower curve. If θ_1 is now decreased, the lower curve is followed until the intersection point at $\theta_1 = 0.927$ is reached, when the frequency jumps to values corresponding to points on the upper section of the full-line curve of Fig. 214. Thus a sort of hysteresis occurs in the variation in frequency, often called a *drag loop*, and is shown in Fig. 215 for $[\tau] = 0.6$ and $[\eta_2] = -0.5$. Such a drag loop can occur only when there are overlapping regions of oscillation, as already explained in connection with Fig. 213, *i.e.*, only when $-[\eta_2] < \eta_1$ and when $[\tau] > \eta_1$. The drag loop is of considerable importance in the study of power oscillators.

155. Conditions for Oscillation at Any Frequency $\omega/2\pi$ with Constant Regenerative Effect. Case *b*, Constant Resistance.—The theory thus far given applies to circuits in which the resist-

ance is proportional to the frequency for the range of frequencies concerned in the analysis. This assumption is approximately true if the range of frequencies considered is small, as is true if the resistances and couplings are small as in most practical cases. The values of η and τ used in the examples just cited are unusually large, so that the frequency range in any practical case would be much less than that used in these examples.

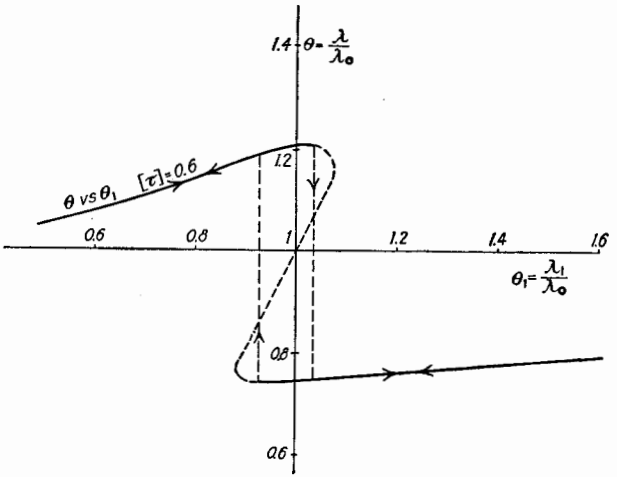


FIG. 215.—Drag loop. $[\eta_2] = -0.5$. $\eta_1 = 0.4$. $[\tau] = 0.6$.

In order to show the limitation of the theory due to the assumption of constant η, the other extreme of constant resistance will be considered as Case *b*. Every practical case lies somewhere between these two extremes.

If we assume that the resistances R_1 and $[r_2]$ remain constant, η_1 and η_2 will signify the ratios $R_1/L_1\omega_0$ and $[r_2]/[l_2]\omega_0$. Equations (430-*R*) and (431-*R*) expressed in terms of θ_1, $[\theta_2]$, and θ become

$$[\eta_2] + \frac{[\tau]^2\eta_1}{\eta_1^2\theta^2 + \left(1 - \frac{\theta^2}{\theta_1^2}\right)^2} = 0 \qquad (437\text{-}R)$$

$$1 - \frac{\theta^2}{[\theta_2]^2} - \frac{[\tau]^2\left(1 - \frac{\theta^2}{\theta_1^2}\right)}{\eta_1^2\theta^2 + \left(1 - \frac{\theta^2}{\theta_1^2}\right)^2} = 0 \qquad (438\text{-}R)$$

The theory for this case of constant resistance will be developed from these equations.

156. Oscillation-boundary Curve. Case *b*, Constant Resistance.— The boundary curve may be found for the condition of constant resistance by eliminating from Eqs. (437-*R*) and (438-*R*) the quantity θ which is the only quantity containing ω. The

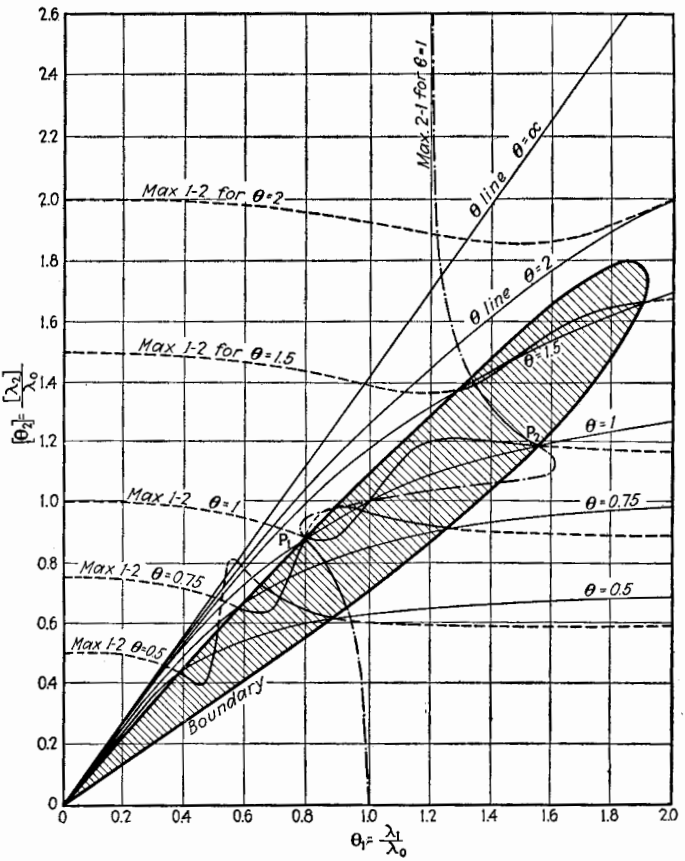


FIG. 216.—Oscillation-boundary curve for constant regenerative effect. Condition of constant resistances. $\eta_1 = 0.4$. $[\eta_2] = -0.2$. $[\tau] = 0.5$.

elimination gives the following expression for the boundary line between oscillation and non-oscillation.

$$\left[\eta_2(\eta_1+\eta_2)\theta_2^2+\frac{\eta_2}{\eta_1}+\tau^2\right]\theta_1^4+\left[\eta_2^2\left(1+\frac{\eta_2}{\eta_1}\right)\theta_2^2-2\frac{\eta_2}{\eta_1}(1-\tau^2)\right]\theta_1^2\theta_2^2 + \left[\frac{\eta_2}{\eta_1}\left(1+\tau^2\frac{\eta_2}{\eta_1}\right)\right]\theta_2^4=0 \quad (439\text{-}R)$$

where η_2, θ_2, and τ are the equivalent values. This boundary equation reduces to the straight line $\theta_1 = \theta_2$ when $[\eta_2] = -\eta_1$. This was also the case with the boundary Eq. (432-*R*). At critical coupling for $\theta = 1$, Eq. (439-*R*) does not reduce to a straight line. This is not surprising because critical coupling is now defined by the relation $[\tau]^2 = -\eta_1[\eta_2]\theta^2$. Critical coupling can exist for only one value of θ for any given values of M, L_1, and L_2.

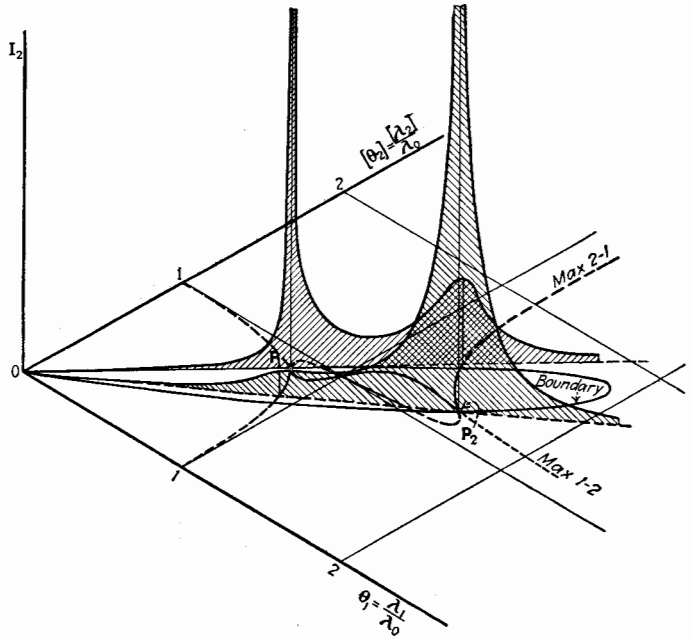


FIG. 217.—Sections of space model for secondary current. $[R_{21}] = 0$. $\eta_1 = 0.4$. $[\eta_2] = -0.2$. $[\tau] = 0.5$.

The boundary Eq. (439-*R*) is plotted in Fig. 216 for $[\tau] = 0.5$, $\eta_1 = 0.4$, and $[\eta_2] = -0.2$. This figure should be compared with Fig. 211 where η is constant, but the other conditions are the same. Figure 217 shows sections through the space model which give the approximate secondary current when the circuits are set at either of the points *a a* of maximum regeneration and the incoming wave length is varied. This figure is applicable also to the case of constant η if the boundary curve is replaced by the straight boundary lines shown in Fig. 211.

Figure 218 is similar to Fig. 216 except that it is calculated for critical coupling when $\theta = 1$. The numerical value of this coupling for the constants given is $\sqrt{0.08}$. In this case the two points a a of maximum regeneration for $\theta = 1$ of Fig. 216 have approached each other and have fused to a single point.

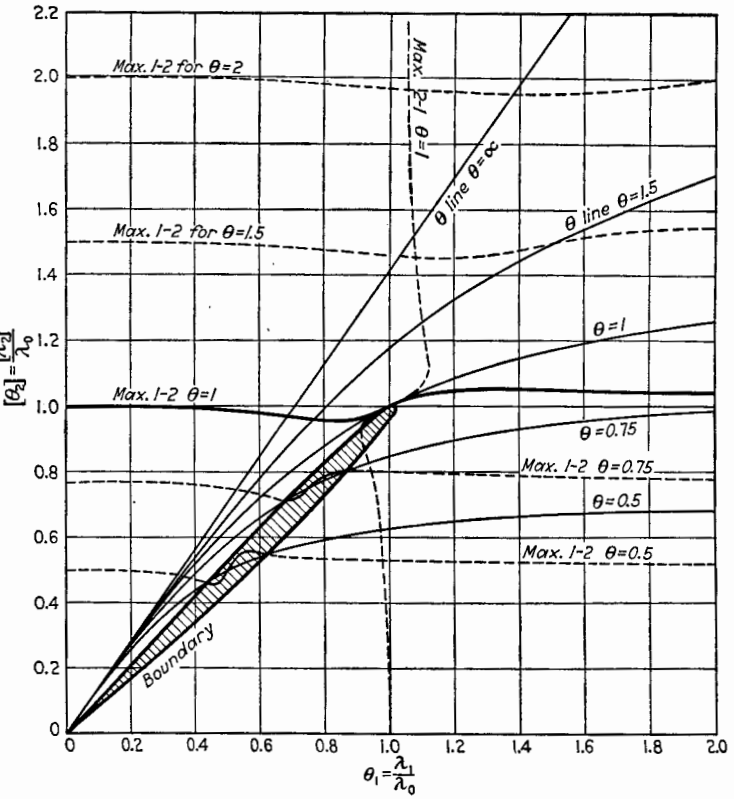


FIG. 218.—Oscillation-boundary curve for constant regenerative effect. Condition of constant resistances. Critical coupling. $\eta_1 = 0.4$. $[\eta_2] = -0.2$. $[\tau]^2 = 0.08$.

157. Changes of Frequency of Oscillation of Secondary Circuit When Tuning the Primary Circuit. Case *b*.—The equation corresponding to Eq. (435-*R*), which gives the variation of θ along the $[\theta_2] = 1$ line, is obtained for the condition of constant resistance by solving Eq. (438-*R*). The result is

$$\frac{1}{\theta_1^2} = \frac{1}{\theta^2}\left(1 - \frac{[\tau]^2 \pm \sqrt{[\tau]^4 - 4\eta_1^2\theta^2(1-\theta^2)^2}}{2(1-\theta^2)}\right) \qquad (440\text{-}R)$$

This equation differs from Eq. (435-*R*) only in the presence of an additional θ^2 under the radical. Qualitatively, the graph of Eq. (440-*R*) is similar to that of Fig. 214.

The elimination of $[\tau]$ from Eqs. (437-*R*) and (438-*R*) gives exactly the same equation as is obtained for constant η, *i.e.*, Eq. (436-*R*). The dotted lines of Fig. 214 are therefore the same for Cases *a* and *b*.

Figures 212 and 213 are drawn for $\theta = 1$ and hence these two figures and the discussion of them are applicable to the present case of constant resistance as well as to the former case of constant η.

Condition II. Variable Regenerative Effect

The two cases just discussed were calculated on the assumption that the coefficient of regeneration is constant. This is approximately true if the secondary circuit is tuned by a reactance which is not a part of the coupling either to the primary circuit or to the triode. Generally, however, both circuits are tuned by the variable condensers, C_1 and C_2 of Fig. 210, the voltage across C_2 being the grid voltage of the triode. Under these conditions the coefficient of regeneration h varies as the secondary circuit is tuned. The purpose of the following discussion is to indicate approximately such alterations in the theory of the first condition as are necessary because of this varying regeneration.

Using the approximate expression for the coefficient of regeneration as given in Eq. (421), replace the variable C_2 by its equivalent in terms of ω. Hence,

$$h' = \frac{u_p M_b L_2 \omega^2}{r_p^2} \tag{441}$$

$$= \frac{u_p \tau_b}{\eta_p^2} \sqrt{\frac{L_2}{L_b}} \cdot \frac{1}{\theta_2^2} \cdot \frac{L_2}{L_b} \tag{441-R}$$

where $\eta_p = r_p / L_b \omega_0$ and $\tau_b = M_b / \sqrt{L_2 L_b}$.
Equation (441-*R*) may be written

$$h' = \frac{f}{\theta_2^2} \cdot \frac{L_2}{L_b} \tag{442-R}$$

where the factor f is the term $\frac{u_p \tau}{\eta_p^2} \sqrt{\frac{L_2}{L_b}}$.

158. Conditions for and Value of Max. Secondary Current with Variable Regenerative Effect.—The secondary current due to the e.m.f. ΔE_0 is given by Eq. (423). Since θ_2 occurs in $[r_2]$ as well as in $[x_2]$, the equation of the max. 1-2 line cannot be obtained by equating $[x_{21}]$ to zero. It is necessary to express the absolute value of ΔI_2 and then to put its derivative with respect to θ_2 equal to zero. The absolute value of ΔI_2, expressed in ratios, is

$$\Delta I_2 = \frac{\dfrac{\tau \Delta E_0}{\sqrt{L_1 L_2 \omega_0}}}{\sqrt{\eta_1^2 + \left(1 - \dfrac{1}{\theta_1^2}\right)^2}\sqrt{\left[\eta_2 - \dfrac{f\eta_p}{\theta_2^2} + \dfrac{\tau^2 \eta_1}{\eta_1^2 + \left(1 - \dfrac{1}{\theta_1^2}\right)^2}\right]^2 + \left[1 - \dfrac{1-f}{\theta_2^2} - \dfrac{\tau^2\left(1 - \dfrac{1}{\theta_1^2}\right)}{\eta_1^2 + \left(1 - \dfrac{1}{\theta_1^2}\right)^2}\right]^2}} \qquad (443\text{-}R)$$

Where $\eta_1 = R_1/L_1\omega_0$ and $\eta_2 = R_2/L_2\omega_0$. In the discussion of Condition I, $[\eta_2]$ denoted the effective η of the secondary circuit *including* the effect of regeneration. In the present case, it is important to note that the effective η of the secondary circuit, including the effect of regeneration, is $[\eta_2] = \eta_2 - \dfrac{f\eta_p}{\theta_2^2}$ and varies with θ_2.

Equating to zero the derivative of ΔI_2 with respect to θ_2 gives the following equation for the max. 1-2 line. This gives the locus of max. ΔI_2 when the order of adjustment is first primary and then secondary. The locus is

$$1 - f + f\eta_2\eta_p - \frac{(1-f)^2 + f^2\eta_p^2}{\theta_2^2} - \frac{\tau^2\left(1 - f - f\eta_1\eta_p - \dfrac{1-f}{\theta_1^2}\right)}{\eta_1^2 + \left(1 - \dfrac{1}{\theta_1^2}\right)^2} = 0 \qquad (444\text{-}R)$$

The equation for the max. 2-1 line can be obtained by the simple process of setting $[x_{12}]$ equal to zero, since $[r_{12}]$ does not contain θ_1. The equation is

$$1 - \frac{1}{\theta_1^2} - \frac{\tau^2\left(1 - \dfrac{1-f}{\theta_2^2}\right)}{\left(\eta_2 - \dfrac{f\eta_p}{\theta_2^2}\right)^2 + \left(1 - \dfrac{1-f}{\theta_2^2}\right)^2} = 0 \qquad (445\text{-}R)$$

At any frequency $\omega/2\pi$, the values of the equivalent resistance and reactance of the system, as viewed from the secondary circuit to which the regenerative tube is attached, are, when expressed in ratios,

$$\frac{r_{21}}{L_2\omega_0} = \eta_2 - \frac{f\eta_p}{\theta_2^2} + \frac{\tau^2\eta_1}{\eta_1^2\theta^2 + \left(1 - \dfrac{\theta^2}{\theta_1^2}\right)^2} \qquad (446\text{-}R)$$

$$\frac{x_{21}}{L_2\omega_0} = 1 - \frac{\theta^2}{\theta_2^2} + \frac{f}{\theta_2^2} - \frac{\tau^2\left(1 - \dfrac{\theta^2}{\theta_1^2}\right)}{\eta_1^2\theta^2 + \left(1 - \dfrac{\theta^2}{\theta_1^2}\right)^2} \qquad (447\text{-}R)$$

These equations assume that all resistances are constant. If it is assumed that R_1 and R_2 vary proportionally with ω, so that η_1 and η_2 remain constant, Eqs. (446-R) and (447-R) are altered in the following respects: the θ^2 following η_1 in the denominators is omitted and the term $f\eta_p/\theta_2^2$ becomes $f\eta_p\theta/\theta_2^2$. Only the case of constant resistance will be considered.

159. Conditions for Oscillation at Any Frequency $\omega/2\pi$. Oscillation-boundary Curve with Variable Regenerative Effect.— The conditions for oscillation are obtained by equating r_{21} and x_{21} to zero. If θ is eliminated from Eqs. (446-R) and (447-R), the equation of the boundary curve is

$$\begin{aligned}
&\theta_1^4\{\eta_1\eta_2(\eta_1 + \eta_2)\theta_2^4 + [\tau^2\eta_1 + \eta_2 + f\eta_1(\eta_1\eta_2 - \eta_1\eta_p - 2\eta_2\eta_p)]\theta_2^2 \\
&\qquad + f\eta_p[f\eta_1(\eta_p - \eta_1) - 1]\} \\
&+ \theta_1^2\{\eta_2^2(\eta_1 + \eta_2)\theta_2^6 + \eta_2[f(\eta_1\eta_2 - 2\eta_1\eta_p - 3\eta_2\eta_p) - 2(1 - \tau^2)]\theta_2^4 \\
&\qquad + f[f\eta_p(\eta_1\eta_p + 3\eta_2\eta_p - 2\eta_1\eta_2) + 2(\eta_p - \eta_2) - 2\tau^2\eta_p]\theta_2^2 \\
&\qquad + f^2\eta_p[2 - f\eta_p(\eta_p - \eta_1)]\} \\
&+ \left\{\eta_2\left[1 + \frac{\tau^2\eta_2}{\eta_1}\right]\theta_2^6 + f\left[2\eta_2 - \eta_p - \frac{2\tau^2\eta_2\eta_p}{\eta_1}\right]\theta_2^4\right. \\
&\qquad \left. + f^2\left[\eta_2 - 2\eta_p + \frac{\tau^2\eta_p^2}{\eta_1}\right]\theta_2^2 - f^3\eta_p\right\} = 0 \qquad (448\text{-}R)
\end{aligned}$$

Equation (448-R) is so complex that mere examination of the equation gives little idea of the shape of the boundary. The

equation is plotted in Figs. 219, 220, and 221 for certain selected values of the coefficients.

The frequency of oscillation can be obtained by equating $x_{21}/L_2\omega_0$, Eq. (447-R), to zero. With constant regenerative effect, the locus of max. ΔI_2 (max. 1-2 line) and the locus over which x_{21} for $\theta = 1$ is zero (locus of oscillation outside boundary for λ_0) are identical. In the case being considered, the two loci are

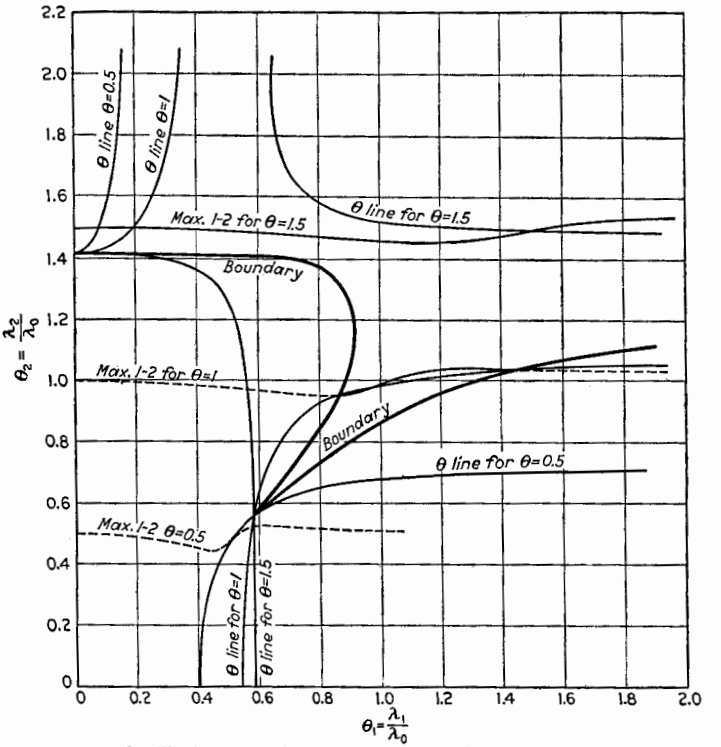


FIG. 219.—Oscillation-boundary curve for variable regenerative effect. $\eta_1 = 0.5$; $\eta_2 = 0.1$; $\eta_p = 10$; $\tau = 0.3$; $f = 0.02$.

different and are given by Eqs. (444-R) and (447-R). On substituting in Eqs. (444-R) and (447-R) the numerical values of the coefficients used in Figs. 219, 220, and 221, the two curves for the two equations are practically coincident and in all cases must exactly coincide at the boundary curve. In view of the closeness of the two curves, Eq. (447-R) alone has been plotted in Figs. 219, 220, and 221, and, for $\theta = 1$, the curve gives very closely the max. 1-2 line.

The coupling τ can be eliminated by combining Eqs. (446-R) and (447-R). This gives Eq. (449-R), which corresponds to Eq. (436-R) for the condition of constant regenerative effect.

$$\theta^2 = \theta_1^2 \frac{f(\eta_p - \eta_1) - \theta_2^2(\eta_1 + \eta_2)}{f\eta_p - \eta_1\theta_1^2 - \eta_2\theta_2^2} \qquad (449\text{-}R)$$

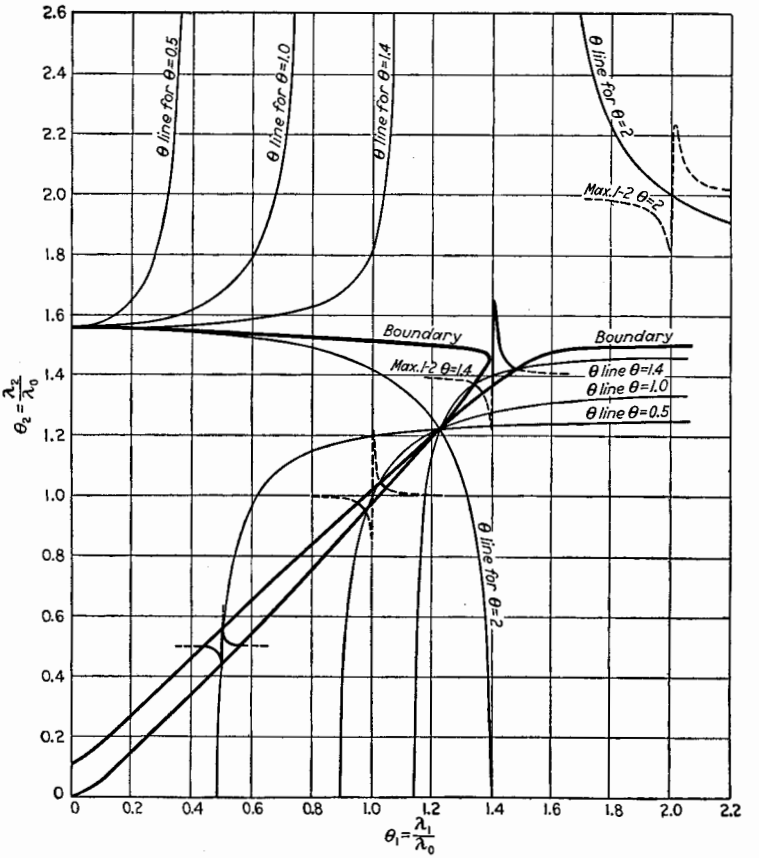


FIG. 220.—Oscillation-boundary curve for variable regenerative effect. $\eta_1 = 0.005$; $\eta_2 = 0.008$; $\eta_p = 72$; $\tau = 0.064$; $f = 0.00027$.

If θ is unity, this equation plotted to θ_1 and θ_2 gives the locus of the maximum regeneration points as the coupling is varied. These are the points of intersection of the boundary curve and the max. 1-2 and the max. 2-1 lines for λ_0. This locus is called the *θ line* for $\theta = 1$ and is shown plotted in Figs. 219, 220, and 221.

If θ is given other constant values and the coupling is varied, Eq. (449-R) gives the locus of points on the boundary where oscillation begins at a wave length $\theta\lambda_0$. Several other θ lines are plotted in Figs. 219, 220, and 221.

The theoretical curves of Fig. 219 were calculated for unusually large values of coupling and of circuit resistance in order to open up the curves and thus to show their characteristics more clearly.

The curves of Figs. 219, 220, and 221 will now be described in detail. In Fig. 219, the boundary curve, instead of giving a

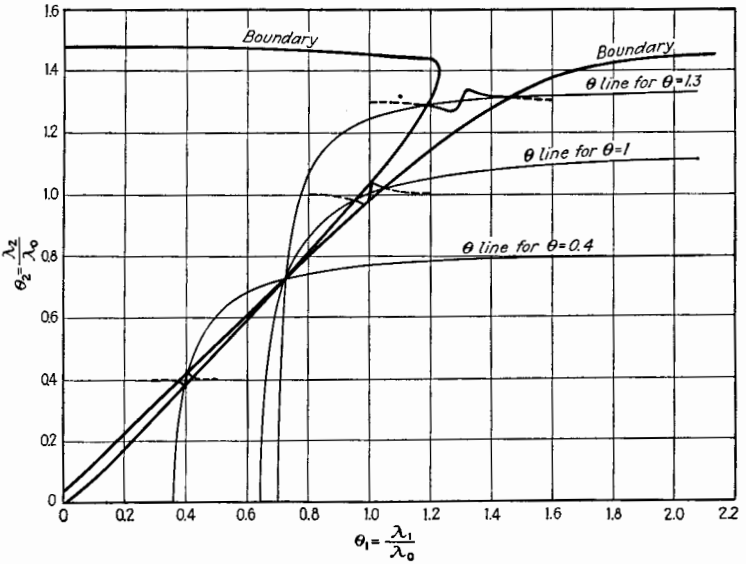


FIG. 221.—Oscillation-boundary curves for variable regenerative effect. $\eta_1 = 0.025$; $\eta_2 = 0.008$; $\eta_p = 70$; $\tau = 0.06$; $f = 0.00025$.

closed figure extending to the origin as in Figs. 216 and 218 or a region limited by two infinite straight lines as in Fig. 211, is now an open curve usually coming down to a sharp point at its lowest part. The region above the curve and within the acute angle is the region of non-oscillation; the part of the diagram below the boundary curve is the region of oscillation. The intercept on the θ_2 axis is the setting for which the secondary circuit, as a single circuit, begins to oscillate. The boundary curve does not extend to the origin because, as the secondary capacitance is decreased, the regenerative tendency rapidly

increases. For values of the secondary capacitance below a certain value (*i.e.*, for the lower extremity or tip of the boundary curve), the system oscillates no matter what the primary-circuit setting is.

In some cases the boundary Eq. (448-*R*) has imaginary roots for θ_1 for all values of θ_2 below the tip. This is the case for Fig.

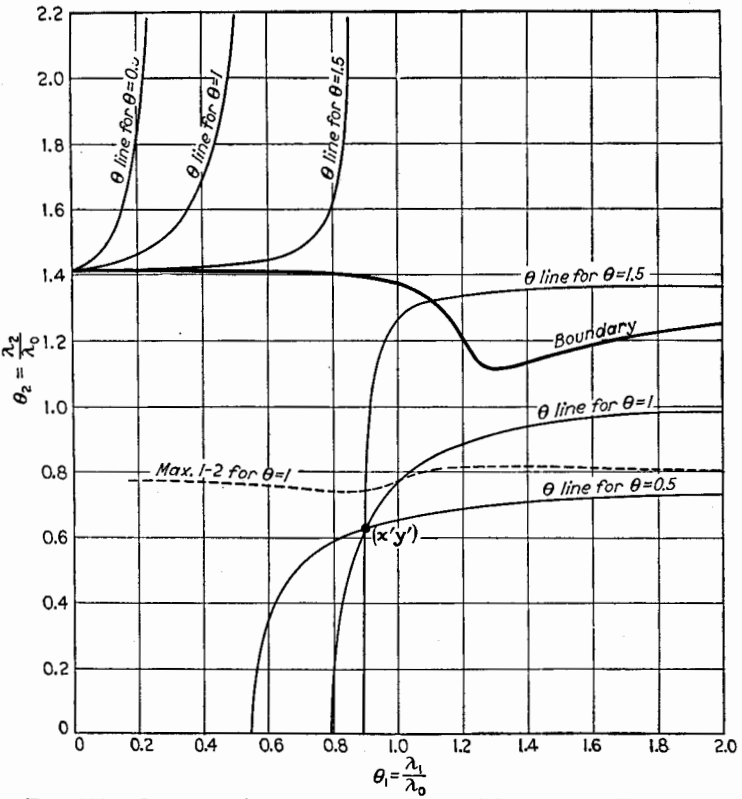


FIG. 222.—Oscillation-boundary curve for variable regenerative effect. $\eta_1 = 0.4$; $\eta_2 = 0.2$; $\eta_p = 1.0$; $\tau^2 = 0.08$; $f = 0.4$.

219. That this is not always so is shown by Figs. 220 and 221, where values of θ_2 below the tip give real values of θ_1. The boundary equation in these figures encloses a narrow region below the tip which is an overlapping of two regions of oscillation. Hence, it is a region where the system may oscillate at either one of two frequencies and where the drag loop is obtained if θ_1 is varied.

The tip of the boundary line is usually coincident with the common intersection point of all the θ lines. The coordinates of this point can be obtained from Eq. (449-R).

Coordinates of tip of boundary line

$$\left.\begin{aligned}\theta_1' &= \sqrt{\frac{f(\eta_p + \eta_2)}{\eta_1 + \eta_2}}\\ \theta_2' &= \sqrt{\frac{f(\eta_p - \eta_1)}{\eta_1 + \eta_2}}\end{aligned}\right\} \qquad (450\text{-}R)$$

The position of the tip is dependent upon f and is independent of coupling until the coupling becomes very weak. Then the tip suddenly draws away from the point determined by Eq. (450-R), and the boundary line may lose its point and show merely a dip, as in Fig. 222. An experimentally obtained plot

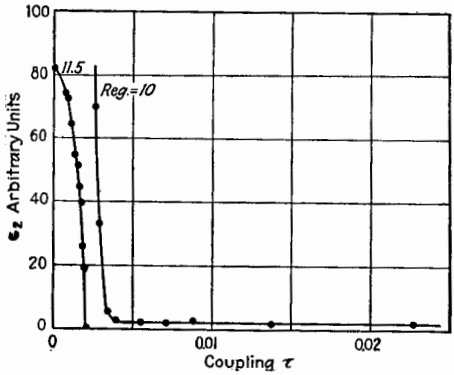


FIG. 223.—Position of tip of oscillation-boundary curve.

of the position of the tip for various couplings, indicated by the value of C_2, is shown in Fig. 223. Plots are given for two values of f or degrees of regeneration. The constancy of position of the tip is clearly shown by the curve for regeneration arbitrarily marked 10. For the other curve of Fig. 223, the regenerative coupling was much reduced, so that the settings for that tip are not on the plot. The rising portion touches the axis at a value of C_2 equal to 82 divisions on its scale. This value corresponds to the largest capacitance for which the secondary circuit oscillates alone and corresponds to the horizontal portion of the boundary line.

Although this condition of variable regenerative effect is much more complicated than the simple condition of constant regenera-

tive effect, there are some general features which are similar in both conditions. If f, which is always small in comparison with unity, is neglected in Eq. (447-R), the expressions of Eqs. (446-R) and (447-R) are the same as for Case b of constant regenerative effect, except that the effective η of the secondary circuit as affected by regeneration is

$$[\eta_2] = \eta_2 - \frac{f\eta_p}{\theta_2^2} \qquad (451\text{-}R)$$

and is a function of θ_2 instead of being constant as in the condition of constant regenerative effect. The boundary Eq. (448-R)

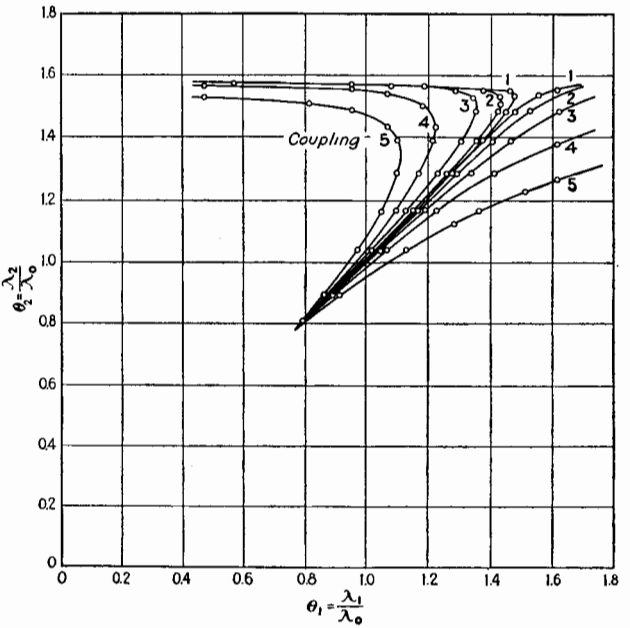


FIG. 224-1.—Experimental boundary curves. Effect of varying the coupling. $\lambda_0 = 600$ meters.

is approximately Eq. (439-R) for Case b, Condition I, except that $[\eta_2]$ is a function of θ_2. In the case of variable regenerative effect, $[\eta_2]$ varies on the $\theta_1 - \theta_2$ plane, becoming more negative as θ_2 decreases. In Cases a and b under constant regenerative effect, the region of no oscillation vanishes when $[\eta_2] = -\eta_1$, and the two edges of the boundary lines come together. In Condition II, the case of variable regeneration, the two boundary lines come together or cross at a point where $[\eta_2] = -\eta_1$. Here,

as in Cases a and b under constant regeneration, a drag loop is obtained when $\theta_2 = 1$ and as the primary circuit is tuned, provided $\tau > \eta_1$ and also provided $-\tau^2/\eta_1 < [\eta_2] < -\eta_1$. The boundary equation has real roots in the region below the tip if the coupling is greater than critical as defined by the relation $\tau^2 = -\eta_1[\eta_2]\theta^2$, and also if $[\eta_2] < -\eta_1$. At critical coupling, the edges of the boundary curve come together to enclose a vanishingly small strip of region of no oscillation, and as the coupling is decreased the tip rapidly rises and disappears.

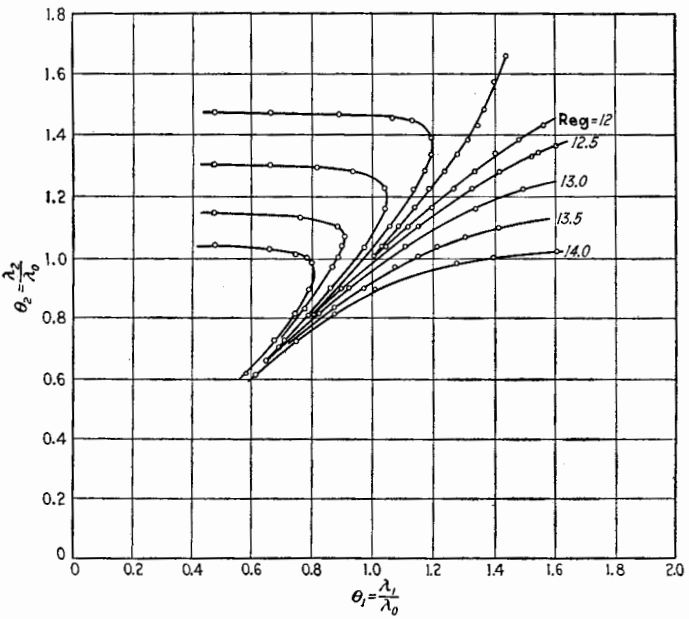


FIG. 224-2.—Experimental boundary curves. Effect of varying regeneration. $\lambda_0 = 600$ meters.

As given in Sec. 154, the discussion of the shape of the drag loop and of the action on tuning a primary circuit coupled to an oscillating secondary circuit applies unchanged to the condition of variable regeneration, since in that discussion the setting of the secondary circuit was assumed to be fixed.

160. Experimentally Determined Boundary Curves.—Some experimentally determined boundary curves confirming the theory are shown in Figs. 224-1, 224-2, and 224-3. Figure 224-1 demonstrates the constancy of the position of the tip and also

the shrinkage of the region of no oscillation as the coupling is decreased and approaches critical coupling. Figure 224-2 shows the effect of varying the coefficient of regeneration and hence

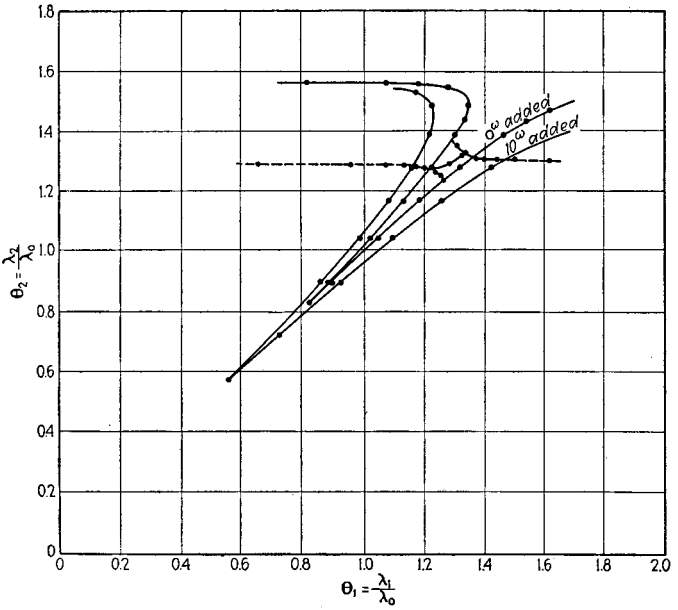


FIG. 224-3.—Experimental boundary curves. Effect of varying R_1. $\lambda_0 = 600$ meters.

the effect of varying the value of $[r_2]$. Figure 224-3 shows the effect of varying the primary resistance. According to Eq. (450-R), the position of the tip varies with η_1. In Fig. 224-3, portions of the max. 1-2 line and max. 2-1 line are shown.

CHAPTER XV

COMBINATIONS OF TRIODES AS LOW-POWER RESISTANCE-COUPLED AMPLIFIERS

Amplification by the use of a single triode was discussed in Chap. XII. When a greater amplification is desired than can be obtained by a single tube or stage, triodes may be combined in various ways to form multistage or cascade amplifiers. The plate circuit of each tube in a multistage amplifier must be closed through some form of impedance. The voltage across this impedance load is then impressed upon the grid of the next tube or stage. Multistage amplifiers are classified in three main groups, according to the character of the plate-load impedance:

I. Resistance-coupled amplifiers.
II. Impedance-coupled amplifiers.
III. Transformer-coupled amplifiers.

The first group comprises the arrangements in which the plate load is essentially a pure high resistance. It will be seen later that generally the capacitances of the tubes and stopping condenser taken in combination with the load resistance make the coupling a capacitance-resistance coupling.

The second group includes the arrangements in which the plate load is an untuned or a tuned inductive resistance, used as direct coupling to the next stage.

The third group includes all amplifiers in which the stages are coupled through either tuned or untuned transformers.

Amplifiers used for the increase of radio-frequency voltages are often treated separately from those used for audio-frequency amplification. Since the underlying theory is the same for all frequencies, no distinction will be made as to frequency until the final analysis is reached under each group. Then, the special sizes of elements and any peculiarities attributable to particular frequency ranges will be considered.

The discussion, except where otherwise qualified, deals with combinations of stages of amplifiers, each stage being an

amplifier of Class *A*, as described in Chap. XII. The principal criterion for the design of multistage amplifiers is high amplification with minimum distortion. The types of distortion were discussed in Chap. XII, where it was pointed out that there are two main types of distortion: *amplitude distortion*, wherein the amplitude of the output variations is not linearly related to the amplitude of the input variations; and *frequency distortion*, wherein variations at different frequencies are not equally amplified. Amplitude distortion is due primarily to a nonlinear relation between the input and output voltages or, in other words, to a nonlinear path of operation. A serious consequence of amplitude distortion is the presence in the output of harmonics and of sum and difference frequencies which are not present in the input. *Frequency distortion* is primarily due to a variation of the external impedances with frequency.

These two main types of distortion assume different relative importance according as the amplification is at radio frequencies or at audio frequencies. When the electric variations are at audio frequencies and it is desired to amplify them, both types of distortion are serious, amplitude distortion being probably more serious than frequency distortion. When, however, a high-frequency oscillation is modulated so that the envelope drawn through the peaks of the radio-frequency amplitudes is a true picture of a sound wave form, it is the sound wave form which must be preserved during amplification, and the wave form of the radio-frequency oscillation is of little importance. Amplitude distortion, in so far as it affects the radio-frequency wave form, is then unimportant; but if the amplitude distortion affects the shape of the envelope, this type of distortion is serious. Frequency distortion also is important in this case, because a modulated radio-frequency potential or current which forms a message is not merely of one frequency but is made up of several or many frequencies included within a range of frequencies 15,000 cycles wide perhaps. If the potentials or currents of these component frequencies are unequally amplified, the message is distorted. We see, therefore, that amplitude distortion is more serious at audio frequencies and frequency distortion at radio frequencies.

161. Fundamental Theory of Resistance-coupled Amplifiers.—The common arrangement of the resistance-coupled amplifier of three stages is shown in Fig. 225. The three stages, separated by the dotted lines, are denoted by S_1, S_2, and S_0. The load

Z_{b0} in the plate circuit of the last tube is usually very different from the load in the plate circuits of the other stages. The last stage may be used to deliver the maximum undistorted power to a loud-speaker, whereas the purpose of the other stages is to deliver the maximum undistorted voltage to the next stage beyond. The adjustment of voltages, and even the tube, of the last stage may be different from those of the previous stages. The last stage presents a problem in itself, which is solved by the methods of Chap. XII. Since each stage before the last, such as S_1 and S_2, performs the same function, we shall assume all stages except the last to be similar.

The condenser C is used merely to insulate the grid from the steady-voltage drop across the plate-circuit resistance R_b. Its

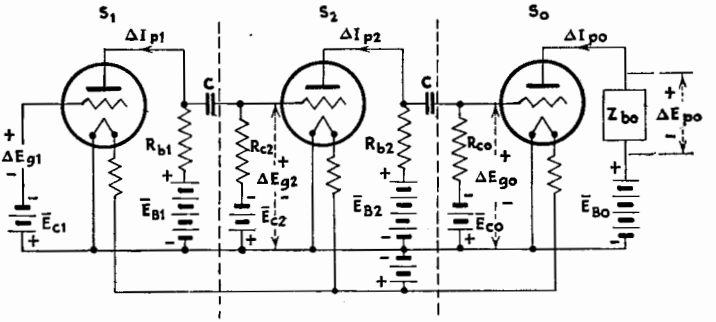


FIG. 225.—Multistage resistance-coupled amplifier.

capacitance is so chosen that its reactance is small in comparison with the resistance R_c. At very low frequencies, however, this capacitance cannot be made negligible, and the effect of the reactance of C requires study. R_c is made very large because its sole purpose is to fix the steady potential of the grid at the value $\bar{E}_c$. If the condenser C has little leakage, R_c may be as high as 10 megohms. Often there is so much leakage through C that the flow of the leakage current through R_c causes the potential of the grid to be different from $\bar{E}_c$, and $\bar{E}_g$ may be positive instead of negative.

Although separate B- and C-batteries are shown for each stage, the wiring can easily be altered to make one B-battery and one C-battery act for all stages. When this is done, it is absolutely necessary to shunt these batteries by low-impedance condensers in order to prevent regenerative effects from the last stages to the earlier stages.

In Fig. 225, ΔE_{g1}, ΔE_{g2}, and ΔE_{g0} represent the r.m.s. values of the grid voltages of the first, second, and last stages. ΔE_{p1}, ΔE_{p2}, and ΔE_{p0} represent the alternating plate voltages.

The voltage amplification of the nth stage, exclusive of the last stage, is

$$(\text{V.A.})_n = \frac{\Delta E_{g(n+1)}}{\Delta E_{gn}} \tag{452}$$

The voltage amplification of n stages is

$$(\text{V.A.}) = (\text{V.A.})_1 \cdot (\text{V.A.})_2 \cdot \cdot \cdot (\text{V.A.})_n \tag{453}$$

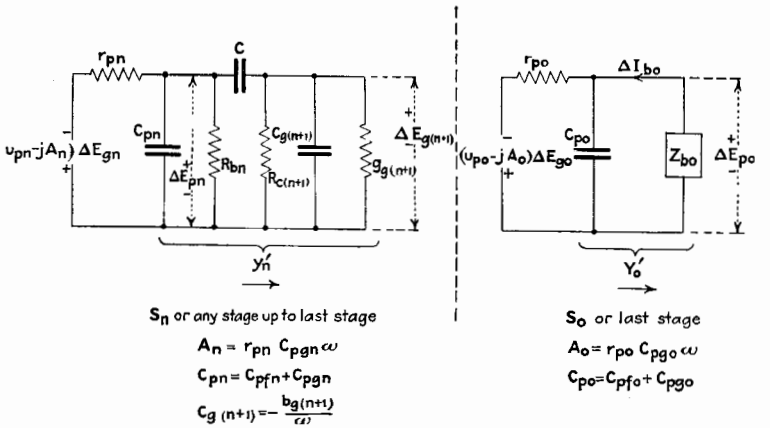


FIG. 226.—Equivalent a-c. circuits of the nth stage and last stage of a resistance-coupled amplifier.

The voltage amplification of the last stage or output stage, denoted by subscript 0, is

$$(\text{V.A.})_0 = \frac{\Delta E_{p0}}{\Delta E_{g0}} = -\frac{(\Delta I_{p0})Z_{b0}}{\Delta E_{g0}} \tag{454}$$

The performance of the whole amplifier, including the last stage, may be expressed by the over-all transconductance s, where

$$s = \frac{\Delta I_{b0}}{\Delta E_{g1}} \tag{455}$$

Although, in general, only the magnitude of the voltage amplification is of interest, it is sometimes convenient to express the voltage amplification in complex form from which the magnitude can be obtained when the components of the complex impedances or admittances are known.

The calculation of (V.A.) or s, even in the most simple arrangement, is complicated. The analysis is made of the nth stage, which applies to any stage up to the last, and of the last stage, using the equivalent circuits shown in Fig. 226. All grids are assumed to be negatively polarized, so that u_g and k_g are zero. The extension of the e-p-c. theorem, given in Eqs. (284) and (285), Chap. XI, page 282, must be used and the equivalent circuits of Fig. 226 are drawn to accord with the extended theory.

The following abbreviations are used in the analysis:

$$\left.\begin{aligned} A_n &= \frac{C_{pgn}\omega}{k_{pn}} = C_{pgn}\omega r_{pn} \\ C_{pn} &= C_{pfn} + C_{pgn} \\ C_{gn} &= -\frac{b_{gn}}{\omega} \\ k_{pn} &= \frac{1}{r_{pn}} \\ K_{bn} &= \frac{1}{R_{bn}} \\ K_{cn} &= \frac{1}{R_{cn}} \\ \mathbf{Y}_{b0} &= \frac{1}{\mathbf{Z}_{b0}} \end{aligned}\right\} \tag{456}$$

The admittances y'_n and Y'_0 are defined by the diagrams of Fig. 226, the admittances being evaluated in the direction of the arrows.

Using Eq. (284) and the abbreviations just given, the ratio of plate voltage to grid voltage for the nth stage is

$$\frac{\mathbf{\Delta E}_{pn}}{\mathbf{\Delta E}_{gn}} = -\frac{s_{pn} - jC_{pgn}\omega}{k_{pn} + \mathbf{y}'_n} \tag{457}$$

The ratio of the grid voltage of the $(n + 1)$st stage to the plate voltage of the nth stage is

$$\frac{\mathbf{\Delta E}_{g(n+1)}}{\mathbf{\Delta E}_{pn}} = \frac{jC\omega}{K_{c(n+1)} + g_{g(n+1)} + j(C_{g(n+1)} + C)\omega} \tag{458}$$

Hence, the exact expression for the voltage amplification of the nth stage is the product of Eqs. (457) and (458), or

$$(\mathbf{V.A.})_n = -\frac{s_{pn} - jC_{pgn}\omega}{\left[k_{pn} + \mathbf{y}'_n\right]\left[\dfrac{K_{c(n+1)} + g_{g(n+1)}}{jC\omega} + \left(\dfrac{C_{g(n+1)}}{C} + 1\right)\right]} \tag{459}$$

$$= -\frac{u_{pn} - jA_n}{\left[\frac{\mathbf{y}'_n}{k_{pn}} + 1\right]\left[\left(\frac{C_{g(n+1)}}{C} + 1\right) - j\frac{K_{c(n+1)} + g_{g(n+1)}}{C\omega}\right]} \tag{460}$$

Similarly, the voltage amplification of the last stage is

$$(\mathbf{V.A.})_0 = -\frac{u_{p0} - jA_0}{\frac{\mathbf{Y}'_0}{k_{p0}} + 1} \tag{461}$$

The expressions of Eqs. (459), (460), and (461) are general in that the character of the plate loads, denoted in Fig. 226 by $\mathbf{y}'_n$ and Z_{b0}, is not specified. By expanding $\mathbf{y}'_n$ and $\mathbf{Z}_{b0}$, we may apply Eqs. (460) and (461) to the special circuits shown in Fig. 226.

$$(\text{V.A.})_n = -\frac{\sqrt{u_{pn}^2 + A_n^2}}{\sqrt{\left[\left(\frac{K_{bn}}{k_{pn}} + 1\right)\left(\frac{C_{g(n+1)}}{C} + 1\right) + \frac{K_{c(n+1)} + g_{g(n+1)}}{k_{pn}}\left(\frac{C_{pn}}{C} + 1\right)\right]^2 + \left[\frac{C_{pn}\omega\left(\frac{C_{g(n+1)}}{C} + 1\right) + C_{g(n+1)}\omega}{k_{pn}} - \left(\frac{K_{bn}}{k_{pn}} + 1\right)\left(\frac{K_{c(n+1)} + g_{g(n+1)}}{C\omega}\right)\right]^2}} \tag{462}$$

If, as is generally the case, $C_{g(n+1)}/C$ and C_{pn}/C are very small in comparison with unity, Eq. (462) takes the simpler approximate form.

$$(\text{V.A.})_n = -\frac{\sqrt{u_{pn}^2 + A_n^2}}{\sqrt{\left[\frac{K_{bn} + K_{c(n+1)} + g_{g(n+1)}}{k_{pn}} + 1\right]^2 + \left[\frac{(C_{pn} + C_{g(n+1)})\omega}{k_{pn}} - \left(\frac{K_{bn}}{k_{pn}} + 1\right)\left(\frac{K_{c(n+1)} + g_{g(n+1)}}{C\omega}\right)\right]^2}} \tag{463}$$

Equation (461), expanded, becomes

$$(\text{V.A.})_0 = -\frac{\sqrt{u_{p0}^2 + A_0^2}}{\sqrt{\left(\frac{G_{b0}}{k_{p0}} + 1\right)^2 + \left(\frac{C_{p0}\omega - B_{b0}}{k_{p0}}\right)^2}} \tag{464}$$

We shall now study the voltage amplification of resistance-coupled amplifiers under certain special conditions.

162. Case 1. Resistance-coupled Amplifier at Low or Audio Frequencies.—Assume that the range of frequencies at which the amplifier operates is from 100 to 10,000 cycles per second. At these frequencies, A_n^2 is generally negligible in comparison with u_{pn}^2, and the ratio $(C_{pn} + C_{g(n+1)})\omega/k_{pn}$ is generally small in comparison with unity. Neglecting these terms, an approximate form of Eq. (463) suitable for use in the audio range of frequencies is

$$(\text{V.A.})_n = -\frac{u_{pn}}{\sqrt{\left[\frac{K_{bn} + K_{c(n+1)} + g_{g(n+1)}}{k_{pn}} + 1\right]^2 + \left[\left(\frac{K_{bn}}{k_{pn}} + 1\right)\left(\frac{K_{c(n+1)} + g_{g(n+1)}}{C\omega}\right)\right]^2}} \tag{465}$$

If C is of the order of a tenth of a microfarad or more, the second squared bracket in the denominator of Eq. (465) is negligible except at very low audio frequencies. For the present we shall neglect this bracket.

To obtain high amplification, each of the factors K_{bn}/k_{pn}, $K_{c(n+1)}/k_{pn}$, and $g_{g(n+1)}/k_{pn}$ should be small in comparison with unity. K_{bn} and $K_{c(n+1)}$ can be chosen by the designer. $K_{c(n+1)}$ is merely for the purpose of fixing the polarizing potential of the next stage at an appropriate negative value, and, since there is no appreciable steady grid current, $K_{c(n+1)}$ can be made very small or its reciprocal very large. For example, $R_{c(n+1)}$ may be of the order of from 1 to 10 megohms. Thus the factor $K_{c(n+1)}/k_{pn}$ can be made so small as to be of little effect in reducing amplification.

It is well to point out that, if $R_{c(n+1)}$ is large, it is necessary that condenser C be a mica condenser in order to reduce the leakage current through C. The leakage current through a paper condenser when used for C is almost always great enough to cause a substantial voltage drop through $K_{c(n+1)}$ and to polarize the grid of the next stage to a positive potential.

The choice of K_{bn} depends upon the tube used. As K_{bn} is made smaller, the steady-voltage drop through it is increased and hence, for a given plate-battery voltage, the plate voltage of the tube is decreased. This decrease in plate voltage is best given by the $i_p - e_p$ diagram of Fig. 227. Let $\bar{E}_B$ be the voltage of the plate battery. Various resistance lines are drawn for R_{bn}, and

their intersections with the curve for the chosen grid-polarizing potential $\bar{E}_{cn}$ give the quiescent points for the several values of R_{bn}. (The subscripts n are omitted in Fig. 227.) The grid-polarizing potential should be just sufficiently negative to insure that the largest grid-voltage variation will not cause grid current to flow. The negative polarizing potential may be less for the first stages than for the last stages. The quiescent points for the several values of R_b give the values of $\bar{I}_p$ and $\bar{E}_p$. As R_b is increased or K_b decreased, k_p also decreases. The only way to determine how the ratio K_b/k_p varies as K_b is decreased is to

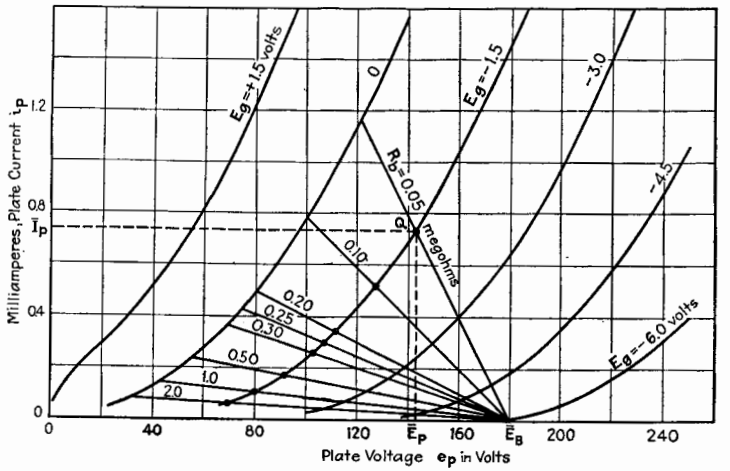


FIG. 227.—Plate-current plate-voltage diagram for a resistance-coupled amplifier.

plot the ratio from measurements with the particular tube used. It is usually more convenient to plot the reciprocal of this ratio, or k_p/K_b. Figure 228 gives a family of curves for this ratio for various grid-polarizing voltages and plate voltages. A certain commercial high-u tube known as UX240 (see Plate II*d*, opposite page 144) was used for these measurements. The curves of Fig. 228 show that as R_b is increased the curve of k_p/K_b rises rapidly at first and then flattens out. Equation (465) shows that the larger the value of k_p/K_b, the greater the amplification. With ordinary triodes this ratio is never much greater than 10, no matter how large R_b is made. It seldom pays to make the ratio much greater than 3 or 4, because the gain in amplification beyond these values is slight.

The third factor $g_{g(n+1)}/k_{pn}$ which occurs in the denominator of Eq. (465) may have a considerable effect upon the voltage amplification of the nth stage, particularly at high audio frequencies. The value of $g_{g(n+1)}$ depends upon the load in the plate circuit of the next stage. The expression for calculating $g_{g(n+1)}$ was given in Eq. (278), page 277, Chap. XI, and is reproduced

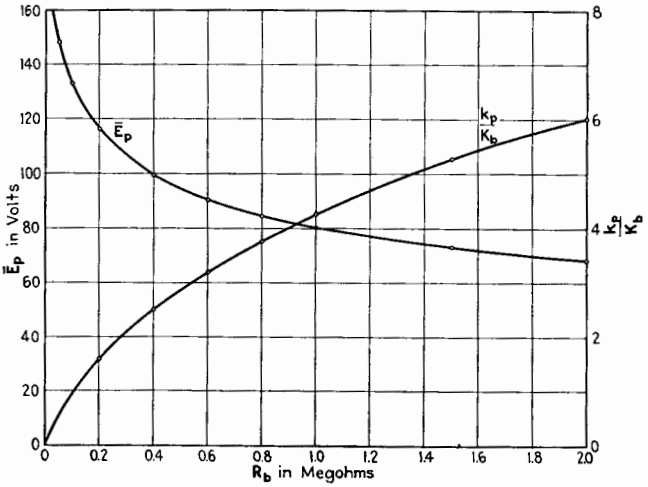


FIG. 228.—Variation of k_p/K_b and $\bar{E}_p$ with R_b for a UX 240 triode. $\bar{E}_c = -1.5$ volts.

here in a slightly modified form which is somewhat more convenient for use in this particular problem.

$$\frac{g_{g(n+1)}}{k_{pn}} = \left(\frac{C_{pg(n+1)}\omega}{k_{p(n+1)}}\right)^2 \cdot \frac{k_{p(n+1)}}{k_{pn}} \cdot \frac{1 + \dfrac{G_{b(n+1)}}{k_{p(n+1)}} + u_{p(n+1)}\left(1 + \dfrac{C_{pf(n+1)}\omega - B_{b(n+1)}}{C_{pg(n+1)}\omega}\right)}{\left(1 + \dfrac{G_{b(n+1)}}{k_{p(n+1)}}\right)^2 + \left(\dfrac{C_{pg(n+1)}\omega}{k_{p(n+1)}}\right)^2\left(1 + \dfrac{C_{pf(n+1)}\omega - B_{b(n+1)}}{C_{pg(n+1)}\omega}\right)^2} \tag{466}$$

where $G_{b(n+1)}$ and $B_{b(n+1)}$ are the conductance and susceptance of the total plate load of the next stage. To give an idea of the approximate value of $g_{g(n+1)}/k_{pn}$ at 10,000 cycles per second, we may substitute in Eq. (466) the following values for an actual amplifier, the test of which will be described in the next section. A type UX240 triode has the following approximate character-

istics when operated with a plate-load resistance of 250,000 ohms, and a plate-battery voltage of 180 volts.

$$\left.\begin{aligned} \mu_p &= 30 \\ r_p &= 125{,}000 \text{ ohms} \\ C_{pg} &= 10\mu\mu\text{f} \\ C_{pf} &= 4\mu\mu\text{f} \\ C_{gf} &= 6\mu\mu\text{f} \end{aligned}\right\} \text{including socket and wires}$$

The term $B_{b(n+1)}/C_{pg(n+1)}\omega$ in Eq. (466) is very important because it can have very large values, depending on the type of

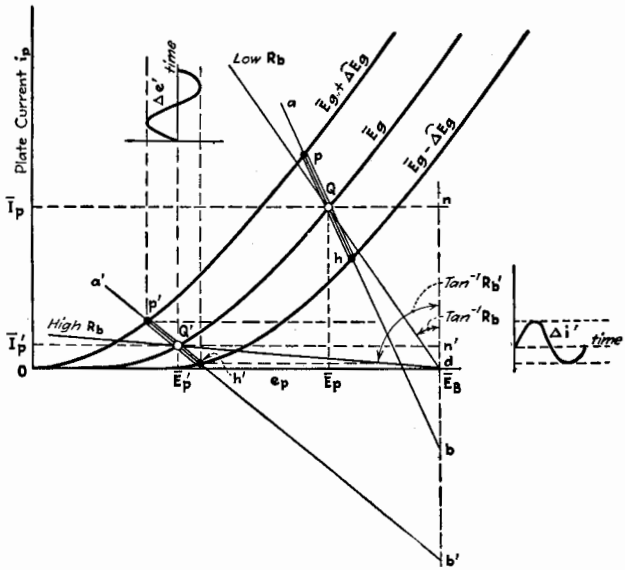


FIG. 229.—Effect of varying R_b of a resistance-coupled amplifier.

load in the plate circuit. When the plate load of the next stage is the same as for the stage being treated, a reasonable value is 8.

Assuming the values just given for the terms in Eq. (466), the formula gives for $g_{g(n+1)}/k_{pn}$ the value 0.775 at 10,000 cycles per second. Formula (465) gives 13.2 for the amplification. This checks well with the experiments. The value of $g_{g(n+1)}/k_{pn}$ decreases rapidly as the frequency is decreased, so that its effect is noticeable only at high frequencies.

An effect which causes distortion when R_b is made large is worthy of notice. Figure 229 is similar to Fig. 227 except that

only two resistance lines are shown, one for a fairly low value of R_b and one for a high value of R_b. Because of the stopping condenser C, the plate-load resistance for alternating current is less than that for steady current. The difference between the two plate-load resistances is accentuated if $K_{c(n+1)} + g_{g(n+1)}$ is increased. The path of operation for low R_b is not along the resistance line for R_b, but along a line such as ab, Fig. 229, drawn through the quiescent point. This path of operation makes with the horizontal an angle whose tangent is $K_{bn} + K_{c(n+1)} + g_{g(n+1)}$ and intersects the vertical through $\bar{E}_B$ at point b. Since $\overline{dn}/\overline{nQ} = K_{bn}$ and $\overline{bn}/\overline{nQ} = K_{bn} + K_{c(n+1)} + g_{g(n+1)}$, then $\overline{bd}/\overline{nQ} = K_{c(n+1)} + g_{g(n+1)}$. Now suppose that the high R_b is used and assume that $K_{c(n+1)} + g_{g(n+1)}$ remains the same. The path of operation is along a line such as $a'b'$ drawn so that $\overline{b'd}/\overline{n'Q'} = \overline{bd}/\overline{nQ}$. Since the plate-current curves for constant grid voltages are practically similar in shape, curves for equal increments of grid voltage are approximately equidistant along any horizontal line, whether the line is near the axis of abscissas or higher up on the diagram. Hence, if operation were along the resistance line for a very high R_b, the path would be so nearly horizontal that the plate-voltage variation would be essentially linear in its relation to the grid-voltage variation. But the plate-current lines for constant grid voltages are not equidistant when measured vertically in the region where the characteristic graphs are curved. Consequently, the more nearly vertical is the path of operation, which cuts through the curved region of the characteristic surface, the more nonlinear is the relation between plate-voltage variation and grid-voltage variation. The high slope of the path $a'b'$ results in wave distortion of the output voltage, as indicated by the curves for $\Delta e'_p$ and $\Delta i'_p$. If the variation of grid voltage exceeds that indicated by the path $p'h'$, the plate current is zero during part of the cycle, which results in excessive distortion.

It can be stated, therefore, that for a high-quality amplifier it is inadvisable to increase R_b above the value which makes $R_b/r_p = k_p/K_b$ greater than 2 or 3. If this ratio is made large, both amplitude and frequency distortion are introduced. Furthermore, $K_{c(n+1)} + g_{g(n+1)}$ should be less than one-fifth of k_{pn} to avoid the distortion described with reference to Fig. 229.

There is a practical advantage in using a large R_b, which, however, is of little importance except when dry batteries are

used as sources of power. A large value of R_b results in a small steady plate current. Consequently, there is less load on the plate-circuit source of power and the cathode can be operated to give low emission, resulting in a saving of filament-heating power and an increase in the life of the tube.

The second stage of the amplifier of Fig. 225 differs from the first stage in one or two particulars. In the first place, the grid-voltage variations of the second stage are much greater than for the first stage because of the amplification of the first stage. Consequently the negative polarizing potential of the second stage should be greater than for the first stage, unless, to make the two stages alike, the grid of the first stage is polarized negatively by the same amount demanded by the second stage. The second difference between the two stages arises from the fact that the input admittance of the second stage, depending as it does on the plate load of the third stage, may be quite different from the input admittance of the first stage. If the third stage is the final stage, its plate load may be a loud-speaker, whose impedance varies greatly with frequency and may be highly inductive at some frequencies. An inductive plate load may give a negative input conductance, so that g_{g0} may vary greatly and be negative for some frequencies. This variation of g_{g0} may cause the amplification curve of the second stage to be quite different from that of the first stage.

Before ending this section, a few words are relevant concerning the effect of the stopping condenser C, Fig. 225, when its reactance is not negligible. If C is 0.1μf, its reactance at 1,000 cycles per second is 1,590 ohms, at 100 cycles per second 15,900 ohms, and at zero cycles per second is infinite. This stopping condenser affects the amplification if the term $\frac{K_{c(n+1)} + g_{g(n+1)}}{C\omega}$ in Eq. (465) is appreciable compared with unity. For example, if C is 0.1μf and $K_{c(n+1)} + g_{g(n+1)}$ is 10^{-6} mho, this term has a value of only 0.159 at 10 cycles per second. Unless C is made very small, or $K_{c(n+1)} + g_{g(n+1)}$ is unusually large, the last term in the denominator of Eq. (465) is negligible from 100 to 10,000 cycles per second, and Eq. (465) becomes

$$(\text{V.A.})_n = -\frac{u_{pn}}{\frac{K_{bn} + K_{c(n+1)} + g_{g(n+1)}}{k_{pn}} + 1} \tag{467}$$

Taking into account all of the effects which introduce frequency distortion in a resistance-coupled audio-frequency amplifier, we should expect the voltage amplification to be a maximum for some rather low frequency, to fall off rapidly as zero frequency is approached, and to fall off less rapidly at high frequencies.

Figure 230 shows experimentally determined voltage-amplification curves for a three-stage, resistance-coupled amplifier, the

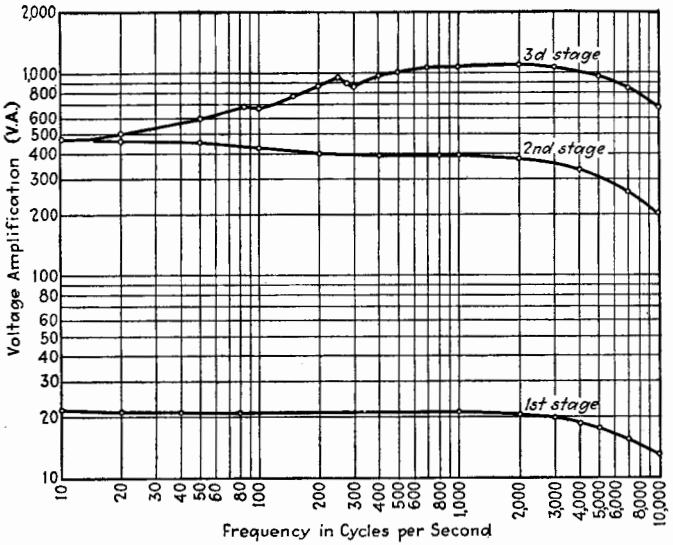


FIG. 230.—Voltage amplification of a three-stage resistance-coupled amplifier measured by a vacuum-tube voltmeter.

FIRST AND SECOND STAGES	THIRD STAGE
UX 240 triode	UX 271A triode
$\mu_p = 30$	$\mu_p = 3$
$R_b = 250{,}000$ ohms	$\bar{E}_C = -22.5$ volts
$R_c = 5$ megohms	$\bar{E}_B = 135$ volts
$C = 0.08\mu$f	Plate load = Western Electric Co. cone-type loud speaker.
$\bar{E}_C = -1.5$ volts	
$\bar{E}_B = 180$ volts	

specifications for which are given under the figure. Both scales of Fig. 230 are logarithmic. The abscissa scale of frequencies is logarithmic, first, because such a scale spreads out the low-frequency end so that the performance of the amplifier can be shown on the same chart for both low frequencies and high frequencies; and, second, because the ear evaluates sounds of various frequencies in terms of ratios of frequencies and not in accordance with a uniform scale. The ordinate

scale is logarithmic because, within the ranges of comfortable intensities of sound, the *intensity of sensation* is practically proportional to the logarithm of the sound intensity or energy. Since energy is proportional to the square of the amplitude of voltage or current producing the sound, the logarithmic plot of voltage amplification or of over-all transconductance of an amplifier has the same shape as the logarithmic plot of energy and better represents relative sensations than does a plot on a uniform scale.

The decrease in amplification due to the effect of $g_{g(n+1)}$ at the high frequencies is evident in the curves of Fig. 230 for the first and second stages. The condenser C was sufficiently large to prevent a decrease in amplification at the low frequencies above 10 cycles. The curves would drop to zero at zero frequency. For the amplifier tested, this drop would occur considerably below 10 cycles per second.

The curve of voltage amplification for the third stage shows some irregularities due to resonance of the diaphragm of the loud-speaker. The cone speaker used in the test had a maximum sensitivity at frequencies of about 2,000 cycles per second, as shown by the maximum of the voltage-amplification curve.

163. Experimental Determination of the Performance of a Resistance-coupled Audio-frequency Amplifier.—Although the theory helps in the understanding and design of an amplifier, the real performance is determined by experiment. Two methods of measuring the amplification of a multistage amplifier are given.

A. Measurement of Voltage Amplification by Vacuum-tube Voltmeter.—The first and most obvious method is to impress a known small voltage upon the input and measure the output voltage by means of a suitable vacuum-tube voltmeter (to be described in Chap. XXII). Two alternative procedures are possible in the execution of this method. One procedure is to maintain the input voltage constant and measure the output voltage. The other alternative, which is usually preferred, is so to adjust the input voltage by means of a variable attenuator as to give always the same value of output voltage. In both it is usually desirable to measure both input and output voltages by means of the same vacuum-tube voltmeter in order to eliminate the possibility that two different voltmeters may have different frequency errors.

An arrangement shown diagrammatically in Fig. 231 was used to obtain the curves of Fig. 230. A beat-frequency oscillator supplied current to a high-resistance potentiometer. The voltage across the potentiometer or the attenuator could be measured by throwing switch S to the left. This voltage was kept constant at 4 volts. When the voltage amplification of the entire amplifier was measured, the attenuator was adjusted to give an output voltage of 4 volts. In order to reduce the input voltage sufficiently, it was necessary to add 90,000 ohms in series with the attenuator, Fig. 231. The voltage amplifications of the first stage and of the first and second stages were

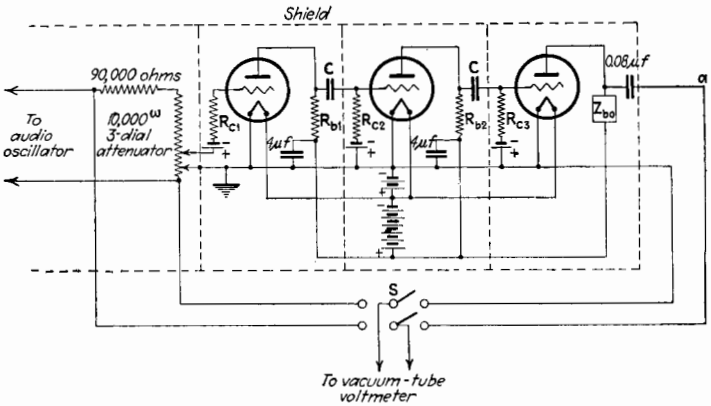


FIG. 231.—Measurement of the voltage amplification of a multistage amplifier by means of a vacuum-tube voltmeter.

measured by connecting the wire a from the switch to the grid of the second and third stages.

B. Measurement of Voltage Amplification by Balance Method.—The second method of measuring the performance of an amplifier is a balance method and can be used for determining the voltage amplification of any single stage or of any number of stages. The diagram of connections for this method, as applied to the measurement of the over-all voltage amplification of a three-stage amplifier, is shown in Fig. 232. An audio oscillator, capable of giving any audio frequency, is connected to send a current ΔI_0 through resistances R_1 and R_2 and through one coil of a variable mutual inductance M. R_1 is a low resistance of the order of 1 ohm, and the voltage drop across it is introduced into the grid circuit of the first stage. R_2 is a variable resistance

made up of units-, tens-, hundreds-, and thousands-ohm dials. M should be such that the maximum $M\omega$ is several thousand ohms. The voltage variation across Z_{b3} is balanced in magnitude and phase by the voltage across R_2 and the secondary winding of the mutual inductance, the balance being indicated by silence in the high-impedance telephone receivers. The voltage amplification is

$$(\text{V.A.}) = \frac{\sqrt{R_2^2 + (M\omega)^2}}{R_1} \tag{468}$$

The condenser in series with the telephone receivers is used to prevent the flow of any steady current from the plate battery through the telephone receivers and R_2.

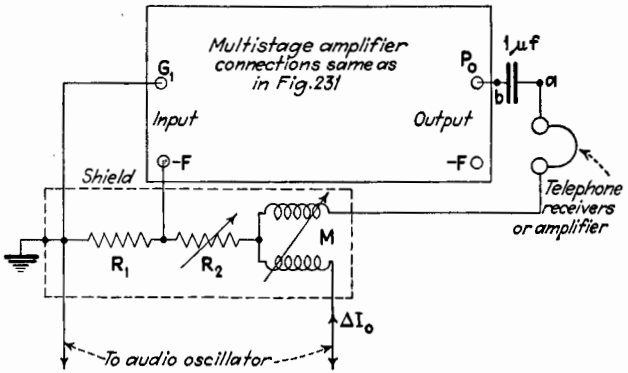


FIG. 232.—Balance method of measuring the voltage amplification of a multistage amplifier.

By attaching terminal b of the condenser to the grid of the second stage instead of to Z_{b3}, the voltage amplification of the first stage can be measured. Even though the amplification of the first stage only is measured, it is best to maintain the other stages in operating condition in order that all cross effects which normally exist between stages may be present when the measurements are made.

The voltage amplification of any other stage, taken alone, can be determined similarly by connecting R_1 in the grid circuit of the stage to be tested, at the same time disconnecting the condenser C from the grid and attaching terminal b to the grid of the next stage.

The voltage amplification of an even number of stages, such as the first two stages, can be measured by a slight shift

of connections. If the terminal a of the telephone receivers were connected to the grid of the third stage, balance would not be obtained because of the wrong phase of the plate-voltage variation of the second tube with respect to the input voltage. By interchanging the input terminals connected to R_1, balance can be obtained. In this case Eq. (468) should read

$$(\text{V.A.}) = \frac{\sqrt{(R_1 + R_2)^2 + (M\omega)^2}}{R_1} \tag{469}$$

R_1 is usually so small in comparison with R_2 that Eq. (468) is sufficiently accurate.

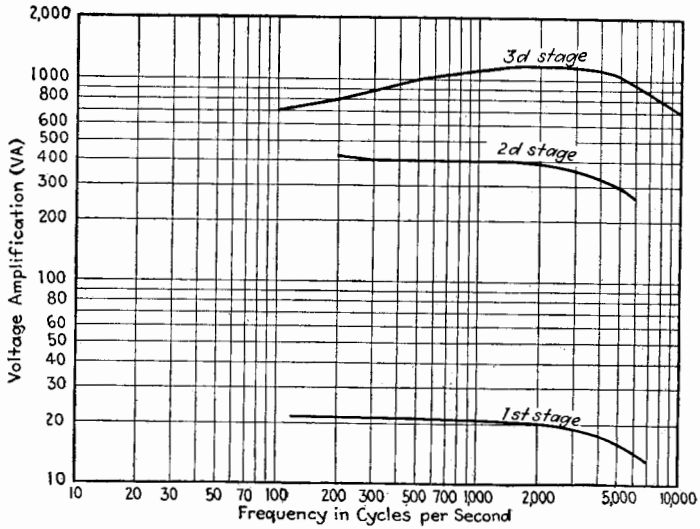


FIG. 233.—Voltage amplification of a three-stage resistance-coupled amplifier measured by the balance method. Constants of the amplifier as given in Fig. 230.

The curves of Fig. 233 were obtained by the balance method. Since both the ear and telephone receivers become insensitive at very low and at very high frequencies, the range of measurement by this method is limited, unless some other method of detection is used. A vibration galvanometer may be used at low frequencies. A better method of detection is the use of an amplifier terminated by a detector or vacuum-tube voltmeter. With this arrangement, the method is rapid and accurate.

164. Use of Common Plate Battery.—Separate plate-circuit batteries for each stage are shown in Fig. 225. It is much more convenient if the same battery can be used for all stages. This

can be done for a few stages if means are provided to prevent coupling between stages due to the common electrical path through the battery.

Consider the effect of this common path in a resistance-coupled amplifier. If the potential of the grid of the first tube is *increased*, the plate current of the first tube *increases*. This causes a *decrease* in the potential of the grid of the second tube and hence a *decrease* in the plate current of the second tube. The plate-current variations of the second tube are opposite in phase to the plate-current variations of the first tube. The plate-current variations of the second tube, passing through the impedance of the common battery, tend to reduce the voltage variations of the grid of the second tube. The plate-current variations of the third tube, however, are in the same direction as the plate-current variations of the first tube and, passing through the impedance of the common battery, tend to increase the grid-voltage variations of the second tube. Regeneration thus takes place.

The regenerative effect can be considerably reduced, especially at high frequencies, by large condensers connected across the battery, as shown in Fig. 231. The impedance, even of very large condensers, is too great at the very low frequencies to prevent regeneration. The increase of amplification at low frequencies due to this regeneration is evident in the curve for the second stage, Fig. 230. Regeneration at low frequencies may cause the multistage amplifier to oscillate at these frequencies. These oscillations may build up charges in the grid-stopping condensers C, owing to rectification in the grid circuit. The polarizing potential thus builds up until the oscillations stop. The charge in the condenser then leaks off and the oscillations begin again. This periodic starting and stopping of oscillation is often known as "motor boating," because of the similarity in the sound emitted by the loud-speaker and by the exhaust of a motor boat. "Motor boating" can be stopped by reducing the amplification at low frequencies or by using two or more sources of plate-circuit power. If the amplifier has more than three or four stages, it is usually necessary to supply the first two or three stages from one source of plate voltage and the remaining stages from another source.

165. Case 2. D-c. Amplifier.—A d-c. amplifier is one which operates at zero frequency and is used to amplify a voltage which varies slowly and which may have periods of no change. The

amplifiers just described are unsuitable as d-c. amplifiers because of the infinite reactance of the condenser C which couples the grid circuit to the plate circuit.

The condenser C can be omitted if provision is made for maintaining the proper steady grid potentials. Figure 234 shows one arrangement suitable for a so-called d-c. amplifier. Separate, well-insulated B-batteries are required. Some point in the B-battery has nearly the same potential as the filament, and the grid of the next stage can be connected near this point to give the proper negative grid potential. The adjustment is

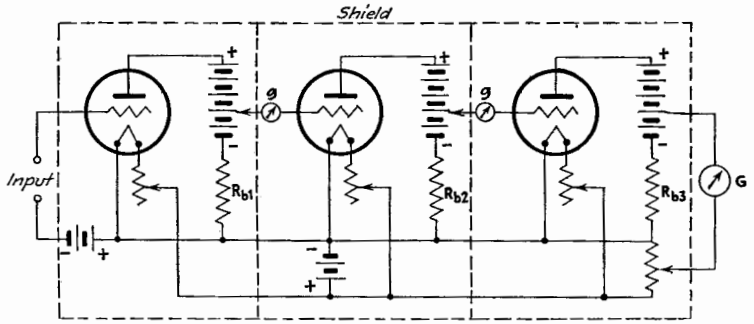


FIG. 234.—Direct-current amplifier.

facilitated by observing a galvanometer (shown at g) in the grid circuit. If the adjustment of the tap is such that the galvanometer shows a current, the grid potential is positive. The tap is then moved by the proper amount toward the negative end of the battery starting from the position which just fails to show grid current. The final detecting device, such as a galvanometer, is connected so as to balance out the steady current. Rough adjustment is made by the tap on the last B-battery, and final balance is obtained by a potential divider across the A-battery. The author has used a two-stage amplifier of this type for measuring the electric potentials produced in the retina of the eye when the retina is illuminated. An Einthoven galvanometer was used at G.[1]

166. Kallirotron: Regenerative Resistance-coupled Amplifier.—A type of regenerative resistance-coupled amplifier, suitable for use as a d-c. amplifier, was devised by Turner.[2] The connec-

[1] CHAFFEE, BOVIE, and HAMPSON, Electric Potential of the Retina under Stimulation by Light, *J. Optical Soc. Am.*, **7**, 1 (1923).

[2] TURNER, *Radio Rev.*, April, 1920.

tions are shown in Fig. 235. Let Δe_0 be the instantaneous value of the small impressed e.m.f. Let Δe_1 be the resulting amplified variations in the plate circuit of the first tube. Let Δi_{p_1} and Δi_{p_2} represent the instantaneous values of the variations in plate current of the first and second tubes.

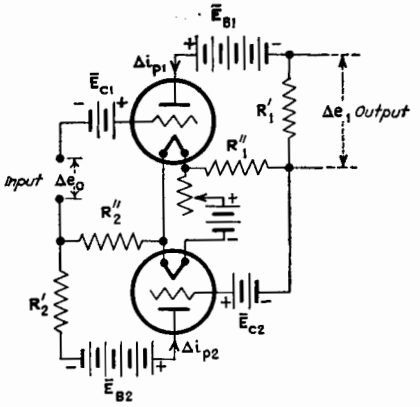


FIG. 235.—Turner's kallirotron.

By the e-p-c. theorem

$$\left.\begin{aligned}\Delta i_{p_1} &= \frac{u_{p_1}(\Delta e_0 + R_2''\Delta i_{p_2})}{r_{p_1} + R_1 + R_1''} \\ \Delta i_{p_2} &= \frac{u_{p_2}R_1''\Delta i_{p_1}}{r_{p_2} + R_2' + R_2''}\end{aligned}\right\} \tag{470}$$

These can be combined to give

$$(\text{V.A.}) = \frac{\Delta e_1}{\Delta e_0} = \frac{u_{p_1}R_1'(r_{p_2}+R_2'+R_2')}{(r_{p_1}+R_1'+R_1'')(r_{p_2}+R_2'+R_2'') - u_{p_1}u_{p_2}R_1''R_2''} \tag{471}$$

The negative sign in the denominator allows it to be made small, giving a large value of (V.A.). The system is apt to be unstable if (V.A.) is made large, unless great care is used in selecting similar tubes and in adjusting the battery potentials so that the quiescent points are at points of maximum transconductance.

167. Case 3. Resistance-coupled Amplifiers at High Frequencies.—We have considered the operation of multistage resistance-coupled amplifiers only at frequencies below 10,000 cycles per second. The amplification at the higher frequencies

of the audio-frequency range is decreased in consequence of the increase of input admittance with increasing frequency. As a result, frequency distortion is present, especially when large resistances are used in the plate circuit. As the frequency at which the amplifier is to operate is made greater than 10,000 cycles per second, the disturbing effects of input admittance are accentuated and the amplification attainable is reduced. Usually, at frequencies above the audio-frequency range, gradual frequency distortion is of small concern because, at the higher frequencies, the frequency range over which the amplifier operates is fractionally very small. For example, the audio amplifier must operate from 100 to 10,000 cycles per second, which is over six octaves. A radio-frequency amplifier is seldom expected to operate over more than one or two octaves, and at any one time the range of frequencies amplified covers a very small fraction of a single octave.

The real obstacle to the use of resistance-coupled amplifiers at high frequency is the low amplification obtainable. That the large value of the input admittance of a triode at high frequencies is the cause of this deficient action is shown by Eq. (463). As an example, assume that the frequency to be amplified is a million cycles per second and that $C_{pgn} = C_{pfn} = 5\mu\mu f$. Let $k_{pn} = 1/16{,}000$ mho. Then A_n is equal to 0.5. If K_{bn} and $K_{b(n+1)}$ are each 1/16,000 mho, Fig. 150, page 276, shows that $g_{g(n+1)}/k_{pn} = 0.8$ and $C_{g(n+1)}\omega/k_{pn} = 2.5$. With these values for the factors in Eq. (463), the value of $(\text{V.A.})_n$ is approximately $u_{pn}/4.25$. If K_{bn} and $K_{b(n+1)}$ are 1/32,000 mho, the voltage amplification is $u_{pn}/4.7$, while if these plate conductances are reduced to 1/8,000 mho, the voltage amplification is $u_{pn}/3.76$. At high frequencies, the greatest voltage amplification occurs generally at a relatively large value of K_b/k_p, that is for values of R_b/r_p less than unity. With common tubes having a voltage ratio of about 8 and internal capacitances of the order of $5\mu\mu f$, it is difficult at radio frequencies to obtain a voltage amplification per stage greater than 2, and it is usually less than this value. Specially designed triodes to have low internal capacitances give much better results. Since impedance-coupled and transformer-coupled amplifiers give so much greater voltage amplification at radio frequencies, resistance-coupled amplifiers are of little practical importance at these frequencies.

CHAPTER XVI

COMBINATIONS OF TRIODES AS LOW-POWER IMPEDANCE-COUPLED AMPLIFIERS

The main advantage of a resistance-coupled amplifier is the relatively uniform amplification over a considerable range of frequencies. The principal disadvantage of such an amplifier is the smaller total amplification attainable compared with the degree of amplification possible with other types using the same number of stages.

The next simple type of amplifier to be considered utilizes, in place of the plate resistance R_b of the resistance-coupled amplifier, an impedance consisting of either a tuned or an untuned inductive resistance. Usually, the inductance of this plate

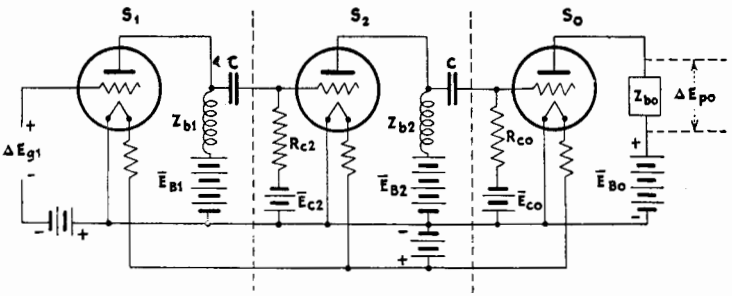


FIG. 236.—Multistage impedance-coupled amplifier.

impedance is the prominent factor, the resistance being small. In studying the general theory of the impedance-coupled amplifier, no restrictions will be made at first as to frequency.

168. Fundamental Theory of Impedance-coupled Amplifier.—The connections of this type of amplifier are shown in Fig. 236, a three-stage amplifier being used for illustration. The equivalent a-c. circuit diagrams for any stage except the last, and also for the last stage, are shown in Fig. 237. Figure 237 is similar to Fig. 226, page 399, except for the addition of L_{bn} in series with R_{bn}. It should be noted that the distributed capacitance C'_{bn} of the impedance Z_{bn}, and the capacitance C_{bn} of any added

physical condenser are grouped with C_{pfn} and C_{pgn}. The sum of the four capacitances is denoted by C_{pn}, or $C_{pn} = C_{pfn} + C_{pgn} + C'_{bn} + C_{bn}$. Furthermore, the value of R_{bn} to alternating current, denoted by $\tilde{R}_{bn}$, may vary with the frequency because of iron losses, eddy currents, and skin effect.

The last stage is assumed to be the same as in Fig. 226 for the resistance-coupled amplifier and will not be discussed again. Refer to Eqs. (454), (461), and (464), pages 399 and 401,

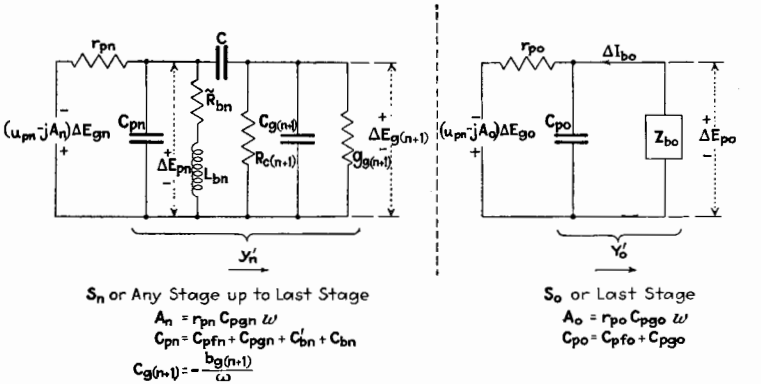


FIG. 237.—Equivalent a-c. circuits of the nth stage and last stage of an impedance-coupled amplifier.

which apply to the last stage of any multistage amplifier unless this stage feeds a detector, in which case, this stage may be similar to the earlier stages, but with slightly different constants for its plate load.

The abbreviations used in the analysis of the impedance-coupled amplifier are in most cases the same as those used for the resistance-coupled amplifier. For convenience they are given here.

$$\left.\begin{aligned} A_n &= \frac{C_{pgn}\omega}{k_{pn}} = C_{pgn}\omega r_{pn} \\ C_{pn} &= C_{pfn} + C_{pgn} + C'_{bn} + C_{bn} \\ C'_{bn} &= \text{distributed capacitance of } Z_{bn} \\ C_{bn} &= \text{capacitance of added condenser} \\ C_{gn} &= -\frac{b_{gn}}{\omega} \\ k_{pn} &= \frac{1}{r_{pn}} \end{aligned}\right\} \qquad (472)$$

$$\left.\begin{aligned} K_{cn} &= \frac{1}{R_{cn}} \\ \eta_{bn} &= \frac{\tilde{R}_{bn}}{L_{bn}\omega} \end{aligned}\right\} \qquad (472)$$

The voltage amplification of the nth stage is given by Eq. (460), page 401. When y'_n is expanded $(\text{V.A.})_n$ takes the form given below

$$(\text{V.A.})_n = \frac{-\sqrt{u_{pn}^2+A_n^2}}{\sqrt{\left[\left(\frac{\tilde{R}_{bn}}{k_{pn}(\tilde{R}_{bn}^2+L_{bn}^2\omega^2)}+1\right)\left(\frac{C_{g(n+1)}}{C}+1\right)+\frac{K_{c(n+1)}+g_{g(n+1)}}{k_{pn}}\left(\frac{C_{pn}}{C}+1-\frac{L_{bn}}{C(\tilde{R}_{bn}^2+L_{bn}^2\omega^2)}\right)\right]^2+\left[\frac{1}{k_{pn}}\left(C_{pn}\omega-\frac{L_{bn}\omega}{\tilde{R}_{bn}^2+L_{bn}^2\omega^2}\right)\left(\frac{C_{g(n+1)}}{C}+1\right)+\frac{C_{gn}\omega}{k_{pn}}-\left(\frac{\tilde{R}_{bn}}{k_{pn}(\tilde{R}_{bn}^2+L_{bn}^2\omega^2)}+1\right)\left(\frac{K_{c(n+1)}+g_{g(n+1)}}{C\omega}\right)\right]^2}} \qquad (473)$$

Except in very special cases η_{bn} is less than 0.1, so that η_{bn}^2 may be neglected in comparison with unity. Equation (473) then reduces to the simpler approximate form

$$(\text{V.A.})_n = \frac{-\sqrt{u_{pn}^2+A_n^2}}{\sqrt{\left[\left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)\left(\frac{C_{g(n+1)}}{C}+1\right)+\frac{K_{c(n+1)}+g_{g(n+1)}}{k_{pn}}\left(\frac{C_{pn}}{C}+1-\frac{1}{L_{bn}C\omega^2}\right)\right]^2+\left[\frac{1}{k_{pn}}\left(C_{pn}\omega-\frac{1}{L_{bn}\omega(1+\eta_{bn}^2)}\right)\left(\frac{C_{g(n+1)}}{C}+1\right)+\frac{C_{g(n+1)}\omega}{k_{pn}}-\left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)\left(\frac{K_{c(n+1)}+g_{g(n+1)}}{C\omega}\right)\right]^2}} \qquad (474)$$

Before discussing this equation, it should be noted that, since $\tilde{R}_{bn}$ is usually small in comparison with r_{pn}, there is very little steady-voltage drop in the external plate load and hence the plate voltage is practically equal to the plate-battery voltage

$$(\text{V.A.})_n = \frac{u_{pn}}{\sqrt{\left[\left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)+\frac{K_{c(n+1)}+g_{g(n+1)}}{k_{pn}}\left(1-\frac{1}{L_{bn}C\omega^2}\right)\right]^2 + \left[\frac{(C_{pn}+C_{g(n+1)})\omega}{k_{pn}} - \frac{1}{k_{pn}L_{bn}\omega(1+\eta_{bn}^2)} - \left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)\left(\frac{K_{c(n+1)}+g_{g(n+1)}}{C\omega}\right)\right]^2}} \quad (475)$$

For a specific numerical example of the use of Eq. (475), let $u_{pn} = 10$, $r_{pn} = 15{,}000$ ohms, $\eta_{bn} = 0.1$, $L_{bn} = 100$ henries, $K_{c(n+1)} = 0.2 \times 10^{-6}$ mho, $g_{g(n+1)} = 0$, $C = 0.1\mu$f, and $C_{pn} + C_{g(n+1)} = 150\mu\mu$f. These constants substituted in Eq. (475) give the following values of $(\text{V.A.})_n$:

Frequency, Cycles per Second	$(\text{V.A.})_n$
10	3.65
100	9.57
1,000	9.94
10,000	9.87

The low amplification at low frequencies can be increased by the use of a larger value for L_{bn}. In the example, $g_{g(n+1)}$ is neglected. The effect of $g_{g(n+1)}$ is pronounced only at high audio frequencies where it tends generally to reduce $(\text{V.A.})_n$.

In general terms, the voltage amplification of an impedance-coupled amplifier for audio frequencies drops off at very low frequencies and also at high audio frequencies. The curve of amplification plotted against the frequency (or the logarithm of the frequency) varies more with frequency than the corresponding curve for the resistance-coupled amplifier.

We shall examine the several terms of Eq. (475) in order to determine which factors are responsible for the variation of amplification with frequency. Except at low frequencies, the term $\eta_{bn}/k_{pn}L_{bn}\omega$ is usually of little importance. At frequencies of about 100 cycles per second and less, this term is appreciable. As the frequency decreases, this term is one of the most important

in decreasing the amplification. To reduce its effect, L_{bn} should be increased. The term $K_{c(n+1)}/k_{pn}$, as in the case of the resistance-coupled amplifier, can be made very small by using for $R_{c(n+1)}$ a resistance of several megohms. Except at the very high audio frequencies, the factor $g_{g(n+1)}/k_{pn}$ is of little importance. It does act to reduce the amplification at high frequencies. The parenthesis $\left(1 - \frac{1}{L_{bn}C\omega^2}\right)$ is zero at resonance of L_{bn} with C, which usually occurs at a very low frequency. Below this resonant frequency, the parenthesis is negative. At higher frequencies it is positive and, when multiplied by the fraction preceding it in Eq. (475), reduces amplification at high frequencies when $g_{g(n+1)}$ is appreciable.

Consider the second bracket under the radical sign in Eq. (475). The first two terms add up to zero at the frequency for which L_{bn} is resonant with the effective capacitance across it, $C_{pn} + C_{g(n+1)}$. This resonance usually occurs at a medium frequency and causes the total amplification to be a maximum for that frequency. At lower frequencies, these terms may reach a value of -4 or -5, in which case they act to reduce the amplification considerably. At high frequencies these two terms are usually of much less importance. The sharpness of resonance, as expressed in the second bracket, is decreased by making L_{bn} large and the capacitances small. The product of the last two parentheses of the second bracket is usually negligible except at low frequencies.

The general conclusions regarding impedance-coupled amplifiers at audio frequencies are: The voltage amplification usually varies somewhat more with frequency than in resistance-coupled amplifiers. The value of the maximum amplification is not greater than u_{pn} and hence not much greater than can be obtained with resistance-coupled amplifiers. Consequently, the impedance-coupled amplifier at audio frequency has little advantage over the resistance-coupled amplifier and gives much less amplification than the transformer-coupled type to be discussed in the next chapter.

170. Experimental Determination of the Performance of an Impedance-coupled Amplifier at Audio Frequencies.—Either of the two methods described in Chap. XV for measuring the voltage amplification of a resistance-coupled amplifier may be used at audio frequencies with an impedance-coupled amplifier.

171. Case 2. Impedance-coupled Amplifier at High Frequencies.—At high frequencies, the impedance-coupled amplifier works somewhat better than the resistance-coupled amplifier. The low input admittance of the triodes which caused the failure of the resistance-coupled amplifier can be resonated by the inductive reactance of the coupling impedance, thereby raising the equivalent impedance of the plate load so that an amplification approaching u_{pn} is attained. The amplifier operates on a resonance peak, which may be made sharp or dull according to requirements. If it is desired that without tuning the amplifier operate over an octave of frequencies, the resonance peak must be dull, but this dullness is obtained only with a sacrifice in voltage amplification. However, the amplifier may be constructed

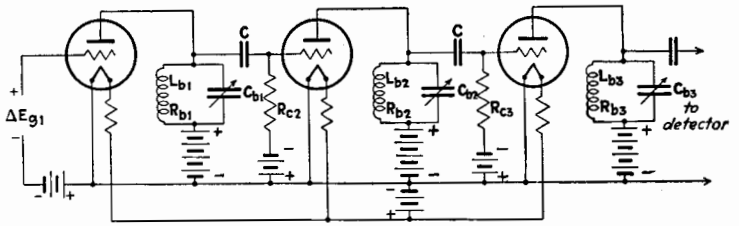


FIG. 238.—Multistage tuned impedance-coupled amplifier for radio frequencies.

with additional variable condensers, one connected across each coupling impedance. These condensers are tuned for any chosen frequency. In this case, the resonance peak can be made sharper and yet broad enough to include all of the frequencies emitted by one station, and the consequent voltage amplification per stage can be made to attain a fairly large fraction of u_{pn}.

The connections of the tuned amplifier are given in Fig. 238. The last stage is assumed to have a plate load similar in structure to that of the first two stages. The voltage across the plate load of this last stage is impressed upon a detector in the usual arrangement. The capacitances C_{b1}, C_{b2}, and C_{b3} of the variable air condensers are included in the capacitances C_{p1}, C_{p2}, and C_{p3}, which appear as C_{pn} in Eq. (474). Generally the inductances L_{bn} are coils without iron, wound with fine wire on cylinders of a few centimeters diameter. These coils have a small inductance, of the order of 0.1 millihenry, and a resistance of the order of

10 ohms. The stopping condenser C might have a capacitance of the order of 0.01μf.

The approximations used in reducing Eq. (474) to Eq. (475) cannot be used at radio frequencies. We shall therefore refer to Eq. (474). At radio frequencies, A_n cannot be neglected, but its square is usually small in comparison with u_{pn}^2. The input admittance of the next triode, expressed in terms of the components $g_{g(n+1)}$ and $b_{g(n+1)}$, is of considerable importance at radio frequencies. The most important terms under the radical of Eq. (474) are the first two of the second bracket. The sum of these terms passes through zero at the resonant frequency and gives a maximum amplification at this frequency. On account of the appreciable values of some of the other terms, the maximum value of the amplification is smaller than u_{pn}. The maximum voltage amplification per stage is usually of the order of one-third to one-half of u_{pn}. The transformer-coupled amplifier gives somewhat higher values of amplification and is more often employed in practice.

Referring again to Eq. (474), if L_{bn} is increased, the various terms containing L_{bn} are decreased. The amplification is thus increased, but the resonance peak is less sharp. A large inductance demands a small tuning condenser. Consequently, the distributed capacitance of the coil added to the tube capacitances, $C_{pf} + C_{pg}$, constitutes a larger fraction of the total capacitance across the coil. The range of frequencies which the tuning condenser is able to cover is thus reduced. In general, the plate-circuit inductance L_{bn} should be made as large as possible, consistent with obtaining the required tuning range, and η_b should be as small as practicable.

The input conductance of the triodes in a resistance-coupled amplifier is positive, because the plate load is a resistance-capacitance circuit. In the multistage impedance-coupled amplifier, the plate load may be highly inductive for some frequencies. As shown in Chap. XI, the input conductance then may be negative. For example, g_{g2} may be so strongly negative that the losses in the plate load of the first stage are supplied by the second triode, and self-oscillation takes place in the plate load at some frequency for which the reactance is zero. Self-oscillation may also arise in the tuned circuit connected to the grid of the first stage if the input conductance of the first stage is sufficiently negative and the react-

ance of the circuit is zero. Since the input conductance of any stage may be negative for frequencies lower than the frequency for which its plate load has zero reactance, the tendency to oscillate is accentuated if the condenser of the tuned circuit connected to the plate circuit is tuned to a *higher* frequency than the resonant frequency of the tuned circuit connected to the grid.

This tendency to self-oscillation together with the very variable and unstable performance, due to the coupling between stages through the C_{pg} of the triodes, makes it difficult to design a stable and at the same time efficient amplifier of the type discussed in this chapter. In Chap. XVIII, methods will be presented for reducing or eliminating the effects due to the grid-to-plate capacitance of the triode. The design of stable radio-frequency amplifiers then becomes a much simpler process. The values of $g_{g(n+1)}$ and $C_{g(n+1)}$ to be used in the equations developed in this chapter, when these neutralizing schemes are used, will be pointed out in Chap. XVIII.

172. Experimental Determination of the Performance of Impedance-coupled Amplifier at Radio Frequencies.—The curve of voltage amplification, or the logarithm of (V.A.), plotted against frequency (*not* log of frequency for radio-frequency amplifiers) is best obtained by the first method given in Chap. XV. The input voltage is generally so varied by a thoroughly shielded attenuator as to give always the same reading of the vacuum-tube voltmeter which is substituted for the usual detector. An obvious objection to this method is that the vacuum-tube voltmeter may not provide the same plate-circuit impedance as that of the detector for which the voltmeter is substituted. If this objection is important and the performance of the amplifier as a part of a receiving set is desired, the detector and audio amplifier of the set may be used as the detecting device. A modulated signal, varied by an attenuator, may be impressed on the input to give always the same audio-frequency output voltage. This output voltage may be determined either by a vacuum-tube voltmeter or by balancing the output against a constant current drawn from the source of modulating current, as shown in Fig. 239. R_a, L_a, and C_a are adjusted to simulate the input circuit of the amplifier. The elements R_1, R_2, and M_2 are set to correspond to a reasonable output voltage of the audio-frequency amplifier, and the attenuator is adjusted

$$(\text{V.A.})_n =$$

$$\frac{u_{pn}}{\sqrt{\left[\left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)+\frac{K_{c(n+1)}+g_{g(n+1)}}{k_{pn}}\left(1-\frac{1}{L_{bn}C\omega^2}\right)\right]^2 + \left[\frac{(C_{pn}+C_{g(n+1)})\omega}{k_{pn}} - \frac{1}{k_{pn}L_{bn}\omega(1+\eta_{bn}^2)} - \left(\frac{\eta_{bn}}{k_{pn}L_{bn}\omega}+1\right)\left(\frac{K_{c(n+1)}+g_{g(n+1)}}{C\omega}\right)\right]^2}} \quad (475)$$

For a specific numerical example of the use of Eq. (475), let $u_{pn} = 10$, $r_{pn} = 15{,}000$ ohms, $\eta_{bn} = 0.1$, $L_{bn} = 100$ henries, $K_{c(n+1)} = 0.2 \times 10^{-6}$ mho, $g_{g(n+1)} = 0$, $C = 0.1\mu$f, and $C_{pn} + C_{g(n+1)} = 150\mu\mu$f. These constants substituted in Eq. (475) give the following values of $(\text{V.A.})_n$:

Frequency, Cycles per Second	$(\text{V.A.})_n$
10	3.65
100	9.57
1,000	9.94
10,000	9.87

The low amplification at low frequencies can be increased by the use of a larger value for L_{bn}. In the example, $g_{g(n+1)}$ is neglected. The effect of $g_{g(n+1)}$ is pronounced only at high audio frequencies where it tends generally to reduce $(\text{V.A.})_n$.

In general terms, the voltage amplification of an impedance-coupled amplifier for audio frequencies drops off at very low frequencies and also at high audio frequencies. The curve of amplification plotted against the frequency (or the logarithm of the frequency) varies more with frequency than the corresponding curve for the resistance-coupled amplifier.

We shall examine the several terms of Eq. (475) in order to determine which factors are responsible for the variation of amplification with frequency. Except at low frequencies, the term $\eta_{bn}/k_{pn}L_{bn}\omega$ is usually of little importance. At frequencies of about 100 cycles per second and less, this term is appreciable. As the frequency decreases, this term is one of the most important

in decreasing the amplification. To reduce its effect, L_{bn} should be increased. The term $K_{c(n+1)}/k_{pn}$, as in the case of the resistance-coupled amplifier, can be made very small by using for $R_{c(n+1)}$ a resistance of several megohms. Except at the very high audio frequencies, the factor $g_{g(n+1)}/k_{pn}$ is of little importance. It does act to reduce the amplification at high frequencies. The parenthesis $\left(1 - \frac{1}{L_{bn}C\omega^2}\right)$ is zero at resonance of L_{bn} with C, which usually occurs at a very low frequency. Below this resonant frequency, the parenthesis is negative. At higher frequencies it is positive and, when multiplied by the fraction preceding it in Eq. (475), reduces amplification at high frequencies when $g_{g(n+1)}$ is appreciable.

Consider the second bracket under the radical sign in Eq. (475). The first two terms add up to zero at the frequency for which L_{bn} is resonant with the effective capacitance across it, $C_{pn} + C_{g(n+1)}$. This resonance usually occurs at a medium frequency and causes the total amplification to be a maximum for that frequency. At lower frequencies, these terms may reach a value of -4 or -5, in which case they act to reduce the amplification considerably. At high frequencies these two terms are usually of much less importance. The sharpness of resonance, as expressed in the second bracket, is decreased by making L_{bn} large and the capacitances small. The product of the last two parentheses of the second bracket is usually negligible except at low frequencies.

The general conclusions regarding impedance-coupled amplifiers at audio frequencies are: The voltage amplification usually varies somewhat more with frequency than in resistance-coupled amplifiers. The value of the maximum amplification is not greater than u_{pn} and hence not much greater than can be obtained with resistance-coupled amplifiers. Consequently, the impedance-coupled amplifier at audio frequency has little advantage over the resistance-coupled amplifier and gives much less amplification than the transformer-coupled type to be discussed in the next chapter.

170. Experimental Determination of the Performance of an Impedance-coupled Amplifier at Audio Frequencies.—Either of the two methods described in Chap. XV for measuring the voltage amplification of a resistance-coupled amplifier may be used at audio frequencies with an impedance-coupled amplifier.

171. Case 2. Impedance-coupled Amplifier at High Frequencies.—At high frequencies, the impedance-coupled amplifier works somewhat better than the resistance-coupled amplifier. The low input admittance of the triodes which caused the failure of the resistance-coupled amplifier can be resonated by the inductive reactance of the coupling impedance, thereby raising the equivalent impedance of the plate load so that an amplification approaching u_{pn} is attained. The amplifier operates on a resonance peak, which may be made sharp or dull according to requirements. If it is desired that without tuning the amplifier operate over an octave of frequencies, the resonance peak must be dull, but this dullness is obtained only with a sacrifice in voltage amplification. However, the amplifier may be constructed

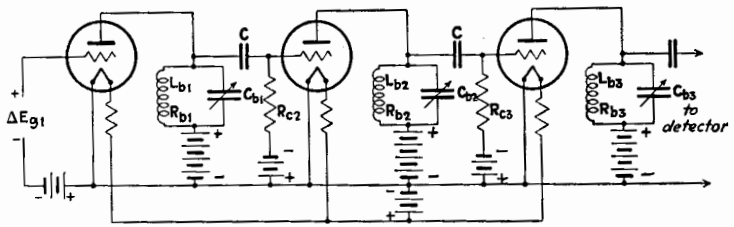


FIG. 238.—Multistage tuned impedance-coupled amplifier for radio frequencies.

with additional variable condensers, one connected across each coupling impedance. These condensers are tuned for any chosen frequency. In this case, the resonance peak can be made sharper and yet broad enough to include all of the frequencies emitted by one station, and the consequent voltage amplification per stage can be made to attain a fairly large fraction of u_{pn}.

The connections of the tuned amplifier are given in Fig. 238. The last stage is assumed to have a plate load similar in structure to that of the first two stages. The voltage across the plate load of this last stage is impressed upon a detector in the usual arrangement. The capacitances C_{b1}, C_{b2}, and C_{b3} of the variable air condensers are included in the capacitances C_{p1}, C_{p2}, and C_{p3}, which appear as C_{pn} in Eq. (474). Generally the inductances L_{bn} are coils without iron, wound with fine wire on cylinders of a few centimeters diameter. These coils have a small inductance, of the order of 0.1 millihenry, and a resistance of the order of

10 ohms. The stopping condenser C might have a capacitance of the order of 0.01μf.

The approximations used in reducing Eq. (474) to Eq. (475) cannot be used at radio frequencies. We shall therefore refer to Eq. (474). At radio frequencies, A_n cannot be neglected, but its square is usually small in comparison with u_{pn}^2. The input admittance of the next triode, expressed in terms of the components $g_{g(n+1)}$ and $b_{g(n+1)}$, is of considerable importance at radio frequencies. The most important terms under the radical of Eq. (474) are the first two of the second bracket. The sum of these terms passes through zero at the resonant frequency and gives a maximum amplification at this frequency. On account of the appreciable values of some of the other terms, the maximum value of the amplification is smaller than u_{pn}. The maximum voltage amplification per stage is usually of the order of one-third to one-half of u_{pn}. The transformer-coupled amplifier gives somewhat higher values of amplification and is more often employed in practice.

Referring again to Eq. (474), if L_{bn} is increased, the various terms containing L_{bn} are decreased. The amplification is thus increased, but the resonance peak is less sharp. A large inductance demands a small tuning condenser. Consequently, the distributed capacitance of the coil added to the tube capacitances, $C_{pf} + C_{pg}$, constitutes a larger fraction of the total capacitance across the coil. The range of frequencies which the tuning condenser is able to cover is thus reduced. In general, the plate-circuit inductance L_{bn} should be made as large as possible, consistent with obtaining the required tuning range, and η_b should be as small as practicable.

The input conductance of the triodes in a resistance-coupled amplifier is positive, because the plate load is a resistance-capacitance circuit. In the multistage impedance-coupled amplifier, the plate load may be highly inductive for some frequencies. As shown in Chap. XI, the input conductance then may be negative. For example, g_{g2} may be so strongly negative that the losses in the plate load of the first stage are supplied by the second triode, and self-oscillation takes place in the plate load at some frequency for which the reactance is zero. Self-oscillation may also arise in the tuned circuit connected to the grid of the first stage if the input conductance of the first stage is sufficiently negative and the react-

ance of the circuit is zero. Since the input conductance of any stage may be negative for frequencies lower than the frequency for which its plate load has zero reactance, the tendency to oscillate is accentuated if the condenser of the tuned circuit connected to the plate circuit is tuned to a *higher* frequency than the resonant frequency of the tuned circuit connected to the grid.

This tendency to self-oscillation together with the very variable and unstable performance, due to the coupling between stages through the C_{pg} of the triodes, makes it difficult to design a stable and at the same time efficient amplifier of the type discussed in this chapter. In Chap. XVIII, methods will be presented for reducing or eliminating the effects due to the grid-to-plate capacitance of the triode. The design of stable radio-frequency amplifiers then becomes a much simpler process. The values of $g_{g(n+1)}$ and $C_{g(n+1)}$ to be used in the equations developed in this chapter, when these neutralizing schemes are used, will be pointed out in Chap. XVIII.

172. Experimental Determination of the Performance of Impedance-coupled Amplifier at Radio Frequencies.—The curve of voltage amplification, or the logarithm of (V.A.), plotted against frequency (*not* log of frequency for radio-frequency amplifiers) is best obtained by the first method given in Chap. XV. The input voltage is generally so varied by a thoroughly shielded attenuator as to give always the same reading of the vacuum-tube voltmeter which is substituted for the usual detector. An obvious objection to this method is that the vacuum-tube voltmeter may not provide the same plate-circuit impedance as that of the detector for which the voltmeter is substituted. If this objection is important and the performance of the amplifier as a part of a receiving set is desired, the detector and audio amplifier of the set may be used as the detecting device. A modulated signal, varied by an attenuator, may be impressed on the input to give always the same audio-frequency output voltage. This output voltage may be determined either by a vacuum-tube voltmeter or by balancing the output against a constant current drawn from the source of modulating current, as shown in Fig. 239. R_a, L_a, and C_a are adjusted to simulate the input circuit of the amplifier. The elements R_1, R_2, and M_2 are set to correspond to a reasonable output voltage of the audio-frequency amplifier, and the attenuator is adjusted

to give silence in the telephone receivers. By this procedure, the detector and audio-frequency amplifier are always operated with the same intensity of signal and hence do not in any way affect the results. A vacuum-tube voltmeter may be substituted for the telephone receivers or may be connected directly

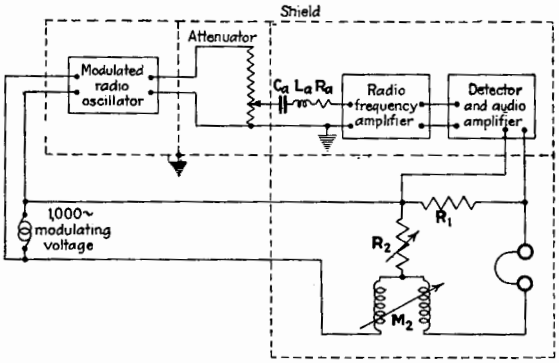


FIG. 239.—Method of measuring the voltage amplification of a radio-frequency amplifier.

across the output of the audio-frequency amplifier. It may be well to reduce the modulation frequency to a low value in order that the radio-frequency spectrum may be narrow.

Thorough shielding is of the utmost importance in all measurements of this type at radio frequencies. Stray effects may be difficult to eliminate, but with care in arranging the apparatus the method may be made to work very satisfactorily.

CHAPTER XVII

COMBINATIONS OF TRIODES AS LOW-POWER TRANSFORMER-COUPLED AMPLIFIERS

Perhaps the most important type of coupling between the stages of an amplifier is transformer coupling. This type of coupling is used at both high and low frequencies and may have very different forms. The transformer may be without iron or it may have an iron core. Either or both of the windings of the transformer may be tuned.

173. Fundamental Theory of Transformer-coupled Amplifier. The simple approximate theory of transformer coupling, in which the effects of tube capacitances and distributed capacitances are neglected, was given in Chap. XII.

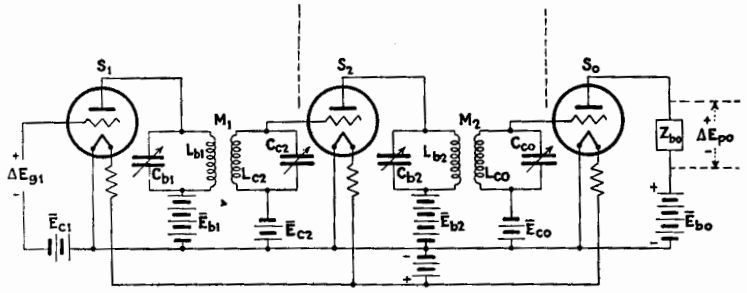


FIG. 240.—Multistage transformer-coupled amplifier.

The general theory will now be presented, from which the special cases can be deduced. The diagram of connections is shown in Fig. 240. If the amplifier is for audio frequencies, the last stage may supply a loud-speaker or similar load, denoted by Z_{bo}. In this case, the last stage does not differ from the last stage as treated in Chap. XV. If the audio amplifier is used to feed a line, or if the amplifier is for radio frequencies and feeds a detector, Z_{bo} is replaced by a third transformer, the constants of which depend upon the impedance of the line or detector which follows. As in Chap. XVI, only the stages before the last will be discussed.

Figure 241 gives the equivalent-circuit diagram of the nth stage of the amplifier of Fig. 240. In this diagram C_{pn} is the total equivalent capacitance across the primary coil and comprises C_{pfn}, C_{pgn}, the distributed capacitance C'_{bn} of the coil L_{bn}, and any added tuning capacitance C_{bn}. Similarly, $C_{g(n+1)}$ comprises the distributed capacitance of coil $L_{c(n+1)}$, the equivalent input capacitance of the next tube, and any added tuning capacitance.

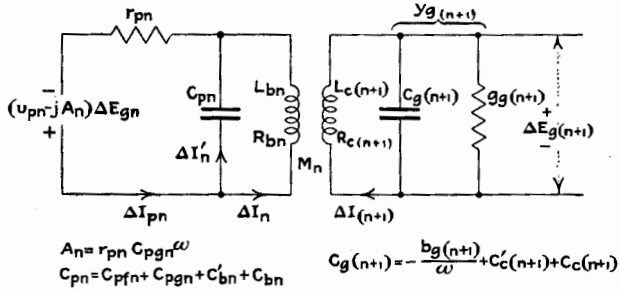


FIG. 241.—Equivalent a-c. circuits of the transformer-coupled amplifier.

The admittance of $C_{g(n+1)}$ and $g_{g(n+1)}$ in parallel, denoted by $\mathbf{y}_{g(n+1)}$ is

$$\mathbf{y}_{g(n+1)} = g_{g(n+1)} + jC_{g(n+1)}\omega \tag{476}$$

The total impedance of the secondary circuit is

$$\begin{aligned}\mathbf{Z}_{(n+1)} &= \left(R_{c(n+1)} + \frac{g_{g(n+1)}}{g^2_{g(n+1)} + C^2_{g(n+1)}\omega^2}\right) \\ &\qquad + j\left(L_{c(n+1)}\omega - \frac{C_{g(n+1)}\omega}{g^2_{g(n+1)} + C^2_{g(n+1)}\omega^2}\right) \\ &= R_{(n+1)} + jX_{(n+1)}\end{aligned} \tag{477}$$

where $R_{c(n+1)}$ and $X_{c(n+1)}$ are the equivalent series resistance and reactance of the secondary circuit.

The equations for the solution of the circuits of Fig. 241 are

$$\left.\begin{aligned} r_{pn}\Delta\mathbf{I}_{pn} - \frac{j\Delta\mathbf{I}'_n}{C_{pn}\omega} &= (u_{pn} - jA_n)\Delta\mathbf{E}_{gn} \\ (R_{bn} + jL_{bn}\omega)\Delta\mathbf{I}_n + jM_n\omega\Delta\mathbf{I}_{(n+1)} + \frac{j\Delta\mathbf{I}'_n}{C_{pn}\omega} &= 0 \end{aligned}\right\} \tag{478}$$

$$\left.\begin{aligned} &j M_n \omega \Delta \mathbf{I}_n + \mathbf{Z}_{(n+1)} \Delta \mathbf{I}_{(n+1)} = 0 \\ &\Delta \mathbf{I}_{pn} = \Delta \mathbf{I}_n + \Delta \mathbf{I}_n' \\ &\mathbf{y}_{g(n+1)} \Delta \mathbf{E}_{g(n+1)} = \Delta \mathbf{I}_{(n+1)} \end{aligned}\right\} \quad (478)$$

If currents $\Delta \mathbf{I}_n'$ and $\Delta \mathbf{I}_{pn}$ are eliminated, the following equations result and are in the canonical form for simple coupled circuits.

$$\left.\begin{aligned} \Delta \mathbf{I}_n &= \frac{\mathbf{Z}_{(n+1)} \Delta \mathbf{E}_n}{\mathbf{Z}_n \mathbf{Z}_{(n+1)} + M_n^2 \omega^2} \\ \Delta \mathbf{I}_{(n+1)} &= \frac{-j M_n \omega \Delta \mathbf{E}_n}{\mathbf{Z}_n \mathbf{Z}_{(n+1)} + M_n^2 \omega^2} = \Delta \mathbf{E}_{g(n+1)} \mathbf{y}_{g(n+1)} \end{aligned}\right\} \quad (479)$$

where $\mathbf{Z}_{(n+1)}$ is given by Eq. (477), and $\mathbf{Z}_n$ and $\Delta \mathbf{E}_n$ are given by the following equations:

$$\mathbf{Z}_n = \left[R_{bn} + \frac{r_{pn}}{(r_{pn} C_{pn} \omega)^2 + 1} \right] + j \left[L_{bn} \omega - \frac{1}{C_{pn} \omega \left(1 + \frac{1}{(r_{pn} C_{pn} \omega)^2} \right)} \right] \quad (480)$$

$$\Delta \mathbf{E}_n = \frac{u_{pn} - j A_n}{1 + j r_{pn} C_{pn} \omega} \Delta \mathbf{E}_{gn} \quad (481)$$

From the second equation of Eq. (479), we obtain *in complex form* the expression for the voltage amplification, or

$$(\mathbf{V.A.})_n = \frac{\Delta \mathbf{E}_{g(n+1)}}{\Delta \mathbf{E}_{gn}} = \frac{-j M_n \omega (u_{pn} - j A_n)}{(1 + j r_{pn} C_{pn} \omega)(g_{g(n+1)} + j C_{g(n+1)} \omega)(\mathbf{Z}_n \mathbf{Z}_{(n+1)} + M_n^2 \omega^2)} \quad (482)$$

The *magnitude* of the voltage amplification for any stage n can now be deduced from Eq. (482) and is

$$(\text{V.A.})_n = \frac{M_n \omega \sqrt{u_{pn}^2 + A_n^2}}{\sqrt{[1 + (r_{pn} C_{pn} \omega)^2][g_{g(n+1)} + C_{g(n+1)}^2 \omega^2][R_n + X_n^2]}} \cdot \frac{1}{\sqrt{\left[R_{(n+1)} + \frac{M_n^2 \omega^2 R_n}{R_n^2 + X_n^2} \right]^2 + \left[X_{(n+1)} - \frac{M_n \omega^2 X_n}{R_n^2 + X_n^2} \right]^2}} \quad (483)$$

The abbreviations used in Eq. (483) are given here for reference.

$$
\left.
\begin{aligned}
A_n &= C_{pgn}\omega r_{pn} \\
C_{pn} &= C_{pfn} + C_{pgn} + C'_{bn} + C_{bn} \\
C'_{bn} &= \text{distributed capacitance of coil } L_{bn} \\
C_{bn} &= \text{capacitance of physical condenser connected across } L_{bn} \\
C_{g(n+1)} &= -\frac{b_{g(n+1)}}{\omega} + C'_{c(n+1)} + C_{c(n+1)} \\
C'_{c(n+1)} &= \text{distributed capacitance of coil } L_{c(n+1)} \\
C_{c(n+1)} &= \text{capacitance of physical condenser connected across } L_{c(n+1)} \\
R_n &= R_{bn} + \frac{r_{pn}}{(r_{pn}C_{pn}\omega)^2 + 1} \\
X_n &= L_{bn}\omega - \frac{1}{C_{pn}\omega\left(1 + \frac{1}{(r_{pn}C_{pn}\omega)^2}\right)} \\
R_{(n+1)} &= R_{c(n+1)} + \frac{g_{g(n+1)}}{g^2_{g(n+1)} + C^2_{g(n+1)}\omega^2} \\
X_{(n+1)} &= L_{c(n+1)}\omega - \frac{1}{C_{g(n+1)}\omega\left(1 + \frac{g^2_{g(n+1)}}{C^2_{g(n+1)}\omega^2}\right)}
\end{aligned}
\right\} \quad (484)
$$

Equation (483) can be expressed entirely in terms of ratios, giving a form which is much more convenient for evaluation.

$$
(\text{V.A.})_n = \frac{\tau_n\sqrt{u^{pn}+A_n^2}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}}{|g_{g(n+1)}|L_{c(n+1)}\omega\sqrt{\left[1+(r_{pn}C_{pn}\omega)^2\right]\left[1+\left(\frac{C_{g(n+1)}\omega}{g_{g(n+1)}}\right)^2\right]\left[\eta_n^2+\left(1-\frac{\omega_n^2}{\omega^2}\right)^2\right]}}
$$

$$
\sqrt{\left[\eta_{(n+1)}+\frac{\tau_n^2\eta_n}{\eta_n^2+\left(1-\frac{\omega_n^2}{\omega^2}\right)^2}\right]^2+\left[1-\frac{\omega_{(n+1)}^2}{\omega^2}-\frac{\tau_n^2\left(1-\frac{\omega_n^2}{\omega^2}\right)}{\eta_n^2+\left(1-\frac{\omega_n^2}{\omega^2}\right)^2}\right]^2} \quad (485)
$$

Another form for Eq. (485) is

$$(\text{V.A.})_n = \frac{\tau_n \sqrt{u_{pn}^2 + A_n^2} \sqrt{\frac{L_{c(n+1)}}{L_{bn}}} \frac{\omega_{(n+1)}\omega_{c(n+1)}}{\omega^2}}{\sqrt{\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2} \sqrt{\left[\eta_{(n+1)} + \frac{\tau_n^2 \eta_n}{\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2}\right]^2 + \left[1 - \frac{\omega_{(n+1)}^2}{\omega^2} - \frac{\tau_n^2\left(1 - \frac{\omega_n^2}{\omega^2}\right)}{\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2}\right]^2}} \tag{486}$$

In Eqs. (485) and (486), the following new abbreviations appear.

$$\left.\begin{aligned}
\tau_n &= \frac{M_n}{\sqrt{L_{bn} L_{c(n+1)}}} \\
\eta_n &= \eta_{bn} + \frac{r_{pn}/L_{bn}\omega}{(r_{pn} C_{pn}\omega)^2 + 1} \\
\eta_{bn} &= \frac{R_{bn}}{L_{bn}\omega} \quad \text{(where } R_{bn} \text{ is the resistance unenhanced by the distributed capacitance of the coil)} \\
\frac{\omega_n^2}{\omega^2} &= \frac{\omega_{bn}^2}{\omega^2}\left[\frac{(r_{pn} C_{pn}\omega)^2}{(r_{pn} C_{pn}\omega)^2 + 1}\right] = \frac{r_{pn}^2 C_{pn}}{L_{bn}}\left[\frac{1}{(r_{pn} C_{pn}\omega)^2 + 1}\right] \\
\omega_{bn}^2 &= \frac{1}{L_{bn} C_{pn}} \\
\eta_{(n+1)} &= \eta_{c(n+1)} + \frac{1}{g_{g(n+1)} L_{c(n+1)}\omega} \cdot \frac{1}{1 + (C_{g(n+1)}\omega / g_{g(n+1)})^2} \\
\eta_{c(n+1)} &= \frac{R_{c(n+1)}}{L_{c(n+1)}\omega} \quad \text{(where } R_{c(n+1)} \text{ is the resistance unenhanced by the distributed capacitance of the coil)} \\
\frac{\omega_{(n+1)}^2}{\omega^2} &= \frac{\omega_{c(n+1)}^2}{\omega^2}\left[\frac{1}{1 + (g_{g(n+1)}/C_{g(n+1)}\omega)^2}\right] = \\
&\qquad \frac{1}{L_{c(n+1)} C_{g(n+1)}\omega^2 + \frac{g_{g(n+1)}^2 L_{c(n+1)}}{C_{g(n+1)}}} \\
\omega_{c(n+1)}^2 &= \frac{1}{L_{c(n+1)} C_{g(n+1)}}
\end{aligned}\right\} \tag{487}$$

For reference, the equivalent admittance $\mathbf{y}_{bn}$ of the plate load can be obtained from Eq. (478) or directly from the parallel admittance of $C'_{bn} + C_{bn}$ and the transformer. Accordingly,

$$\mathbf{y}_{bn} = \frac{1}{L_{bn}\omega} \cdot$$

$$\frac{\eta_{bn} + \dfrac{\tau_n^2 \eta_{(n+1)}}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}}{\left[\eta_{bn} + \dfrac{\tau_n^2 \eta_{(n+1)}}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}\right]^2 + \left[1 - \dfrac{\tau_n^2\left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}\right]^2}$$

$$+ j\frac{1}{L_{bn}\omega}\left\{\frac{\omega^2}{\omega_{bn}'^2}\right.$$

$$\left. - \frac{1 - \dfrac{\tau_n^2\left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}}{\left[\eta_{bn} + \dfrac{\tau_n^2 \eta_{(n+1)}}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}\right]^2 + \left[1 - \dfrac{\tau_n^2\left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)}{\eta_{(n+1)}^2 + \left(1 - \dfrac{\omega_{(n+1)}^2}{\omega^2}\right)^2}\right]^2}\right\} \tag{488}$$

where $\omega_{bn}'^2 = \dfrac{1}{(C_{pn} + C'_n)L_{bn}}$ and the other quantities are defined by Eq. (487).

Expressions (485) and (486) are so complicated that analysis is difficult, especially since $g_{g(n+1)}$ and $b_{g(n+1)}$ vary rapidly according to the character of the plate load of stage $(n + 1)$. These expressions can be reduced to simpler approximate forms for certain special cases.

174. Case 1. Limiting or Ideal Case. $C_{pn} = 0$; $C_{g(n+1)} = 0$; $g_{g(n+1)} = 0$.—Although the conditions stated never prevail in practice, it is of interest to express the voltage amplification

for this limiting case. The arrangement comprises a transformer with open secondary circuit, both windings having negligible self-capacitance, the primary winding being connected in the plate circuit of a triode. All tube capacitances are assumed negligible and there are no tuning capacitances. In accordance with the conditions,

$$\eta_n = \frac{R_{bn} + r_{pn}}{L_{bn}\omega}$$

$$\frac{\omega_n^2}{\omega^2} = 0$$

$$\eta_{(n+1)} = \infty$$

$$\frac{\omega_{(n+1)}^2}{\omega^2} = \frac{\omega_{c(n+1)}^2}{\omega^2} = \infty$$

$$A_n = 0$$

Putting these values in Eq. (485),

$$(\text{V.A.})_n = \frac{\tau_n u_{pn}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{\eta_n^2 + 1}} \tag{489}$$

Equation (489) may be considered to represent the ideal case. With no capacitances, neither tube capacitances nor circuit capacitances, the coupling should be made as great as possible. That is, τ_n should be nearly unity. If τ_n is unity, $\sqrt{L_{c(n+1)}/L_{bn}}$ is the ratio of turns in the secondary and primary windings. Therefore, it would be advantageous to make $L_{bn}\omega$ large compared with r_{pn}. Under these conditions Eq. (489) reduces to

$$(\text{V.A.})_n = -u_{pn}\frac{N_{c(n+1)}}{N_{bn}} \tag{490}$$

where $N_{c(n+1)}$ and N_{bn} are the numbers of turns in the two windings of the transformer. Barring regeneration, this expression gives the greatest amplification possible to be obtained from one stage of an untuned transformer-coupled amplifier. The negative sign indicates that the grid voltages are opposite in phase if the windings of the transformer are in the same direction.

The three quantities C_{pn}, $C_{c(n+1)}$, and $g_{g(n+1)}$, which have been assumed to be zero in this ideal case, are in practice always present. According to the type of transformer and the frequency, they have different relative effects in causing the amplification

If $g_{g(n+1)}$ is zero, making $\eta_{(n+1)}$ infinite, Eq. (492) reduces to Eq. (489). When $g_{g(n+1)}$ has a value greater than zero, the amplification is less than the ideal value given by Eq. (489). If $R_{c(n+1)}$ is negligible compared with $1/g_{g(n+1)}$, Eq. (492) further reduces to

$$(\text{V.A.})_n = \frac{u_{pn}\dfrac{N_{c(n+1)}}{N_{bn}}}{\sqrt{\eta_n^2 + \left(\dfrac{\eta_n}{\eta_{(n+1)}} + 1\right)^2 - \dfrac{1 - \tau_n^2}{\eta_{(n+1)}^2}}} \tag{493}$$

***176. Case 3.** $(r_{pn}C_{pn}\omega)^2 << 1$; $r_{pn}^2 C_{pn}/L_{bn} << 1$; $(C_{g(n+1)}\omega/g_{g(n+1)})^2 >> 1$.—This case corresponds to a transformer having a variable or fixed capacitance across the secondary winding. The plate-to-filament capacitance is assumed negligible. The real component of the input admittance of the tube connected across the secondary winding is also assumed to be negligible. Although this case is also a limiting case, it is more practical than the previous two cases.

Conditions of Eq. (487) now become

$$A_n = 0$$
$$\eta_n = \frac{R_{bn} + r_{pn}}{L_{bn}\omega}$$
$$\left(\frac{\omega_n}{\omega}\right)^2 << 1$$
$$\eta_{(n+1)} = \eta_{c(n+1)}$$
$$\omega_{(n+1)} = \omega_{c(n+1)}$$

In accordance with these values, Eq. (486) becomes

$$(\text{V.A.})_n = \frac{\tau_n u_{pn}\sqrt{\dfrac{L_{c(n+1)}}{L_{bn}}}\left(\dfrac{\omega_{c(n+1)}}{\omega}\right)^2}{\sqrt{\left[\eta_n^2 + 1\right]\left[\eta_{c(n+1)}^2 + \left(1 - \dfrac{\omega_{c(n+1)}^2}{\omega^2}\right)^2\right] + \tau_n^2\left[\tau_n^2 - 2\left(1 - \dfrac{\omega_{c(n+1)}^2}{\omega^2}\right) + 2\eta_n\eta_{c(n+1)}\right]}} \tag{494}$$

If the capacitance $C_{g(n+1)}$ is small, so that the secondary circuit is not near resonance, or in other words if $(\omega_{c(n+1)}/\omega)^2$ is very large, the effect of $C_{g(n+1)}$ is to reduce the amplification

below the value expressed for the ideal case of Eq. (489). As $C_{g(n+1)}$ is increased, the amplification decreases to a minimum and then rises again as resonance is approached.

If the capacitance across the secondary winding is variable, it may be adjusted to give a maximum voltage amplification. Differentiating Eq. (494), the condition for max. $(\text{V.A.})_n$ is

$$\left(\frac{\omega_{c(n+1)}}{\omega}\right)^2 = \left(1 - \frac{\tau_n^2}{\eta_n^2 + 1}\right) + \frac{\left(\eta_{c(n+1)} + \frac{\tau_n^2 \eta_n}{\eta_n^2 + 1}\right)^2}{1 - \frac{\tau_n^2}{\eta_n^2 + 1}} \tag{495}$$

If this value is substituted in Eq. (494), the maximum voltage amplification is

$$\text{Max. } (\text{V.A.})_n = \frac{\tau_n u_{pn} \sqrt{\frac{L_{c(n+1)}}{L_{bn}}}}{\sqrt{\eta_n^2 + 1}} \sqrt{1 + \frac{\left(1 - \frac{\tau_n^2}{\eta_n^2 + 1}\right)^2}{\left(\eta_{c(n+1)} + \frac{\tau_n^2 \eta_n}{\eta_n^2 + 1}\right)^2}} \tag{496}$$

***177. Case 4.** $(r_{pn} C_{pn} \omega)^2 << 1; r_{pn}^2 C_{pn}/L_{bn} << 1$.—This case is the same as Case 3 except that the input admittance of the next tube is not assumed to be negligible.

The conditions of Eq. (487) become

$$A_n = 0$$

$$\eta_n = \frac{R_{bn} + r_{pn}}{L_{bn}\omega}$$

$$\left(\frac{\omega_n}{\omega}\right)^2 << 1$$

$$\eta_{(n+1)} = \eta_{c(n+1)} + \frac{1}{g_{g(n+1)} L_{c(n+1)} \omega} \cdot \frac{1}{\left(1 + \frac{C_{g(n+1)}\omega}{g_{g(n+1)}}\right)^2}$$

$$\eta_{c(n+1)} = \frac{R_{c(n+1)}}{L_{c(n+1)}\omega}$$

$$\left(\frac{\omega_{(n+1)}}{\omega}\right)^2 = \left(\frac{\omega_{c(n+1)}}{\omega}\right)^2 \cdot \frac{1}{1 + (g_{g(n+1)}/C_{g(n+1)}\omega)^2}$$

$$\omega_{c(n+1)} = \frac{1}{L_{c(n+1)} C_{g(n+1)}}$$

If these values are substituted in the general expression of Eq. (486),

$$(\text{V.A.})_n = \frac{\tau_n u_{pn}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{\left[\eta_n^2+1\right]\left[\left(g_{g(n+1)}L_{c(n+1)}\omega\right)^2+\left(\frac{\omega}{\omega_{c(n+1)}}\right)^4\right]}}$$

$$\sqrt{\left(\eta_{(n+1)}+\frac{\tau_n^2\eta_n}{\eta_n^2+1}\right)^2+\left(1-\frac{\omega_{(n+1)}^2}{\omega^2}-\frac{\tau_n^2}{\eta_n^2+1}\right)^2} \qquad (497)$$

If Eq. (497) is differentiated to find the value of $\omega_{c(n+1)}/\omega$ which gives a maximum, noting that $\eta_{(n+1)}$ and $\omega_{(n+1)}/\omega$ involve $\omega_{c(n+1)}/\omega$, it develops that the value of $\omega_{c(n+1)}/\omega$ is the same as for Case 3 and is given by Eq. (495).

Substituting this value of $\omega_{c(n+1)}/\omega$ in Eq. (497) gives the maximum value of voltage amplification for any value of τ_n. The result is

$$\text{Max. (V.A.)}_n = \frac{\tau_n u_{pn}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{\eta_n^2+1}\sqrt{1-\dfrac{\left(1-\dfrac{\tau_n^2}{\eta_n^2+1}\right)^2}{\left(1-\dfrac{\tau_n^2}{\eta_n^2+1}\right)^2+\left(\eta_{c(n+1)}+\dfrac{\tau_n^2\eta_n}{\eta_n^2+1}\right)^2}}}$$

$$+\, g_{g(n+1)}L_{c(n+1)}\omega\left\{2\left(\eta_{c(n+1)}+\frac{\tau_n^2\eta_n}{\eta_n^2+1}\right)\right.$$

$$\left.+\, g_{g(n+1)}L_{c(n+1)}\omega\left[\left(1-\frac{\tau_n^2}{\eta_n^2+1}\right)^2+\left(\eta_{c(n+1)}+\frac{\tau_n^2\eta_n}{\eta_n^2+1}\right)^2\right]\right\} \qquad (498)$$

If in Eq. (498) $g_{g(n+1)}$ is made zero, Eq. (498) reduces to Eq. (496). It is obvious from Eq. (498) that a positive value of $g_{g(n+1)}$ reduces the amplification, while a negative value increases the amplification and may result in self-oscillation.

***178. Case 5.** $g_{g(n+1)}L_{c(n+1)}\omega << 1$; $(C_{g(n+1)}\omega/g_{g(n+1)})^2 << 1$. This case is similar to Case 3 except that the primary winding of the transformer is tuned instead of the secondary winding.

The conditions of Eq. (487) become

$$\eta_n = \eta_{bn} + \frac{r_{pn}/L_{bn}\omega}{(r_{pn}C_{pn}\omega)^2 + 1}$$

$$\left(\frac{\omega_n}{\omega}\right)^2 = \left(\frac{\omega_{bn}}{\omega}\right)^2 \cdot \frac{(r_{pn}C_{pn}\omega)^2}{(r_{pn}C_{pn}\omega)^2 + 1}$$

$$\eta_{(n+1)} >> 1$$

$$\left(\frac{\omega_{(n+1)}}{\omega}\right)^2 << 1$$

Using these values, Eq. (485) reduces to

$$(\text{V. A.})_n = \frac{\tau_n\sqrt{u_{pn}^2 + A_n^2}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{1 + (r_{pn}C_{pn}\omega)^2}\sqrt{\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2}} \tag{499}$$

$$= \frac{\tau_n\sqrt{u_{pn}^2 + A_n^2}\sqrt{L_{c(n+1)}/L_{bn}}\sqrt{1 - (\omega_n/\omega)^2}}{\sqrt{\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2}} \tag{500}$$

In order that the voltage amplification as given by Eq. (500) may be a maximum, it must follow that

$$1 - \frac{\omega_n^2}{\omega^2} = \eta_n \tag{501}$$

Putting this value in Eq. (500) and making τ_n equal to unity, the maximum amplification is

$$\text{Max. (V.A.)}_n = -\frac{\sqrt{u_{pn}^2 + A_n^2}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{2\eta_n}} \tag{502}$$

***179. Case 6.** $g_{g(n+1)}^2 L_{c(n+1)}/C_{g(n+1)} >> 1$; $(C_{g(n+1)}\omega/g_{g(n+1)})^2 << 1$.—For this case, consider the amplification obtainable if the input admittance of the next stage is not negligible and if the primary winding is tuned. The conditions of Eq. (487) become

$$\eta_n = \eta_{bn} + \frac{r_{pn}/L_{bn}\omega}{(r_{pn}C_{pn}\omega)^2 + 1}$$

$$\left(\frac{\omega_n}{\omega}\right)^2 = \left(\frac{\omega_{bn}}{\omega}\right)^2 \cdot \frac{(r_{pn}C_{pn}\omega)^2}{(r_{pn}C_{pn}\omega)^2 + 1} = \frac{r_{pn}^2 C_{pn}}{L_{bn}}\left[\frac{1}{(r_{pn}C_{pn}\omega)^2 + 1}\right]$$

$$\eta_{(n+1)} = \eta_{c(n+1)} + \frac{1}{g_{g(n+1)}L_{c(n+1)}\omega}$$

$$\left(\frac{\omega_{(n+1)}}{\omega}\right)^2 << 1$$

Equation (485) reduces to

$$(\text{V.A.})_n = \frac{\tau_n\sqrt{u_{pn}^2 + A_n^2}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}\sqrt{1 - \frac{\omega_n^2}{\omega^2}}}{|g_{g(n+1)}|L_{c(n+1)}\omega\sqrt{\left[\eta_{(n+1)}^2 + 1\right]\left[\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2\right]}}$$

$$\overline{\overline{+ 2\tau_n^2\left[\tau_n^2 + 2\eta_n\eta_{(n+1)} - 2\left(1 - \frac{\omega_n^2}{\omega^2}\right)\right]}} \quad (503)$$

In this, as in Case 5, the greatest amplification is obtained when τ_n is as large as possible. Putting τ_n equal to unity in Eq. (503),

$$(\text{V.A.})_n = \frac{\sqrt{u_{pn}^2 + A_n^2}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}\sqrt{1 - \frac{\omega_n^2}{\omega^2}}}{|g_{g(n+1)}|L_{c(n+1)}\omega\sqrt{\left[\eta_{(n+1)}^2 + 1\right]\left[\eta_n^2 + \left(1 - \frac{\omega_n^2}{\omega^2}\right)^2\right]}}$$

$$\overline{\overline{+ 2\left[1 + 2\eta_n\eta_{(n+1)} - 2\left(1 - \frac{\omega_n^2}{\omega^2}\right)\right]}} \quad (504)$$

Equation (504) may be analyzed to find the value of ω_n^2/ω^2 for maximum voltage amplification. This value is found to be given by the relation

$$1 - \frac{\omega_n^2}{\omega^2} = \sqrt{\eta_n^2 + \frac{2(1 + 2\eta_n^2\eta_{(n+1)})}{\eta_{(n+1)}^2 + 1}} \quad (505)$$

Equation (505) substituted in Eq. (504) gives the optimum amplification for this case.

$$\text{Optimum } (\text{V.A.})_n = \frac{\sqrt{u_{pn}^2 + A_n^2}\sqrt{L_{c(n+1)}/L_{bn}}}{|g_{g(n+1)}|L_{c(n+1)}\omega\sqrt{2(\eta_{(n+1)}^2 + 1)\sqrt{\eta_n^2 + \frac{2(1 + \eta_n\eta_{(n+1)})}{\eta_{(n+1)}^2 + 1}} - 4}} \quad (506)$$

If $g_{g(n+1)}$ is zero, Eqs. (505) and (506) reduce to Eqs. (501) and (502) of Case 5.

Here, also, we should recognize the possibility of self-oscillation when $g_{g(n+1)}$ is sufficiently negative.

180. Case 7. $\frac{r_{pn}^2 C_{pn}}{L_{bn}} >> 1$; $g_{g(n+1)}^2 \frac{L_{c(n+1)}}{C_{g(n+1)}} << 1$; $\left(\frac{C_{g(n+1)}\omega}{g_{g(n+1)}}\right)^2 >> 1$; $|g_{g(n+1)} L_{c(n+1)}\omega| << 1$.—The final special case, defined by the conditions given, is one of considerable practical importance because it applies to most tuned radio-frequency amplifiers. The conditions given in Eq. (487) become

$$\eta_n = \frac{r_{pn}/L_{bn}\omega}{(r_{pn}C_{pn}\omega)^2 + 1}$$

$$\frac{\omega_n^2}{\omega^2} >> 1$$

$$\eta_{(n+1)} = \eta_{c(n+1)} + \frac{g_{g(n+1)}}{L_{c(n+1)} C_{g(n+1)}^2 \omega^3}$$

$$\omega_{(n+1)} = \omega_{c(n+1)}$$

The above relations placed in Eq. (486) give for the voltage amplification

$$(\text{V.A.})_n = \frac{\tau_n \sqrt{u_{pn}^2 + A_n^2} \sqrt{\frac{L_{c(n+1)}}{L_{bn}}} \frac{\omega_{c(n+1)}^2}{\omega^2} \sqrt{(r_{pn}C_{pn}\omega)^2 + 1}}{\frac{r_{pn}}{L_{bn}\omega} \sqrt{\left[\eta_{(n+1)} + \frac{\tau_n^2 L_{bn}\omega}{r_{pn}}\right]^2 + \left[1 - \frac{\omega_{c(n+1)}^2}{\omega^2} - \tau_n^2 L_{bn} C_{pn}\omega^2\right]^2}} \tag{507}$$

In Eq. (507), the term $\tau_n^2 L_{bn} C_{pn}\omega^2$ is generally very small compared with unity, so that maximum voltage amplification occurs when the variable capacitance $C_{c(n+1)}$ is set to make $\omega_{c(n+1)}/\omega = 1$. The maximum value of $(\text{V.A.})_n$ obtained by tuning is

$$\text{Max. (V. A.)}_n = \frac{\tau_n \sqrt{u_{pn}^2 + A_n^2} \sqrt{\frac{L_{c(n+1)}}{L_{bn}}} \sqrt{(r_{pn}C_{pn}\omega)^2 + 1}}{\frac{r_{pn}}{L_{bn}\omega_{c(n+1)}} \left[\eta_{(n+1)} + \frac{\tau_n^2 L_{bn}\omega_{c(n+1)}}{r_{pn}}\right]} \tag{508}$$

The max. (V. A.)$_n$, given by Eq. (508), has the greatest value for the particular value of τ_n for which

$$\tau_n^2 = \frac{r_{pn}\eta_{(n+1)}}{L_{bn}\omega_{c(n+1)}} \tag{509}$$

Substituting this value of τ_n in Eq. (508) gives for the optimum value of the voltage amplification

$$\left.\begin{aligned}\text{Max. max. (V. A.)}_n &= \frac{\sqrt{u_{pn}^2 + A_n^2}\,\sqrt{(r_{pn}C_{pn}\omega)^2 + 1}}{2\sqrt{\eta_{(n+1)} \cdot r_{pn}/L_{c(n+1)}\omega}} \\ &= \frac{\sqrt{u_{pn}^2 + A_n^2}\,\sqrt{(r_{pn}C_{pn}\omega)^2 + 1}}{2} \cdot \frac{L_{c(n+1)}}{M}\end{aligned}\right\} \quad (510)$$

The analysis of Case 7 may be compared with the simple treatment in Chap. XII, where the effects of input admittance and tube capacitances were neglected. Equation (509) corresponds to Eq. (316), page 302, the value of τ specified by Eq. (509) being somewhat different from that for the simple theory because of the different value of the resistance of the secondary circuit caused by the input admittance g_g of the next triode. Equation (510) corresponds to Eq. (317), page 302, but gives a somewhat different value of the max. max. (V.A.) because of the larger value of M or τ required for the optimum conditions.

The analysis of the band width or selectivity given in Chap. XII applies to the present theory, except that η_2 of Eq. (325), page 304, must be replaced by $\eta_{c(n+1)}$.

***181. Discussion of the Special Cases.**—We have developed the fundamental theory of transformer-coupled amplifiers with no special reference to the frequency range over which the amplifier is to operate. Some special cases, characterized only by the values of certain terms in the equation for voltage amplification, have been analyzed. We shall now consider these special cases from the point of view of practice.

One of the characteristics of the first four cases is the negligibility of the quantity $(r_{pn}C_{pn}\omega)^2$ in comparison with unity. If r_{pn} is of the order of 10,000 ohms and C_{pn} is less than 150$\mu\mu$f, this condition is satisfied only for frequencies less than about 10,000 cycles per second. The quantity $r_{pn}^2 \frac{C_{pn}}{L_{bn}}$ may be neglected in comparison with unity if r_{pn} is of the order of 10,000 ohms, C_{pn} of the order of 100$\mu\mu$f, and L_{bn} greater than about 1 henry. These conditions indicate that Cases 2, 3, and 4 are restricted to audio-frequency amplifiers. Case 3 applies to an audio-frequency amplifier having a tuned secondary circuit. Case 4

applies to most untuned audio-frequency transformer-coupled amplifiers.

The term $(C_{g(n+1)}\omega/g_{g(n+1)})^2$ becomes $(1/g_{g(n+1)}L_{c(n+1)}\omega)^2$ if the secondary circuit is tuned to the frequency $\omega/2\pi$. Hence, if $g_{g(n+1)}L_{c(n+1)}\omega << 1$, the term $C_{g(n+1)}\omega/g_{g(n+1)}$ must necessarily be large in comparison with unity. Case 5 cannot then apply to a transformer with a tuned secondary circuit. It may apply, however, to a transformer-coupled amplifier with a tuned primary circuit for any frequency and especially for very low audio frequencies.

The term $g^2_{g(n+1)}L_{c(n+1)}/C_{g(n+1)}$, the value of which is a criterion in Cases 6 and 7, becomes $(g_{g(n+1)}L_{c(n+1)}\omega)^2$ if the secondary circuit is tuned to the frequency $\omega/2\pi$. If Cases 6 and 7 are applied to a transformer-coupled amplifier with a tuned secondary circuit, the criteria reduce to $(g_{g(n+1)}L_{c(n+1)}\omega)^2 >> 1$ for Case 6, and $g_{g(n+1)}L_{c(n+1)}\omega << 1$ and $r^2_{pn}C_{pn}/L_{bn} >> 1$ for Case 7. Considering Case 6 under these conditions, the criterion is satisfied if $L_{c(n+1)}\omega >> 1/10g_{g(n+1)}$, and this is very often true at all frequencies, provided $g_{g(n+1)}$ is not very small. Case 6 has wide application, especially since no restriction is placed upon the value of C_{pn}. The reduced criteria given for Case 7 indicate that this case applies especially to transformer-coupled amplifiers at radio frequencies for which $g_{g(n+1)}$ is very small. This condition is met in amplifiers in which the effects of the tube capacitance C_{pg} are reduced or eliminated, as explained in Chap. XVIII.

182. Transformer-coupled Amplifier at Low or Audio Frequencies.—The audio-frequency transformer used as coupling between stages of an amplifier is necessarily an iron-cored transformer as otherwise it is impossible to obtain sufficiently close coupling and sufficiently high impedance in the primary winding.

In the early days of amplifier construction, transformers were built with insufficient impedance in the primary winding. This primary impedance was so low that, at low frequencies of the order of 100 cycles per second, it was small in comparison with the plate resistance of the tube. As a consequence, there was at these low frequencies a comparatively small voltage across the primary winding and a marked reduction in the amplification.

In the last five or ten years, transformer designers have increased this primary impedance very much, thereby improving the low-frequency amplification. This has been done by increas-

ing the amount and improving the quality of the iron in the core, and by increasing the number of turns in the primary winding. The use of such a large number of turns in the primary winding makes it difficult to wind the secondary coil to obtain a high turns ratio. Consequently, the turns ratio of modern transformers ranges from 2 to 6. Furthermore, because of the large number of turns, the distributed capacitance of the windings often places the natural frequency of both the primary and the secondary coils within the range of audio frequencies. The effect of this upon the amplification will now be studied.

The characteristics of a typical audio-frequency transformer are given below.

CHARACTERISTICS OF HIGH-QUALITY AUDIO-FREQUENCY COUPLING TRANSFORMER (AD-1)

Inductance of primary coil (L_{bn})	= 85 henries
D-c. resistance of primary coil ($\bar{R}_{bn}$)	= 1,390 ohms
A-c. resistance of primary coil ($\bar{R}_{bn}$)	= 1,390 ohms (approx.)
Inductance of secondary coil ($L_{c(n+1)}$)	= 765 henries
D-c. resistance of secondary coil ($R_{c(n+1)}$)	= 6,660 ohms
A-c. resistance of secondary coil ($\bar{R}_{c(n+1)}$)	= 6,660 ohms (approx.)
Distributed capacitance of primary coil (C'_{bn})	= 330$\mu\mu$f
Distributed capacitance of secondary coil ($C'_{c(n+1)}$)	= 90$\mu\mu$f
Coefficient of coupling (τ_n)	= 0.998

The method of measuring these characteristics is given in Appendix B.

Because of the iron core, the inductance of the coils is to some extent a function of the degree of saturation of the iron and hence a function of the steady current flowing through the primary winding. Figure 242 gives the inductance of the primary coil of the transformer just described as a function of the steady current flowing in the primary winding.

The inductance of the coils is also a function of the amplitude of the alternating current flowing through the windings, as shown in Fig. 243. The values of inductance given in the tabulation of data above are for an extremely small alternating current.

Using the values of the constants given above for a typical audio transformer, examine the various factors of Eq. (486). Unfortunately, the approximations which hold for low frequencies do not hold for high frequencies. It is best, therefore, to divide the problem into two parts, one for frequencies below

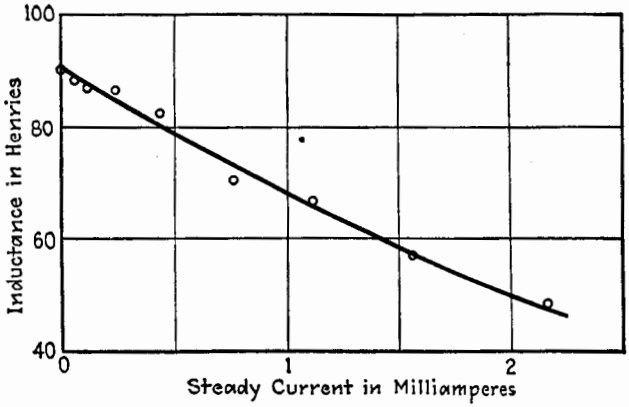


FIG. 242.—Variation of the inductance with steady current of the primary coil of an audio transformer (AD-1).

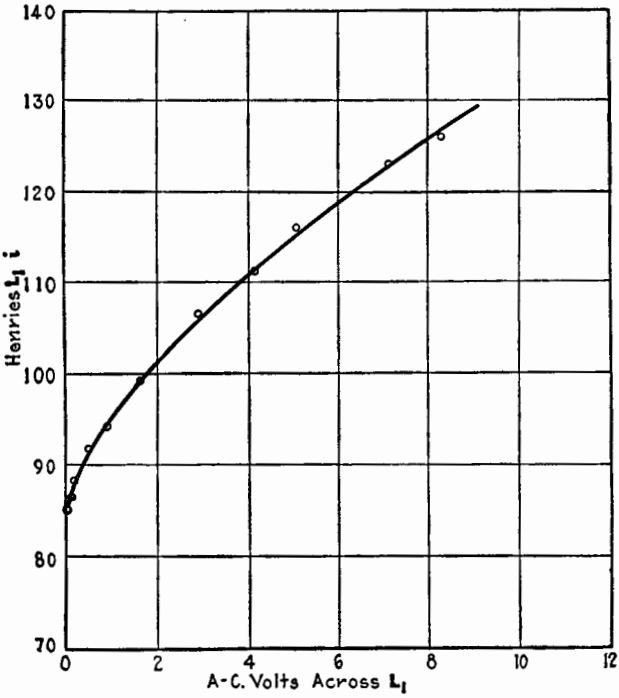


FIG. 243.—Variation of inductance with a-c. voltage across the inductance for the primary coil of an audio-frequency transformer (AD-1).

1,000 cycles per second and the other for frequencies above 1,000 cycles per second.

Frequencies below 1,000 *Cycles per Second.*—Within this range of frequencies, A_n is negligible. The input conductance of the next tube may be neglected; *i.e.*, $g_{g(n+1)}$ may be taken as zero. Then, $\omega_{(n+1)} = \omega_{c(n+1)}$. The factor $(r_{pn}C_{pn}\omega)^2$ is negligible in comparison with unity if r_{pn} is of the order of 10,000 ohms, so that $\eta_n = \dfrac{R_{bn} + r_{pn}}{L_{bn}\omega}$ (see Eq. (487)). Assuming this value of r_{pn}, the ratio $(\omega_n/\omega)^2$ is negligible in comparison with unity, as shown by the fourth equation in Eq. (487). These conditions are the same as the conditions of Case 3. If τ_n is taken as unity, the voltage amplification is

$$(\text{V.A.})_n = \frac{u_{pn}\sqrt{\dfrac{L_{c(n+1)}}{L_{bn}}}\,\dfrac{\omega^2_{c(n+1)}}{\omega^2}}{\sqrt{\eta_n^2+1}\sqrt{\left[\eta_{c(n+1)}+\dfrac{\eta_n}{\eta_n^2+1}\right]^2+\left[1-\dfrac{\omega^2_{c(n+1)}}{\omega^2}-\dfrac{1}{\eta_n^2+1}\right]^2}} \quad (511)$$

Equation (511) is the same as Eq. (494) under Case 3 with τ_n put equal to unity.

At frequencies of about 1,000 cycles per second, η_n^2 and $\eta^2_{c(n+1)}$ are generally very small in comparison with unity, and Eq. (511) reduces to

$$(\text{V.A.})_n = u_{pn}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}} \text{ (for frequencies in the neighborhood of 1,000 cycles per second)} \quad (512)$$

As the frequency is decreased, η_n increases, causing the voltage amplification to decrease; at 10 cycles per second, η_n^2 is of the order of 3 or 4; and the fraction $\omega^2_{c(n+1)}/\omega^2$ is very large. Equation (511) then reduces to the approximate expression

$$(\text{V.A.})_n = \frac{u_{pn}\sqrt{L_{c(n+1)}/L_{bn}}}{\sqrt{\eta_n^2+1}} \text{ (for very low frequencies)} \quad (513)$$

Since Eq. (511) does not involve C_{pn}, the capacitance across the primary coil has no appreciable effect upon the voltage amplification for frequencies below 1,000 cycles per second, if the constants are of the order of magnitude assumed in the example. The capacitance across the secondary coil *does*, however, have an

appreciable effect, as shown by Eq. (511), although at the two extremes of 10 and 1,000 cycles per second this effect is very small.

Frequencies from 1,000 *to* 10,000 *Cycles per Second.*—For the range of frequencies between 1,000 and 10,000 cycles per second, A_n^2 may still be neglected but the other approximations made for frequencies below 1,000 cycles per second do not hold.

At frequencies of the order of 10,000 cycles per second, $g_{g(n+1)}$ may not be negligible, but to simplify the analysis we shall neglect it in the following example. If $g_{g(n+1)}$ is not negligible, its effect is to decrease or increase $\eta_{c(n+1)}$, according as $g_{g(n+1)}$ is negative or positive, and also to make the ratio $\omega_{(n+1)}/\omega$ less than $\omega_{c(n+1)}/\omega$. If r_{pn} is of the order of 10,000 ohms, $(r_{pn}C_{pn}\omega)^2$ is not negligible above about 4,000 cycles per second. The factor η_n is then given by the fourth equation in Eq. (487). The ratio $(\omega_n/\omega)^2$ is negligible in comparison with unity provided r_{pn} is not much greater than 10,000 ohms. The voltage amplification as given by Eq. (486) is

$$(\text{V.A.})_n = \frac{u_{pn}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}\cdot\frac{\omega_{(n+1)}\omega_{c(n+1)}}{\omega^2}}{\sqrt{\eta_n^2+1}\sqrt{\left[\eta_{c(n+1)}+\frac{\eta_n}{\eta_n^2+1}\right]^2+\left[1-\frac{\omega_{(n+1)}^2}{\omega^2}-\frac{\tau_n^2}{\eta_n^2+1}\right]^2}} \tag{514}$$

If r_p is not much greater than 10,000 ohms, and L_{bn} is of the order of 80 henries, η_n^2 is generally negligible in comparison with unity for all frequencies above 1,000 cycles per second. Equation (514) then assumes the simpler form

$$(\text{V.A.})_n = \frac{u_{pn}\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}\cdot\frac{\omega_{(n+1)}\omega_{c(n+1)}}{\omega^2}}{\sqrt{\left[\eta_{c(n+1)}+\eta_n\right]^2+\left[1-\tau_n^2-\frac{\omega_{(n+1)}^2}{\omega^2}\right]^2}} \tag{515}$$

If u_{pn} is unity and r_{pn} is zero, as is true when the voltage ratio of the transformer alone is tested, Eq. (515) gives a maximum $(\text{V.A.})_n$ when $\omega_{c(n+1)}^2/\omega^2 = 1 - \tau_n^2$. The value of the maximum is $\sqrt{\frac{L_{c(n+1)}}{L_{bn}}}\cdot\frac{\omega_{c(n+1)}^2}{\omega^2}\cdot\frac{1}{\eta_{bn}+\eta_{c(n+1)}}$. This maximum is illustrated for a particular transformer in the curves of Fig. 4, page 632, Appendix B. Referring to the figure, if $R_p = 0$, the voltage ratio is 7.7 at a frequency of 10,500 cycles per second.

When r_{pn} is not zero, the maximum is less pronounced. This is shown by the curves of Figs. 2 and 4 of Appendix B and also by curves of $(\text{V.A.})_n$, Fig. 244, taken for an actual amplifier.

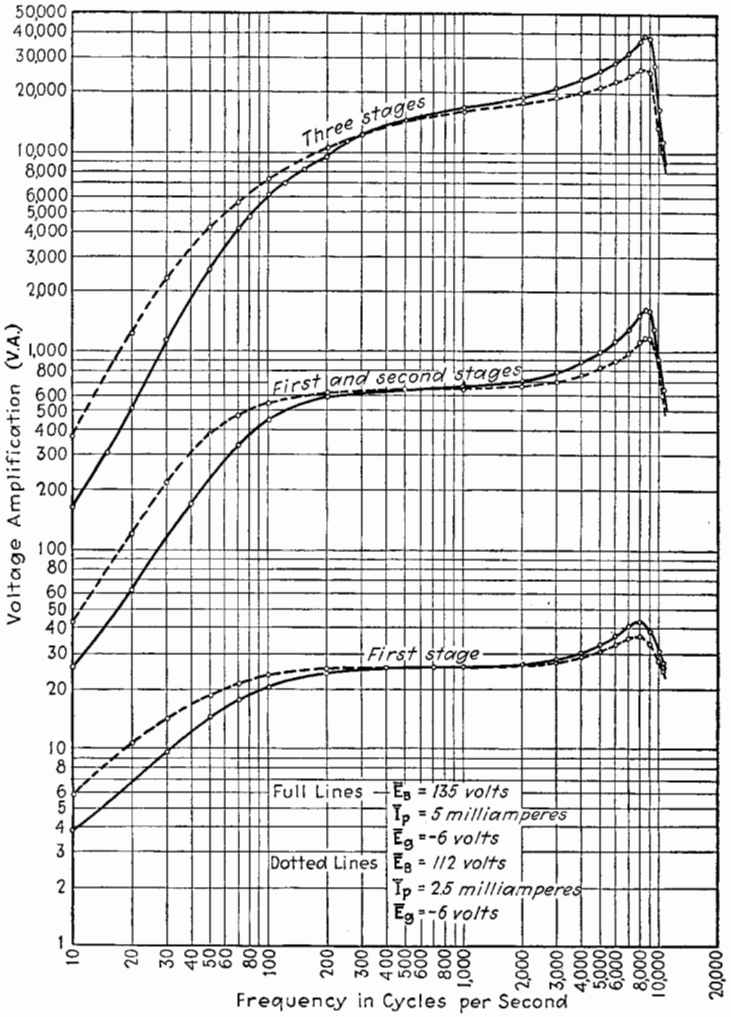


FIG. 244.—Voltage amplification of a three-stage transformer-coupled amplifier. Transformers type TL. UX 201A triode.

183. Experimental Determination of the Performance of a Transformer-coupled Amplifier at Audio Frequencies.—The methods outlined in Chap. XV, Sec. 163, may be used to deter-

mine the curve of voltage amplification plotted against frequency for a transformer-coupled amplifier.

Figure 244 gives typical curves of the way in which the voltage amplification varies with frequency for an audio-frequency transformer-coupled amplifier. The particular amplifier tested comprised three stages with two interstage coupling transformers. These transformers (type TL) were of the same type for which the curves of Fig. 2 of Appendix B were plotted giving the voltage ratio of the transformer alone. The triodes in the first and second stages of the amplifier were type UX 201A, operating at a grid-polarizing potential of -6 volts. The third tube was a type UX 171A, operating at $\bar{E}_B = 110$ volts and $\bar{E}_C = -22.5$ volts. The plate circuit of the last stage included a large cone-type loud-speaker.

The full-line curves of Fig. 244 give the voltage amplification of the first, the first two, and all three stages of the amplifier when the plate-battery voltage of the first two stages was 135 volts. With this value of voltage, the plate current of each of the first two stages was approximately 5 milliamp. and the plate resistance r_p was about 8,000 ohms. The dotted curves of Fig. 244 were taken under the same conditions as for the corresponding full-line curves, except that $\bar{E}_B$ for the first two stages was decreased to 112 volts. Under this condition the plate current of each of the first two stages was about 2.5 milliamp. and r_p was about 11,000 ohms.

Referring to the lower full-line curve of Fig. 244 for the amplification of the first stage with an r_p of 8,000 ohms, the voltage amplification drops off at low frequencies to a much greater extent than in the curve of Fig. 2 of Appendix B for the same transformer with $R_p = 10{,}000$ ohms. This difference is due to the decreased inductance of the coils of the transformer due to the large (5 milliamp.) plate current flowing through the primary winding. That a considerable decrease in inductance resulted, due to saturation of the core, is shown by the curves of Fig. 242. The lower dotted curve of Fig. 244 shows a smaller dropping off of amplification at low frequencies because of the smaller plate current (2.5 milliamp.). Therefore, the magnitude of the plate current is an important consideration in transformer-coupled audio-frequency amplifiers.

184. Transformer-coupled Amplifier at High or Radio Frequencies.—This section deals with the theory of transformer-

coupled amplifiers for radio frequencies. There are three types of transformer used as coupling element at these frequencies. They are

1. Iron-cored untuned transformer.
2. Air-cored untuned transformer.
3. Air-cored tuned-secondary-circuit transformer.

1. *Iron-cored Untuned Transformer.*—The iron-cored untuned transformer is practically obsolete and of little concern. It has been used with fair satisfaction for the lower radio frequencies when the amplifier is designed to operate without tuning over a considerable range of frequencies. Although such a transformer-coupled amplifier is termed an "untuned" amplifier, resonance actually does play an important part. Because of the iron losses and the large resistance of the coils, the resonance peak is very dull, but has its maximum within the range of frequencies for which the amplifier is designed. The coupling between the primary and secondary winding is generally nearly unity.

In order to give an idea of the order of magnitude of the constants of a transformer of this type, values are given for a particular transformer used as an interstage transformer in an early amplifier.

CONSTANTS OF IRON-CORED TRANSFORMER FOR RADIO FREQUENCIES

$L_1 = 1.5$ millihenries
$L_2 = 28.5$ millihenries
$\tau = 0.98$
$R_1 = 33$ ohms
$R_2 = 148$ ohms
Frequency range, 75–150 kc.

Equation (486) should be consulted in calculating the performance of an amplifier with these constants.

2. *Air-cored Untuned Transformer.*—This type of transformer is so similar in electrical characteristics to the iron-cored type as to require no special theoretical treatment. The coils are generally wound with very fine wire. In order to obtain the close coupling required to give a broad resonance curve, the coils are wound so that they are physically very close. The wires of the two coils may even be wound together so that the two wires run parallel to each other in the same bobbin.

3. *Air-cored Tuned Transformer.*—The approximate theory of a single-stage amplifier using an air-cored tuned transformer was presented in Sec. 124, Chap. XII. This approximate theory neglected some factors which, although generally small, may be too large to be disregarded. This section supplements the approximate theory by including all effects. In order to emphasize the importance of this section, it may be mentioned that practically all radio-frequency amplifiers are of the type now to be considered. Usually, there is means to decrease the feed-back effect through the grid-to-plate capacitance of the triode. This feed-back may be decreased by neutralizing the amplifier by methods to be described in the next chapter.

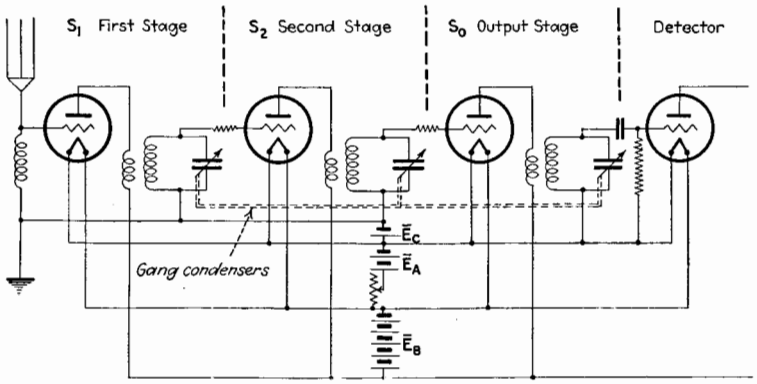


FIG. 245.—Multistage tuned radio-frequency amplifier feeding a detector.

The neutralizing means is either obviously incorporated in the structure of the amplifier or present in the less obvious form of natural capacitances between parts of the circuits, due to their relative positions. The effect of this reduction in feed-back is to decrease the value of $g_{g(n+1)}$ and of $b_{g(n+1)}$. If feed-back is practically eliminated, the input conductance may be assumed to be zero and the input susceptance may be assumed to have some constant value dependent upon the means for eliminating feed-back.

Figure 245 is a diagram of connections including the essential elements and is typical of a multistage radio-frequency amplifier. The grid of the first stage is connected across an inductance in an untuned antenna. There are, however, many different methods of connection to the antenna which may be

tuned or untuned. These methods will not be described because they do not influence materially the performance of the amplifier. The third transformer is shown connected to the grid of a detector. The detector triode may be of a different type from the amplifier triodes. Even if of the same type, the input admittance of the detector tube is generally different from the input admittance of the other triodes. Consequently, the third or output stage of the amplifier has somewhat different characteristics from those of the first and second stages.

The tuning of the several stages is generally accomplished by condensers, the rotating parts of which are either mounted on the same shaft or else belted or geared. Such a system of condensers is known as a "gang condenser."

The design of the transformer varies to some extent according to the frequency range for which the amplifier is constructed and also according to the type of vacuum tube used. The physical dimensions and approximate electrical constants of a particular type of transformer follow. These data are given merely for illustration and are not recommended necessarily as the best design. The primary coil is wound directly over the low-potential end of the secondary winding and is insulated from it by a strip of paper.

DIMENSIONS AND CONSTANTS OF RADIO-FREQUENCY TRANSFORMER

Secondary coil:	
Diameter	1.25 in.
Length	1.87 in.
Number turns	137
Size wire	No. 32 (B. & S. gauge)
Inductance	420μh
D-c. resistance	8.0 ohms
η	0.01 (approx.)
Primary coil:	
Diameter	1.3 in.
Length	0.11 in.
Number turns	16
Size wire	No. 38 (B. & S. gauge)
Inductance	19.2μh
D-c. resistance	3.5 ohms
Mutual inductance	38.7μh
Coefficient of coupling	0.431
Capacitance of tuning condenser	Max. 181$\mu\mu$f
	Min. 11.2$\mu\mu$f
Frequency range	Max. 1407 kc.
	Min. 553 kc.

Using the data given, the theory developed under Case 7 applies, provided $g_{g(n+1)}$ is in absolute value not greater than about 10^{-5} mho. The analysis of Case 7 applies to each stage as a unit, and, if the feed-back is eliminated, each stage operates independently of the other stages. When, however, feed-back exists, the performance of one stage is greatly affected by the next stage, because the value of $g_{g(n+1)}$ changes according to the nature of the plate load of the stage $(n + 1)$. If the capacitance of the tuning condenser of stage $(n + 1)$ is smaller than that for resonance to the incoming signal, the plate load is inductive and the input conductance $g_{g(n+1)}$ may be negative. This negative input conductance affects the amplification of stage n and may be so negative as to cause the nth stage to oscillate. If the capacitance of the tuning condenser in stage $(n + 1)$ is larger than the resonant value, $g_{g(n+1)}$ is positive and the amplification of the nth stage is decreased. It is clear that the performance of an amplifier in which feed-back exists varies greatly according to the settings of the several tuning condensers. Because of the interaction between stages, such an amplifier is often unstable. Feed-back is greater at the high frequencies. Partially to stabilize an amplifier in which this disturbing feed-back exists, resistances of a few hundred ohms are often connected in the wires leading to the grids of the triodes. These resistances introduce losses which are greater at high frequencies.

In a later chapter screen-grid triodes will be considered. These tubes eliminate, to a large extent, the feed-back through the capacitance C_{pg}. But the value of r_p for these tubes is very high, usually of the order of several hundred thousand ohms. The theory of Case 7 still applies but the constants have different relative magnitudes.

General References

DICKEY: Notes on the Testing of Audio Frequency Amplifiers, *Proc. I.R.E.*, **15**, 687 (1927).

DIAMOND and WEBB: Testing of Audio-frequency Transformer-coupled Amplifiers, *Proc. I.R.E.*, **15**, 767 (1927).

FELDTKELLER and STRECKER: Theorie der Niederfrequenz Verstärkerketten, *Arch. Elektrot*, **24**, 425 (1930).

LANG: Die Berechnung der Verstärker Transformatoren, *Elektrotech. Zeits.*, **51**, 1637 (1930).

FORSTMANN: Über die Bemessung verzerrungsfreier Niederfrequenzverstärker mit Transformatorenkopplung, *Elektrotech. Zeits.*, **52,** 599, 660 (1931).

WHEELER and MCDONALD: Theory and Operation of Tuned Radio-frequency Coupling Systems, *Proc. I.R.E.*, **19,** 738 (1931).

SCHOR: An Untuned Radio-frequency Amplifier, *Proc. I.R.E.*, **20,** 87 (1932).

CHAPTER XVIII

METHODS OF REDUCING ENERGY INTERCHANGE BETWEEN GRID AND PLATE CIRCUITS OF A TRIODE. "NEUTRALIZATION"

The ideal repeater or amplifier is one which is perfectly unidirectional. A unidirectional amplifier is one in which there is no energy flow in either direction between the output and input circuits, and the only action is the control of power in the output circuit by the grid voltage.

185. Effects of Energy Interchange between the Grid and Plate Circuits of a Triode.—We may summarize briefly the deleterious effects of energy interchange before considering methods of reducing or eliminating these effects. When there exists any electrical path between the plate circuit and the grid circuit of a triode, energy may flow *from* the grid circuit to the plate circuit, or *into* the grid circuit from the plate circuit. If energy flows *from* the grid circuit, the effective resistance of the grid circuit is increased and the potential impressed upon the grid is less than it would be if this energy flow did not occur. Reduced amplification results. If energy flows *into* the grid circuit, regeneration occurs, and the amplifier may oscillate or be unstable. "Feed-back" is said to take place, although broadly speaking, feed-back may be either positive or negative.

The effects upon amplification of the interchange of energy through the path comprising the mutual capacitance C_{pg} between plate and grid were studied in Chap. XVII. This interchange of energy was there expressed in terms of input admittance $\mathbf{y}_g = g_g - jb_g$. If g_g is positive, energy is taken from the grid circuit by the plate circuit and the resulting amplification is reduced. This was found to be a very large effect at radio frequencies, because the path through C_{pg} is then of comparatively low impedance. The input conductance g_g is always positive for resistance-coupled amplifiers but may be negative when the plate circuit is inductive. When g_g is negative, regen-

eration and often oscillation take place. The amplifier is then unstable and is useless.

Not only is the real component g_g of the input admittance a disturbing factor in the design of stable amplifiers, but the component b_g, also, has its harmful effects. This susceptance component b_g is always equivalent to a shunt capacitance between grid and filament. Since b_g depends largely upon frequency, plate resistance, and other constants of the circuits, this equivalent shunt capacitance is variable and dependent upon the conditions of operation. The existence of both components of the input admittance, if appreciable, makes it difficult to design a single-control, multistage, tuned radio-frequency amplifier, unless some means is applied to eliminate, partially at least, the reaction of the plate circuit upon the grid circuit.

186. Causes of Energy Interchange between the Plate and Grid Circuits of a Triode.—The electrical path through C_{pg}, though usually the most active, is not the only path by which energy may flow between the grid and plate circuits. In fact, there are several causes for this interchange of energy, the most important of which are listed below.

Causes of Energy Interchange between Grid and Plate Circuits

1. Incidental coupling between such portions of the plate circuit and grid circuit as are external to the tube.
2. The flow of the a-c. component of the plate current through the filament of the tube.
3. Reaction of the plate voltage upon the thermionic grid current, expressible in terms of s_g.
4. Ionization or secondary emission which makes the flow of current possible directly from grid to plate or *vice versa*.
5. Coupling between output and input circuits due to capacitance between plate and grid electrodes, including the socket and connections thereto.
6. Coupling between output and input circuits due to reaction of the fluctuating space charge upon the grid.

We shall now consider these causes of energy interchange and means of reducing them..

187. Reduction of Cause 1.—The coupling between external portions of the plate and grid circuits may be electrostatic coupling between wires, between condensers, or between coils in the two circuits; magnetic coupling between coils; or conductive coupling if any portion of the two circuits is common to both.

The magnetic coupling is reduced by placing the grid and plate coils so that their mutual inductance is as near zero as possible. The electrostatic coupling can be sufficiently reduced only by the most thorough electrostatic screening. It is now common practice to enclose the tube and its output circuit, together with the input circuit of the next stage, in a common metal enclosure or in separate metal enclosures made of wire screening or of sheet metal carefully bonded and grounded. Even a screw hole or a crack in the metal screen is sometimes sufficient to allow a disturbing amount of energy to leak through.

Coupling between different stages often exists because of common plate or grid batteries. This conductive coupling can be reduced by shunting the batteries or voltage sources with large condensers. It is very important that there be no wire, no matter how short, which is common to an output and input circuit, unless it is bridged by a large condenser.

The types of coupling described are harmful when one or more of them exists, not only between the output and input circuits of a single stage but also between any output and any input circuit of a multistage amplifier. When the amplification of the several stages is large, an inconceivably small coupling between the final output and initial input circuits may cause the amplifier to oscillate. Great care is necessary to reduce or to eliminate all coupling of the kinds enumerated in this section.

188. Reduction of Cause 2.—The plate circuit is often connected to one end of the filament. In consequence, a fraction of the alternating component of the plate current flows through the filament and, owing to the filament resistance, produces a varying voltage across the filament. To a slight degree, this varying filament potential causes a varying potential between the grid and some portion of the filament. In this way the plate circuit can cause a varying potential in the grid circuit. Although this effect is small unless the filament resistance is large, it is usually desirable in multistage amplifiers to reduce this reaction by shunting the filament with a condenser having a capacitive reactance which is small in comparison with the resistance of the filament. This effect is absent when the triodes have separately heated constant-potential cathodes.

189. Reduction of Cause 3.—The third cause of reaction of the plate potential upon the grid circuit can occur only when

there is a flow of grid conduction current. The reaction is usually in the direction to cause absorption of energy from the grid circuit and hence to cause a reduction of amplification. Energy interchange due to this cause seldom exists in practice, because high-vacuum triodes whose grids are negatively polarized are generally used. Under these conditions no grid conduction current flows and hence s_g is zero.

190. Reduction of Cause 4.—Interchange of energy between the two circuits due to the fourth cause can occur only when electrons or ions can travel from one of the cold electrodes to the other, or when energy changes can take place in the space between electrodes, as by excitation or ionization of a gas. Hence, this cause is eliminated by the use of high-vacuum tubes operated at potentials which allow no secondary-emission effects to occur. In general, the grid is the most negative electrode so that no electrons can pass to the grid and hence no secondary electrons can be emitted therefrom.

191. Reduction of Causes 5 and 6.—Causes 5 and 6 can be considered together. Although the character of the coupling between the plate and grid circuits differs somewhat for the two cases, the effects of the two couplings are the same. The coupling under Cause 5 is through the capacitance between the grid and its connections and the plate and its connections. This capacitance is that which is measured when the filament is not heated. When, however, the tube is in operation, the alternating component of the plate current causes a fluctuation in the space charge which is most dense near the filament. This fluctuating space charge induces a fluctuating charge on the grid. The varying plate potential, therefore, can act upon the charge on the grid not only by the simple electrostatic effect of the plate but also through the varying space charge in the grid-to-filament space. The total effect of the two couplings can be expressed in terms of an equivalent C_{pg} which is a little larger than the pure capacitance between grid and plate. This equivalent C_{pg} must be measured when the tube is lighted and in operating condition. Benham[1] has shown that this equivalent value of C_{pg} may be quite different from the value measured when the filament is cold.

The causes of interchange of energy numbered 1, 2, 3, and 4 can be reduced or eliminated by proper adjustment of the tubes,

[1] BENHAM, *Phil. Mag.*, **5**, 630, 1928; **11**, 457 (1931).

by suitable disposition of the inductances and capacitances, and by the use of electrostatic screens. The coupling through C_{pg} is always present unless the electrostatic screening is extended to the internal structure of the tube. Such a scheme of internal screening will be considered in Chap. XXIII. We shall now study at some length methods of reducing the effects of the coupling through the equivalent capacitance C_{pg}.

The methods of reducing the interchange of energy between the plate and grid circuits through the capacitance C_{pg} can be classified in two fundamental groups, as follows:

1. *Balancing the Feed-back Voltage.*
2. *Balancing the Feed-back Current.*

The method of Group 1 consists in introducing into the grid circuit a voltage equal and opposite in phase to the voltage produced by the flow of the feed-back current through the total grid-circuit impedance.

The method of Group 2 consists in leading to the grid a current equal and opposite in phase to the feed-back current that flows to the grid through C_{pg}, so that the net current which flows through the grid-circuit impedance, due to the action of the plate voltage, is zero. Consequently, no feed-back voltage is produced in the grid circuit.

Probably the earliest method of reducing feed-back is one described by Hartley[2] in 1915 and is an example of Group 1. About two years later, Rice[3] invented a more practical scheme for neutralizing amplifiers. Hazeltine[4] has taken out several patents on various neutralizing circuits which fall in both of the groups given. A very good discussion of the various neutralizing schemes was given by Ballantine[5] in a patent filed in 1923.

192. Group 1. Balancing the Feed-back Voltage Due to C_{pg}. There are several ways of reducing the effect of C_{pg} which are classified under this method. Perhaps the most obvious way is that patented by Hartley and illustrated in Fig. 246. It was shown in Chap. XI that, if the plate load of a triode is inductive, the input admittance may have a negative real component which may cause the input circuit to oscillate. In Fig. 246, the induc-

[2] HARTLEY, U. S. patent 1,183,875 filed September, 1915.

[3] RICE, U. S. patent 1,334,118 filed July, 1917.

[4] HAZELTINE, U. S. patent 1,450,080 filed August, 1919; U. S. patent 1,489,228 filed Dec., 1920; U. S. patent 1,533,858 filed December, 1923.

[5] BALLANTINE, U. S. patent 629,702 filed April, 1923.

tive plate load L_b may cause regeneration or even oscillation in the circuit L_1C_1. If the plate current is passed through a small inductance L'_b, which is coupled to L_1 by mutual inductance M, the coupling being in the opposite direction to that which causes regeneration, the regenerative effect of C_{pg} can be reduced. In this scheme a voltage derived from the plate circuit is introduced into the grid circuit to balance the voltage due to the flow of current through C_{pg}.

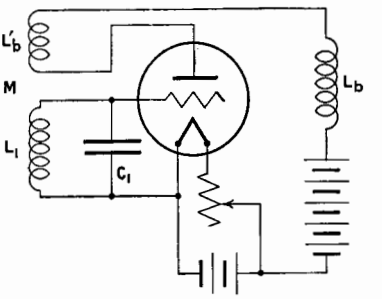


FIG. 246.—Hartley's method of reducing feed-back.

The Hartley method is seldom used because its action is dependent upon frequency. As the frequency is changed by varying C_1, M must also be changed. The most generally used schemes are those which balance the effects of C_{pg} independently of frequency.

Figure 247*a* indicates a way of utilizing the Wheatstone bridge principle to prevent the plate voltage from affecting the grid voltage. The circles P, G, and F represent the plate, grid, and filament terminals of a triode. Between P and F are shown the internal effective elements of a triode, r_p and $u_p\Delta E_g$. The plate load of the triode is Z_1 and Z_2 in series, these two impedances forming also two arms of the bridge. C_{pg}, the grid-to-plate capacitance whose effects it is desired to eliminate, forms the third arm of the bridge. Z_c forms the fourth arm of the bridge. The input voltage is applied across one diagonal of the bridge, while the plate voltage ΔE_p acts across the other diagonal. When the bridge is balanced by a proper choice of the impedances of the arms, ΔE_p can neither increase nor decrease the input voltage ΔE_0, and hence the input circuit, whatever it may be, is electrically isolated from the output circuit. This is true whether the cathode of the triode is heated or cold, *i.e.*, whether r_p is finite or infinite. Because of the conjugate property of a bridge, ΔE_0 has no effect upon ΔE_p when the cathode is cold. When the cathode is heated, however, and r_p is finite, ΔE_p is produced by ΔE_0, not through the bridge network but because of the unilateral amplifying property of the triode which is expressed by the e.m.f. $u_p\Delta E_g$.

The condition of balance of the bridge of Fig. 247*a* is

$$\frac{\mathbf{Z}_1}{\mathbf{Z}_2} = \frac{-j/C_{pg}\omega}{\mathbf{Z}_c} \tag{516}$$

If this condition of balance is to be independent of frequency, either Z_c or Z_1 must be a pure capacitance. If there is some conductance between the grid and plate across the glass seals, C_{pg} is a leaky condenser, and Z_c or Z_1 must likewise be a leaky con-

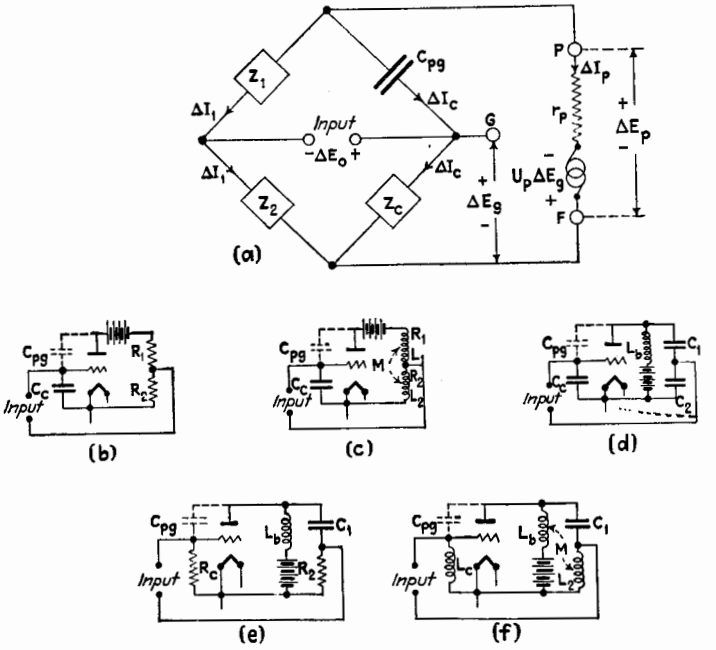


FIG. 247.—Neutralization by voltage balance.

denser. If Z_c is a capacitance C_c, Z_1 and Z_2 may be two resistances, two inductances with or without mutual inductance, or two capacitances. Figures 247*b*, *c*, and *d* show the connections for these three possibilities. The conditions of balance are

$$\left.\begin{aligned} \frac{R_1}{R_2} &= \frac{C_c}{C_{pg}} && \text{(Fig. 247}b\text{)} \\ \frac{\pm M + L_1}{\pm M + L_2} &= \frac{C_c}{C_{pg}} && \text{(Fig. 247}c\text{)} \\ \frac{C_2}{C_1} &= \frac{C_c}{C_{pg}} && \text{(Fig. 247}d\text{)} \end{aligned}\right\} \tag{517}$$

Since C_{pg} is usually small, of the order of from 5 to 10$\mu\mu$f, R_2, L_2, and C_1 may be small for the conditions of balance.

If we make Z_1 a pure or leaky capacitance according as C_{pg} is pure or leaky, two new possible connections arise, as shown in Figs. 247*e* and *f*. The conditions of balance for these two schemes are

$$\left.\begin{aligned} \frac{R_c}{R_2} &= \frac{C_1}{C_{pg}} \qquad \text{(Fig. 247}e\text{)} \\ \frac{\pm M + L_c}{\pm M + L_2} &= \frac{C_1}{C_{pg}} \qquad \text{(Fig. 247}f\text{)} \end{aligned}\right\} \tag{518}$$

Since C_{pg} is small, R_2 and L_2 should be small.

In Figs. 247*d*, *e* and *f*, the impedance Z_1 consists of a capacitance, which makes it necessary to provide a path for the d-c. component of the plate current. This is done by an inductance L_b, as shown in the figures. L_b does not in any way disturb the balance and it may have any constants or indeed be a pure resistance. In fact, L_b may be the main plate load, such as the primary coil of a transformer. This shunt plate load might to advantage be added to Figs. 247*b* and *c* so that the plate battery could be connected in series with L_b instead of being in the positions shown. This arrangement would permit the plate battery to be connected directly to the filament, a more practical position, for then the same plate battery could be used to supply other tubes. Similarly, for the scheme in Fig. 247*d*, since there is shown no conductive path from filament to grid to maintain the proper polarizing potential of the grid, a high resistance must shunt C_c and, for perfect balance, C_2 also.

All of the schemes shown have a serious disadvantage from a practical standpoint. This disadvantage results from the fact that the input circuit is connected from the grid to some point in the plate-circuit network and *not* to the filament. The effective grid voltage ΔE_g is, therefore, not the same as the input voltage, as shown by the following relation which applies to Fig. 247*a*.

$$\mathbf{\Delta E}_g = \frac{\mathbf{Z}_c}{\mathbf{Z}_c + \mathbf{Z}_2}\mathbf{\Delta E}_0 \tag{519}$$

Using the condition of balance given by Eq. (516), Eq. (519) can be written

$$\mathbf{\Delta E}_g = \frac{-j/C_{pg}\omega}{\mathbf{Z}_1 - \frac{j}{C_{pg}\omega}}\mathbf{\Delta E}_0 \tag{520}$$

The effective part of ΔE_0, namely ΔE_g, varies with frequency except for the schemes in Figs. 247*d*, *e* and *f*. For these cases $\Delta E_g = \frac{C_1}{C_1 + C_{pg}}\Delta E_0$, and C_1 should be at least as large as C_{pg} and preferably larger. In the scheme shown in Fig. 247*c*, the effects of R_1 and R_2, the resistances of L_1 and L_2, have been neglected. In calculating by Eq. (520) the fraction of the input voltage which acts upon the grid, the resistances R_1 and R_2 must be considered. For Fig. 247*c*

$$\mathbf{\Delta E}_g = \frac{-j/C_{pg}\omega}{R_1 + j\left(L_1\omega - \frac{1}{C_{pg}\omega}\right)}\mathbf{\Delta E}_0 \tag{521}$$

If in Eq. (521), $L_1\omega - \frac{1}{C_{pg}\omega} = 0$, then $\Delta E_g = \frac{\Delta E_0}{R_1 C_{pg}\omega}$, so that for some frequencies ΔE_g may be much greater than ΔE_0.

It is of considerable practical importance to examine the effect that the balancing schemes just discussed have upon the input admittance of the tube and its system, as determined at the input terminals. For those connections which are fundamentally represented by the bridge diagram, Fig. 247*a*, the input admittance is

$$\mathbf{Y}_g = \frac{1}{\mathbf{Z}_1 - \frac{j}{C_{pg}\omega}} + \frac{1}{\mathbf{Z}_2 + \mathbf{Z}_c} \tag{522}$$

Combining with this equation the condition of balance of Eq. (516), the input admittance becomes

$$\mathbf{Y}_g = \frac{\mathbf{Z}_c - \frac{j}{C_{pg}\omega}}{\mathbf{Z}_c + \mathbf{Z}_2} \cdot jC_{pg}\omega = \frac{1}{\mathbf{Z}_1} \cdot \frac{\mathbf{Z}_2 + \mathbf{Z}_1}{\mathbf{Z}_2 + \mathbf{Z}_c} \tag{523}$$

The input admittance $\mathbf{Y}_g$ is not dependent upon the plate resistance and amplification factor of the tube, as in the case of an unbalanced triode, but is dependent upon frequency and the constants of the balancing circuits. In determining the magnitudes of the impedances Z_1, Z_2, and Z_c, Eqs. (516), (519), and

(523) should be considered. Usually, Z_c should be of the same order of magnitude as $1/C_{pg}\omega$, while Z_1 and Z_2 are from 0.01 to 0.1 of $1/C_{pg}\omega$. This choice of magnitudes gives the ratio $\Delta E_g/\Delta E_0$ a value of at least 0.9, and the input admittance is not much more than twice the value of $jC_{pg}\omega$.

Before leaving this method of balancing the feed-back voltage, two schemes will be described which make use of a mutual inductance. The bridge of Fig. 247*a* may be extended to the form shown in Fig. 248*a*. The actual connections for this scheme

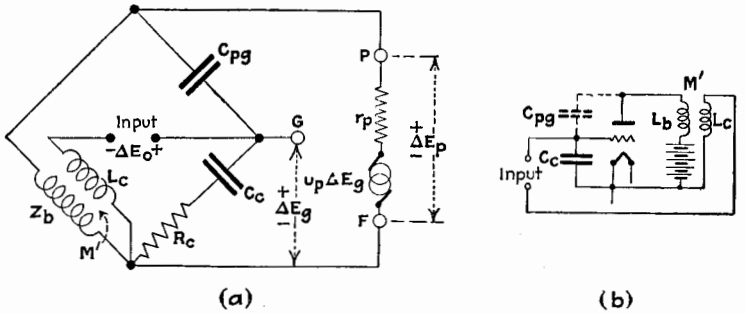


FIG. 248.—Neutralization by voltage balance by means of a mutual inductance.

are shown in Fig. 248*b*. The conditions of balance for the diagram of Fig. 248*a* are

$$\left.\begin{aligned} \frac{R_b R_c C_c + L_b - M'}{M'} &= \frac{C_c}{C_{pg}} \\ R_c(L_b - M') &= \frac{R_b}{C_c} \end{aligned}\right\} \tag{524}$$

If the resistances are negligible, the balance conditions reduce to

$$\frac{L_b - M'}{M'} = \frac{C_c}{C_{pg}} \tag{525}$$

The method given in Fig. 248*a* has an input circuit which does not connect directly from grid to filament but has an impedance L_c in series. Hence, the total input voltage is not effective as grid voltage and the method has the same disadvantage as the methods already discussed. The input admittance does not depend upon r_p and u_p but does depend upon the constants of the bridge elements.

A second method of balancing the feed-back voltage, utilizing a mutual inductance, is shown diagrammatically in Fig. 249 and, as used in practice, in Fig. 250. The condition of balance is expressed by

$$M'^2\omega^2 - \frac{M'}{C_{pg}} + \mathbf{Z}_c\left(\mathbf{Z}'' - \frac{j}{C'\omega}\right) = 0 \tag{526}$$

This equation may be separated into two equations by equating to zero the sum of the real terms and the sum of the imaginary terms. The two conditions of balance for the scheme of Fig. 249 are

$$\left.\begin{aligned}(M'^2 - L_cL'')\omega^2 - \left(\frac{M'}{C_{pg}} - \frac{L_c}{C'}\right) + R_cR'' &= 0\\ R_c\left(L''\omega - \frac{1}{C'\omega}\right) + R''L_c\omega &= 0\end{aligned}\right\} \tag{527}$$

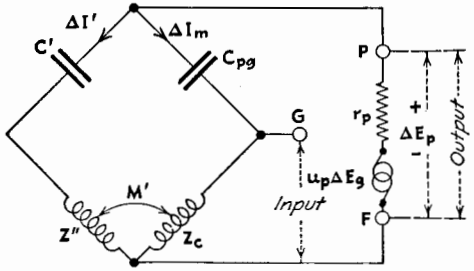


FIG. 249.—Circuit for neutralization by voltage balance

The second equation cannot be satisfied independent of frequency. If, however, the resistance terms, which are small in comparison with the reactance terms, are neglected, the approximate condition of balance is

$$(M'^2 - L_cL'')\omega^2 - \left(\frac{M'}{C_{pg}} - \frac{L_c}{C'}\right) = 0 \tag{528}$$

If L_c and L'' are wound to have unity coupling, $M'^2 = L_cL''$ and the balance is independent of frequency. The condition of approximate balance becomes

$$\frac{C'}{C_{pg}} = \frac{L_c}{M'} \text{ or } \frac{C'}{C_{pg}} = \frac{N_c}{N''} \tag{529}$$

where N_c and N'' are the numbers of turns in L_c and L''.

This last method has advantages over the others previously described in that the input now connects directly from grid to filament and the whole input voltage is effective in acting upon the grid. Furthermore, the conditions of balance are independent of any circuit connected across the output and also of any circuit connected across the input if the circuit has no mutual action upon either L'' or L_c. In Fig. 250, a tuning condenser C_c is shown connected across the input coil Z_c. This method was described by Hazeltine[6] in 1919.

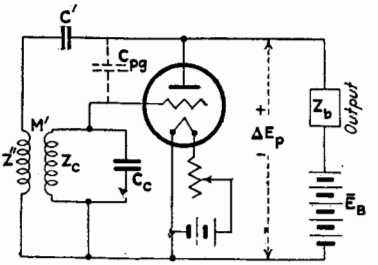


FIG. 250.—Hazeltine's method of neutralization by voltage balance.

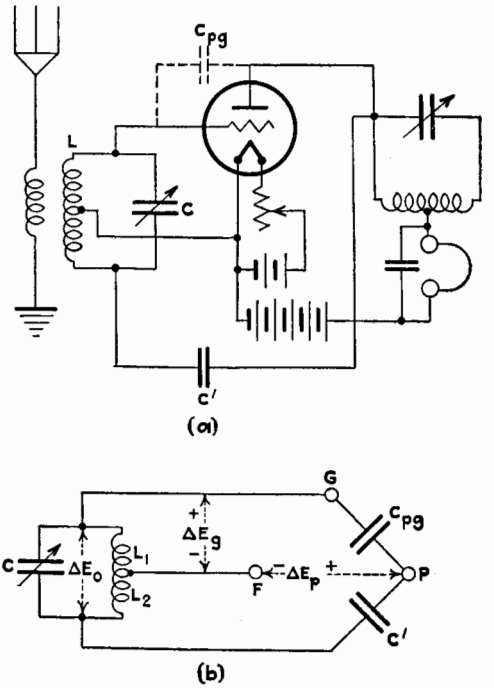


FIG. 251.—Rice's method of neutralization.

Finally, a method of neutralization closely allied to the method just described was proposed by Rice in 1917 before any of the

[6] HAZELTINE, U. S. patent 1,450,080 filed Aug. 7, 1919.

other methods of neutralization, except that of Hartley, were described. Rice's method is shown in Fig. 251*a* which is taken from Fig. 1 of the Rice patent. The symbols are changed and certain shields and dotted lines indicating natural capacitances are omitted. The Rice circuit, when redrawn to show the elements which are important to the type of neutralization now being considered, is shown in Fig. 251*b*. This figure shows that, because of symmetry, the plate voltage ΔE_p produces no current through C and hence no voltage across L, if

$$\frac{C_{pg}}{C'} = \frac{L_2 + M}{L_1 + M} \tag{530}$$

In Eq. (530), M is the mutual inductance between L_1 and L_2, the two portions of L.

Although ΔE_p produces no voltage across L when Eq. (530) holds, this does not mean that ΔE_p produces no grid voltage. True neutralization is, therefore, not secured. The value of ΔE_g, when the adjustments are in accordance with Eq. (530), is

$$\Delta E_g = \frac{\left(L_1 - M\frac{C'}{C_{pg}}\right)\omega}{\left(L_1 - M\frac{C'}{C_{pg}}\right)\omega - \frac{1}{C_{pg}\omega}}\Delta E_p \tag{531}$$

or

$$\Delta E_g = \frac{(L_1L_2 - M^2)\omega^2}{\frac{L_2 + M}{C_{pg}} - (L_1L_2 - M^2)\omega^2}\Delta E_p \tag{532}$$

The first term in the denominator of Eq. (532) is usually very large in comparison with the second term, so that, although theoretically the equation indicates a lack of absolute balance and a variation of the degree of balance with frequency, practically the degree of balance can be made quite satisfactory over a considerable range of frequencies.

One peculiarity of the Rice method is the tendency of the system to oscillate at a very high frequency compared to the frequency of the circuit LC. This oscillation can be prevented without affecting the operation of the system by including a high resistance in the wire from the coil L to the filament.

193. Group 2. Balancing the Feed-back Current.—The method to be described in this section for reducing the interchange

of energy between plate and grid circuits includes the most important and the most used schemes of eliminating feed-back.

The broad principle of operation may be explained by reference to Fig. 252, which shows a triode with plate load Z_b and grid-circuit impedance Z_c. The capacitance which causes feed-back is represented in dotted lines as a condenser C_{pg}. Because of the alternating component ΔE_p of the plate voltage, a current ΔI_m flows through the coupling path C_{pg} and would by itself cause a variation of grid potential due to the impedance voltage $\mathbf{Z}_c\Delta\mathbf{I}_m$. If a connection, usually through a small condenser C', is made from the grid to a point in the system having the proper potential to cause a current $\Delta I'$ to flow to the grid, such that

$$\Delta\mathbf{I}' = -\Delta\mathbf{I}_m \tag{533}$$

then the net current through Z_c due to ΔE_p is zero and the plate voltage ΔE_p causes no change in grid voltage. The problem is to connect a wire a to a suitable point in the plate circuit or to a suitable point in a circuit which is coupled to the plate circuit.

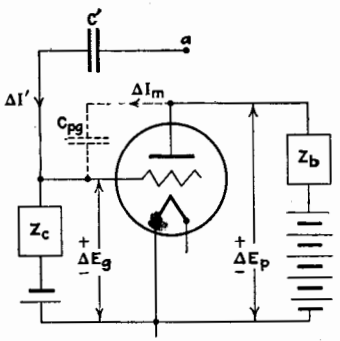


FIG. 252.—Neutralization by current balance.

Some of the circuits which neutralize the feed-back current by an equal and opposite balancing current are variations on the general bridge circuit shown in Fig. 253. In this diagram the input circuit connects directly from grid to cathode.

The equations expressing the condition that ΔE_g is zero for all values of ΔE_p are

$$\left.\begin{aligned} \frac{j\Delta\mathbf{I}_m}{C_{pg}\omega} - \mathbf{Z}_b\Delta\mathbf{I}_b + jM'\omega\Delta\mathbf{I}' &= 0 \\ -jM'\omega\Delta\mathbf{I}_b + (\mathbf{Z}' + \mathbf{Z}'')\Delta\mathbf{I}' &= 0 \\ \Delta\mathbf{I}' = -\Delta\mathbf{I}_m & \end{aligned}\right\} \tag{534}$$

The conditions of balance derived from Eq. (534) are contained in

$$M'^2\omega^2 - \frac{M'}{C_{pg}} + \mathbf{Z}_b(\mathbf{Z}' + \mathbf{Z}'') = 0 \tag{535}$$

Equation (535) may be separated into two equations by equating

to zero the sum of the real terms and the sum of the imaginary terms. The two conditions of balance are

$$\left.\begin{aligned} M'^2\omega^2 - \frac{M'}{C_{pg}} - X_b(X' + X'') + R_b(R' + R'') &= 0 \\ (R' + R'')X_b + R_o(X' + X'') &= 0 \end{aligned}\right\} \quad (536)$$

The second equation can be satisfied only by making one of the reactances negative, in which case the equation is true for only

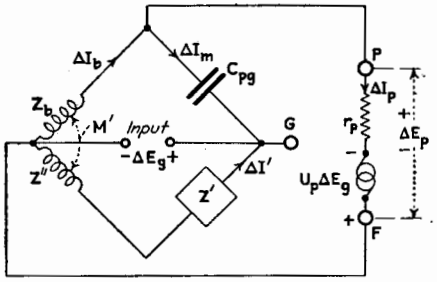


FIG. 253.—General circuit for neutralization by current balance.

one frequency. It is evident, therefore, that any arrangement of circuits corresponding to Fig. 253 cannot be perfectly balanced independent of frequency. Usually, however, the resistances of the circuits are small in comparison with the reactances and can in general be neglected. The condition of balance then reduces to

$$M'^2\omega^2 - \frac{M'}{C_{pg}} - X_b(X' + X'') = 0 \quad (537)$$

One way of satisfying Eq. (537) is to make X_b and X'' inductive, and X' capacitive. Then Eq. (537) becomes

$$(M'^2 - L_bL'')\omega^2 - \left(\frac{M'}{C_{pg}} - \frac{L_b}{C'}\right) = 0 \quad (538)$$

The balance can be made independent of frequency only if

$$M'^2 - L_bL'' = 0 \quad (539)$$

This demands unity coupling between L_b and L'', in which case the value of C' for balance is

$$C' = \frac{L_b}{M'}C_{pg} \quad (540)$$

If $M'^2 = L_b L''$,

$$\left.\begin{aligned} C' &= \sqrt{\frac{L_b}{L''}} C_{pg} \\ &= \frac{N_b}{N''} C_{pg}, \end{aligned}\right\} \tag{541}$$

where N_b and N'' are the numbers of turns in L_b and L''.

The actual connections for the method of balance or neutralization just described are given in Fig. 254. L'' and L_b are usually wound very close together in order to secure the necessary close coupling. They are wound in opposite directions to give the proper phase to the potential induced in L''. C' is called the *neutralizing condenser* and is usually a very small variable air condenser, the capacitance of which is set by methods to be described.

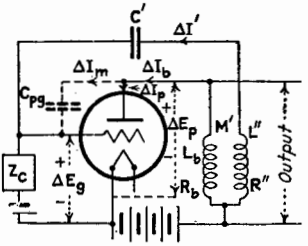


FIG. 254.—Neutralization by current balance (Hazeltine).

It should be noted that the plate current passes through L_b, whereas only the small neutralizing current passes through L''. Any kind of output circuit connected conductively or inductively to L_b may be used, and any type of input circuit may be connected between grid and filament without altering the conditions

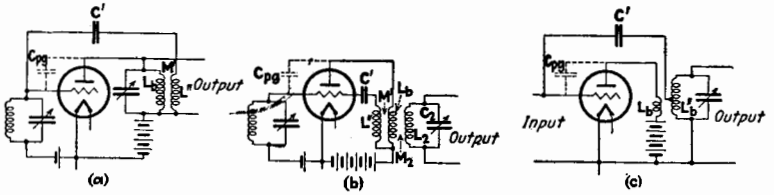


FIG. 255.—Methods of neutralization by current balances.

of balance expressed in Eq. (536). Figures 255, *a*, *b*, and *c*, are connections which conform to the bridge arrangement of Fig. 253, to the conditions of Eq. (534), and to Eq. (540), provided L'' and L_b are closely coupled in each case.

If in Fig. 253 the plate and grid are interchanged, the condition of balance is not altered, but the system is one in which the feed-back voltage is balanced and becomes identical with that shown in Figs. 249 and 250 and discussed under the first group.

The methods of neutralizing feed-back shown in Figs. 254 and 255*a*, *b*, and *c*, were described by Hazeltine. The scheme of Fig. 254 is given in his U. S. patents 1,489,228 and 1,533,858, and the schemes of Fig. 255 in his U. S. patent 1,648,808. Many radio sets, containing radio-frequency amplifiers neutralized by the method just described, have been manufactured and are often called "neutrodynes" in the trade. It should be remembered, however, that the methods of Figs. 254 and 255 all demand essentially unity coupling between the coils in order that the balance may be practically independent of frequency. If the coupling departs considerably from unity, the degree of

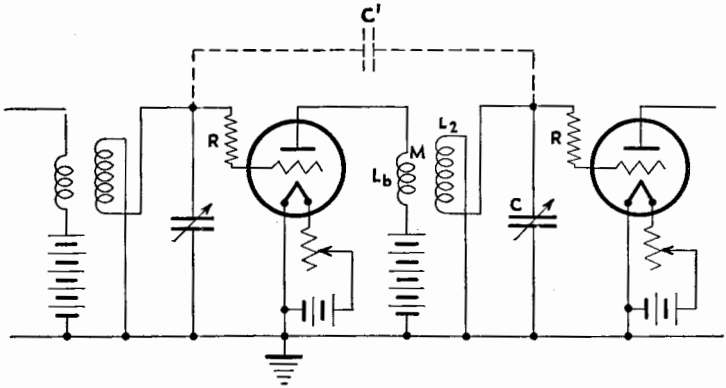


FIG. 256.—Neutralization through inherent capacitance.

balance of the effect of C_{pg} varies with frequency, but even then a considerable benefit is derived from the use of this balancing scheme.

Frequently, a considerable degree of neutralization is obtained because of inherent capacitance between wires and especially between the tuning condensers, as indicated by the dotted lines in Fig. 256. If the secondary coil L_2 is reversely connected, as indicated in the figure, so that the potential of the ungrounded end of L_2 is approximately opposite in phase to the potential of the plate, any natural capacitance, indicated by C', which exists between the grid systems acts to neutralize the effect of C_{pg}, partially at least. The greatest part of C' usually exists between the stators of the tuning condensers, the rotors being at ground potential. The neutralization by the use of natural

capacitances is generally not complete and varies with frequency because of the fact that the coefficient of coupling between L_b and L_2 is usually considerably less than unity. In such a case, regeneration is greatest at high frequencies. To prevent oscillation at these frequencies, a resistance R of several hundred ohms is often inserted in the grid lead, as shown in Fig. 256. When the circuits are tuned for high frequencies, the capacitance of the tuning condensers C is small, and consequently an increased fraction of the oscillatory current passes through R and the capacitance C_{gf}. Thus the equivalent resistance of the oscillatory circuit is increased at the high frequencies by an amount sufficient to offset the increasing regenerative effect as the capacitance of the tuning condensers is reduced.

Before considering neutralizing circuits which depart somewhat from the general scheme shown in Fig. 253, examine the input admittance of the system of this figure. If it were possible to balance the bridge perfectly, a voltage at the input would produce no voltage from plate to filament except the voltage which is due to the amplifying action of the triode. Considering only the bridge circuit, there would be no voltage across Z_b and hence no voltage across Z'' due to ΔE_g. Then, Z' and C_{pg} would be effectively in parallel between grid and filament and the total input admittance would be

$$B_g = -(C_{gf} + C_{pg} + C')\omega \tag{542}$$

provided Z' is a pure capacitance. Actually, as has been pointed out, it is impossible to balance the bridge perfectly for all frequencies, so that the input admittance is not so simple as just given and actually has a real component. However, for practical purposes, Eq. (542) gives closely enough the input admittance for this case.

The scheme shown in Fig. 253 is especially easy of adaptation to push-pull amplifiers, as shown in Fig. 257. Since the potential with respect to the filament of one end of the primary winding is equal and opposite in phase to the potential with respect to the filament of the other end of the winding, the neutralizing capacitances, made equal to C_{pg}, can be cross-connected, as shown in the figure. It is not necessary that the two halves of the primary winding have unity coupling, because the joint action of the two tubes maintains the proper magnitude and phase of the neutralizing potentials.

In the circuit of Fig. 254, L_b is generally large so that the small alternating plate current produces a large output voltage across the coil. It is necessary to have very close coupling between L_b and L'' in order to satisfy the conditions for nearly complete neutralization. On the other hand, the reactances of coils L_c in Fig. 250 and L_b in Fig. 255*a* are normally smaller than the reactance of L_b in Fig. 254. This is because the circuits con-

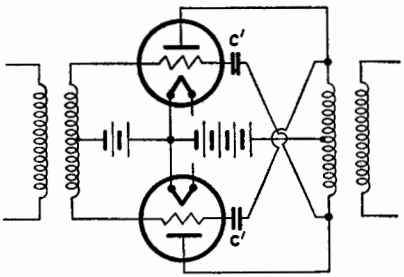


FIG. 257.—Neutralization of "push-pull" amplifier.

taining L_c and L_b in Figs. 250 and 255*a* are tuned by condensers, and relatively large currents circulate through the coils producing relatively large voltages across the coils. In the cases shown by Figs. 250 and 255*a*, unity coupling is necessary between the coil in the oscillatory circuit and the neutralizing winding. In the circuit shown in Fig. 255*b*, however, L_b is relatively small and, being untuned, the primary or plate current alone produces a relatively small voltage across it. Hence, little feed-back results from the primary current alone. The circulating current in the tuned secondary coil L_2 is, on the other hand, relatively large and the voltage induced in L_b by this large current is also relatively large and produces most of the feed-back. In circuits of the type shown in Fig. 255*b*, it is convenient to treat separately the two sources of feed-back which are called the *primary* and *secondary feed-back effects.*

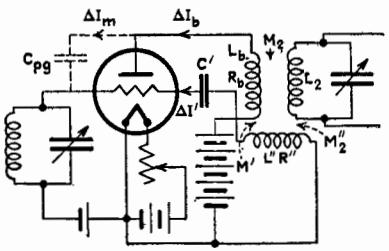


FIG. 258.—Neutralization of feed-back due to both primary and secondary currents.

Usually, because of the small primary feed-back, the coupling between L'' and L_b in Fig. 255*b* is of less importance than the

couplings between L_b and L_2 and between L'' and L_2. The primary regenerative feed-back is made small by making L_b small and sometimes by winding L_b of high-resistance wire. Denote the main mutual inductance between L_b and L_2 by M_2, the mutual inductance between L_b and L'' by M', and the mutual inductance between L'' and L_2 by M_2'', as shown in Fig. 258 which is similar to Fig. 255*b*. The equations expressing the conditions of balance are

$$\left.\begin{aligned} \mathbf{Z}_b\Delta\mathbf{I}_b - jM_2\omega\Delta\mathbf{I}_2 - jM'\omega\Delta\mathbf{I}' - \frac{j\Delta\mathbf{I}_m}{C_{pg}\omega} &= 0 \\ \mathbf{Z}''\Delta\mathbf{I}' + jM_2''\omega\Delta\mathbf{I}_2 - jM'\omega\Delta\mathbf{I}_b - \frac{j\Delta\mathbf{I}'}{C'\omega} &= 0 \\ \Delta\mathbf{I}' = -\Delta\mathbf{I}_m & \end{aligned}\right\} \quad (543)$$

Solving these equations,

$$[M_2M'\omega^2 + j\mathbf{Z}_bM_2''\omega]\Delta\mathbf{I}_2 + \left[M'^2\omega^2 - \frac{M'}{C_{pg}} + \mathbf{Z}_b\left(\mathbf{Z}'' - \frac{j}{C'\omega}\right)\right]\Delta\mathbf{I}' = 0 \quad (544)$$

In order that Eq. (544) may be true for any values of ΔI_2 and $\Delta I'$, each bracket must be zero. Hence,

$$M'^2\omega^2 - \frac{M'}{C_{pg}} + \mathbf{Z}_b\left(\mathbf{Z}'' - \frac{j}{C'\omega}\right) = 0 \text{ (\textit{primary balance})} \quad (545)$$

$$M_2M'\omega^2 + j\mathbf{Z}_bM_2''\omega = 0 \text{ (\textit{secondary balance})} \quad (546)$$

Equation (545) is the same as Eq. (535) and expresses the conditions for primary balance. Equation (546) gives the conditions for secondary balance, since it makes Eq. (544) independent of ΔI_2.

It has been pointed out that primary balance cannot be made perfect except at a single frequency or unless the resistances are zero or are negligible, but that from a practical point of view balance is satisfactory and sufficiently independent of frequency if Eq. (545) is satisfied when resistances are neglected. Examining Eq. (546), it is evident that perfect secondary balance can never be obtained unless R_b is zero. Neglecting R_b, which is quite warranted for practical purposes, the condition of secondary balance is

$$\frac{M_2''}{M_2} = \frac{M'}{L_b} \quad \text{(approx. secondary balance)} \tag{547}$$

Secondary balance is evidently independent of the values of C' and C_{pg} and must be attained by adjusting one or more of the quantities in Eq. (547). It is also apparent that secondary balance can be approximately obtained without primary balance. But if the circuits are adjusted for primary balance (neglecting resistances), secondary balance is automatically obtained. This follows because primary balance demands that $M'^2 = L_b L''$, and substituting this value of M' in Eq. (547) gives $M_2''/M_2 = N''/N_b$, a relation that is necessarily true if L_b and L'' are tightly coupled.

Another scheme often used in practice for obtaining approximate balance by the use of mutual inductance is that shown in Fig. 259. There is no separate neutralizing winding, but the

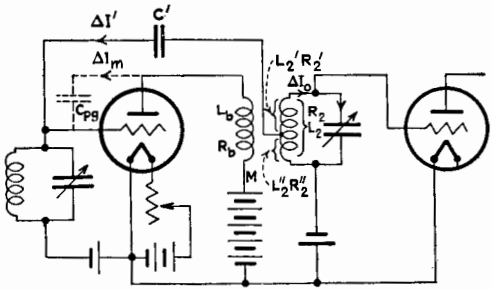


FIG. 259.—Neutralization by current balance using tap on secondary coil.

neutralizing capacitance C' is connected to a point in the secondary winding. Occasionally, one finds that the neutralizing capacitance consists of a single small disk-shaped terminal, adjustably located near the secondary winding, so that the capacitance C' is the capacitance between this disk and the winding itself. There is in this case no conductive connection to L_2. This arrangement is essentially equivalent to that shown in Fig. 259. The arrangement shown in Fig. 255*c*, in which there is close coupling between L'' and L_b, is a special case of the arrangement of Fig. 259.

For deducing the conditions of balance for Fig. 259, denote the inductance of the whole of the secondary winding by L_2, and the two portions above and below the neutralizing connection by L_2' and L_2''. The mutual inductances between the various self-inductances are defined as

M between L_b and L_2
M' between L_b and L_2''
M_2 between L_2' and L_2''

The equations for deriving the conditions of balance are

$$\left.\begin{aligned} \mathbf{Z}_b\Delta\mathbf{I}_b - \frac{j\Delta\mathbf{I}_m}{C_{pg}\omega} - jM\omega\Delta\mathbf{I}_0 - jM'\omega\Delta\mathbf{I}' &= 0 \\ jM'\omega\Delta\mathbf{I}_b - (\mathbf{Z}_2'' + jM_2\omega)\Delta\mathbf{I}_0 - \left(\mathbf{Z}_2'' - \frac{j}{C'\omega}\right)\Delta\mathbf{I}' &= 0 \\ \Delta\mathbf{I}_m = -\Delta\mathbf{I}' & \end{aligned}\right\} \quad (548)$$

where $\mathbf{Z}_2'' = R_2'' + jL_2''\omega$.

Eliminating $\Delta\mathbf{I}_m$ and $\Delta\mathbf{I}_b$ gives

$$(M'M\omega^2 + \mathbf{Z}_b\mathbf{Z}_2'' + j\mathbf{Z}_bM_2\omega)\Delta\mathbf{I}_0 + \left(M'^2\omega^2 - \frac{M'}{C_{pg}} + \mathbf{Z}_b\mathbf{Z}_2'' - \frac{j\mathbf{Z}_b}{C'\omega}\right)\Delta\mathbf{I}' = 0 \quad (549)$$

In Eq. (549), each parenthesis must be equal to zero. The second parenthesis placed equal to zero expresses the condition for primary balance and is the same as Eqs. (535) and (545), if L_2'' is considered the neutralizing winding.

The first parenthesis placed equal to zero is the condition for secondary balance, the more important of the two if L_b is small, as is usually the case. The first parenthesis separates into the two following conditions for secondary balance:

$$\left.\begin{aligned} [M'M - L_b(L_2'' + M_2)]\omega^2 + R_bR_2'' &= 0 \\ R_2''L_b\omega + R_b(L_2'' + M_2)\omega &= 0 \end{aligned}\right\} \quad (550)$$

Perfect secondary balance cannot be obtained because the second equation cannot be satisfied. If, however, R_2'' and R_b are small, the resistance terms are negligible and approximate secondary balance is secured if

$$M'M = L_b(L_2'' + M_2) = \frac{L_b}{2}(L_2 + L_2'' - L_2') \quad (551)$$

This condition of balance does not involve C', so that the only object in varying one or more of the quantities in Eq. (551) is to adjust for approximate secondary balance. Primary balance should be obtained by varying one or more of the

quantities in Eq. (545). Usually, the coupling between L_b and L_2'' is made quite close in order to approximate primary balance. In such a case, L_b is usually fixed in position, and secondary balance is most easily attained by varying the contact to the secondary winding or by moving along the coil the small disk terminal forming the capacitance connection to L_2. If L_b is wound close to L_2'', M should be such as to satisfy the optimum conditions given under Case 7, Chap. XVII.

The scheme of neutralization just described is cheaper to construct than other schemes, but the quantities are so interrelated that the design is more difficult.

If the coefficient of coupling between L_b and L_2'' in Fig. 259 is unity, Eq. (551) is satisfied and the scheme of Fig. 259 becomes that of Fig. 255*c*.

194. Experimental Method of Setting Neutralizing Capacitance.—The connections just described provide means for eliminating practically all of the reaction of the varying plate voltage upon the input or grid voltage of a triode. The ideal way of setting the constants to the correct values for securing balance would be to impress a varying voltage between plate and filament and to adjust the neutralizing capacitance and the mutual inductances when necessary, until no voltage is produced from grid to filament, the triode being in an operating condition. Since this method is inconvenient, the reciprocal property of the bridge schemes is generally used and a voltage produced by an intense signal is caused to act across the input of the triode, the filament of which is cold. The neutralizing capacitance is adjusted until no alternating voltage is produced between plate and filament. If the amplifier is one stage of the radio-frequency amplifier of a receiving set, this method is generally applied by tuning the set for an intense signal and adjusting the neutralizing capacitance until a minimum sound or no sound is heard in the output. We may call this the *forward-balance* method as distinguished from the more nearly correct *backward-balance* method.

It is well to determine if the forward-balance method always gives the correct setting for backward balance. When applying the forward-balance method, the filament must be cold, whereas the backward-balance method should be applied with the filament operating. Hence, if the capacitance C_{pg} is altered by heating the filament, the two methods are not equivalent. There

is evidence that C_{pg} may differ considerably according as the cathode is hot or cold.

The fact that r_p is infinite in one case and not in the other makes no difference in those cases in which the output voltage is the plate voltage ΔE_p. For example, in the bridges of Figs. 247*a*, 248*a*, 249, and 253, the forward balance is valid so far as being affected by the value of r_p. The two methods are not equivalent in the Hartley scheme of Fig. 246. In the schemes shown in Figs. 255*b*, 255*c*, 256, 258, and 259, the output voltage is not the same as the plate voltage and the forward balance, which gives a minimum of *output* voltage, may not give a minimum of plate voltage. In these cases the two methods may not be equivalent. In Fig. 258, it is possible to adjust the mutual inductance between L_b and L_2 and between L'' and L_2 so as to give no output voltage by the forward-balance method and yet have very small mutual inductance between L'' and L_b. In such a case secondary balance exists but primary balance is not complete. The forward-balance method may indicate complete neutralization, yet, when the tube is operated, regeneration may be so great as to make the tube oscillate because of incomplete primary balance. With the circuit constants used in practice, the forward-balance method is usually sufficiently correct.

195. Methods of Testing the Degree of Neutralization.—Two methods will now be described for determining experimentally whether or not an amplifier stage is neutralized, and for giving some measure of the degree of neutralization.

Method 1.—The first method consists in measuring the input voltage when the filament of the tube is cold and also when the tube is operating. If the stage is perfectly neutralized and there is no conduction current to the grid, the input voltage should be the same in the two cases. If the stage is not perfectly neutralized, the input voltage will be greater if the feed-back is regenerative, or less if the feed-back is absorptive. Another way of applying this method is to measure the input voltage when the triode is operating normally and also when the impedance in the plate circuit is short-circuited, the filament being hot during both measurements. One should be careful not to short-circuit the source of steady voltage.

The voltage across the input is measured by a vacuum-tube voltmeter which is of sufficient sensitivity so that an excessive

signal is not required, and which is so connected to the input that no losses are added and the feed-back is not changed by any added capacitances. It is sometimes very difficult to prevent the measuring apparatus from altering the feed-back by a considerable amount, especially when natural capacitances are important in the neutralization of the set, as illustrated in Fig. 256. The author has used an unbased low-capacitance triode for the voltmeter, connected to the grid terminal of the input through a shielded capacitance of 2 or 3$\mu\mu$f. Such a shielded capacitance is shown in Fig. 260. The terminal a was touched to the high-potential side of the tuning condenser in the input circuit.

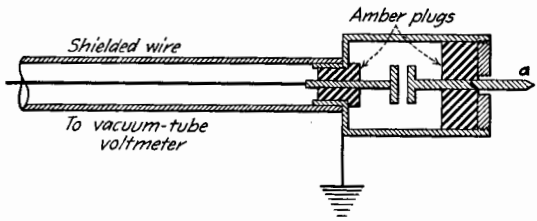


FIG. 260.—Shielded capacitance used in testing amplifiers. Three-quarters natural size.

Method 2.—The second method for giving an approximate idea of the degree of neutralization of a stage of an amplifier has come to be known as the "dummy-tube test." The triode is removed from its socket and a small variable condenser, which is connected between the plate and grid prongs of a tube base, is inserted in the socket. The capacitance of this condenser is varied until a minimum of sound is heard in the output when applying the forward-balance method. The ratio of the capacitance of this dummy tube to the C_{pg} of the triode is arbitrarily defined as the degree of neutralization, usually expressed in percentage. Actually, the figure gives the degree to which the capacitance of the triode itself is balanced but does not measure the neutralization of other capacitances between grid and plate terminals such as the capacitance due to the socket and wiring. If the dummy-tube capacitance for forward balance were zero, the definition would give no neutralization, whereas all of the feed-back capacitances other than the tube capacitance might be neutralized.

For several reasons this method may give results which are incorrect. Sometimes, the capacitances between the elements of the triode and the coils, condensers, or wiring are important, and, unless the dummy tube has the same geometrical form as the removed tube, the feed-back capacitances may be altered. To do away with this source of error, a triode of the same form as the one removed, but with no filament, has been used for the dummy tube. The capacitance was varied by changing the level in the envelope of castor oil which serves as dielectric between the grid and plate electrodes. Unless the grid element is made smaller than that in the normal tube, only capacitances larger than C_{pg} can be obtained in this dummy tube.

The dummy-tube method may also give erroneous results in those schemes in which the forward-balance method is not equivalent to the backward-balance method.

General References

FELDTKELLER: Theorie neutralisierter Verstärkerketten, *Zeits. f. Hochfreuenztechnik*, **35**, 45 (1930).

BELOW: Die Anodenrückwerkung bei verschiedenen Röhrenschaltungen und ihre Verminderung, *Zeits. f. Hochfrequenztechnik*, **37**, 65 (1931).

CHAPTER XIX

SMALL-SIGNAL DETECTION BY A DIODE

Definitions.—A signal cannot be conveyed by an alternating current or wave unless some characteristic of that current or wave is modified. Any modification of the current or wave is called *modulation*, and the current or wave which conveys the signal is often called the *carrier* current or wave.

There are several forms of modulation. If the frequency of a current or wave is varied, the modulation is called *frequency modulation*. The most common form of modulation is a variation of amplitude and is called *amplitude modulation*. The simplest form of amplitude modulation is called *sinusoidal modulation*, in which case the amplitude of the carrier current

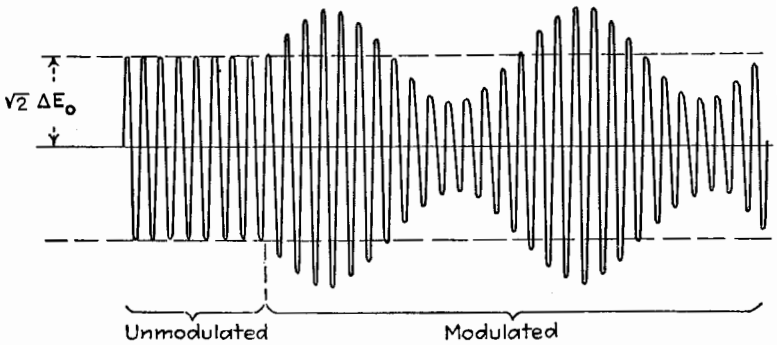


FIG. 261.—Sinusoidally modulated voltage.

or wave varies sinusoidally at a lower or *modulation frequency*. Sinusoidal modulation of a potential, for example, is expressed mathematically as follows.

$$\Delta e_0 = \sqrt{2}\Delta E_0(1 + m \sin \omega_l t) \sin \omega_h t \qquad (552)$$

In Eq. (552), $\sqrt{2}\Delta E_0$ is the unmodulated amplitude of the carrier potential having a frequency $\omega_h/2\pi$. The modulation frequency is $\omega_l/2\pi$. The factor m in Eq. (552) is known as the *degree of*

modulation and can have values ranging from zero to unity. Figure 261 is a time graph of a sinusoidally modulated potential.

The term *detector*, as used in the field of communication, signifies in general terms any device which, when acted upon by an alternating current or potential which is modulated in some way, usually for the purpose of conveying a signal, produces an effect characteristic of this modulation. Sometimes a detector is correctly called a *demodulator* in that it performs the reciprocal function of a modulator. A *modulator* is a device which modifies or modulates the alternating carrier current or potential by impressing upon it a characteristic corresponding to a lower frequency than that of the carrier current or potential.

The general definition of a detector as given includes such devices as a thermophone which consists of a fine platinum wire which is heated by the carrier current. If the amplitude of the carrier current varies at an audible frequency, the heat produced in the platinum wire varies accordingly and a sound wave of the modulation frequency is sent out. A hot-wire ammeter or thermal ammeter is a detector for slow variations in amplitude which the pointer of the ammeter can follow. Similarly, a glow tube actuated by the carrier potential gives out light dependent upon the amplitude of the impressed potential and is therefore a detector. The early Branley coherer and the Marconi magnetic detector are detectors within the scope of the definition given.

In this chapter the treatment is confined to those detectors in which the output is an electric current, the source of which is the carrier current, in part at least. Such detectors are known as *rectifying detectors* and include the thermionic diode as well as crystal and electrolytic detectors.

All rectifying detectors have one property in common, namely, their volt-ampere characteristic is *nonlinear*. This property is in marked contrast to the linear characteristic of an amplifier. The theory of the two-electrode detector is thus the theory of a nonlinear circuit element. This theory will be developed in general terms and can be applied to any two-electrode rectifying detector. In this chapter the theory is limited to small electrical changes in order to make certain simplifying assumptions.

196. Two-terminal Detector with No Load. Unmodulated Signal.—First, some simple relations will be developed concerning nonlinear circuit elements. These will be of use in the more complex theory to follow. Detector D in Fig. 262 is connected

in series with a polarizing battery $\bar{E}_B$ and a small alternating potential Δe_0 of frequency $\omega_h/2\pi$. The detector D has a nonlinear characteristic, given by the general relation

$$i = f(e) \tag{553}$$

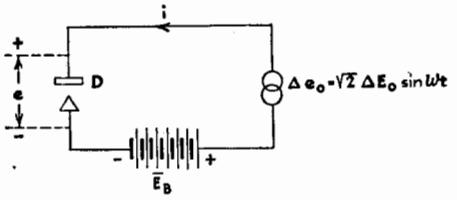


FIG. 262.—Diode detector without load.

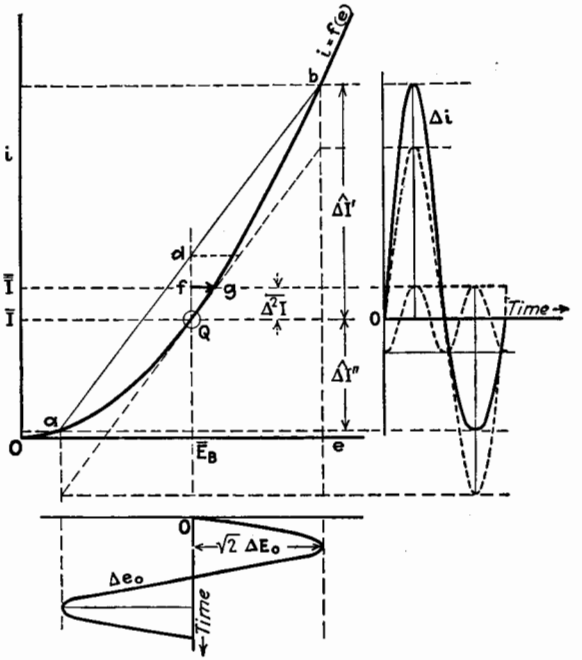


FIG. 263.—Rectification by two-electrode non-linear circuit element without load.

the form of which may be known only graphically, as illustrated in Fig. 263.

If Δe_0 is zero, the steady current that flows is $\bar{I}$ and is determined by substituting $\bar{E}_B$ in Eq. (553). Hence

$$\bar{I} = f(\bar{E}_B) \tag{554}$$

When Δe_0 also acts in series with $\bar{E}_B$, the current that flows is

$$i = f(\bar{E}_B + \sqrt{2}\Delta E_0 \sin \omega_h t) \tag{555}$$

If Δe_0 is assumed so small (the electrical variations are much exaggerated in Fig. 263 and in the figures that follow) that the operating portions of the characteristic curve between points a and b can be represented with sufficient accuracy by an expression containing no terms higher than the second degree, the variations in current and potential measured from the point Q are expressed by

$$\Delta i = A^{\mathrm{I}}\Delta e_0 + \frac{A^{\mathrm{II}}}{2}(\Delta e_0)^2 \tag{556}$$

where $A^{\mathrm{I}} = di/de$ at Q, and $A^{\mathrm{II}} = d^2i/de^2$ at Q. Substituting these values of A^{I}, A^{II}, and Δe_0 in Eq. (556), and noting that $\sin^2 \omega_h t = \frac{1}{2} - \frac{1}{2} \cos 2\omega_h t$,

$$\Delta i = \frac{1}{2} \cdot \frac{d^2i}{de^2}(\Delta E_0)^2 + \frac{di}{de}\sqrt{2}(\Delta E_0) \sin \omega_h t - \frac{1}{2} \cdot \frac{d^2i}{de^2}(\Delta E_0)^2 \cos 2\omega_h t \tag{557}$$

[1] It is instructive to inquire as to the error produced in the several terms of Eq. (557) if the portion of the characteristic curve traveled over by the operating point is not accurately represented by a second-degree equation but requires some higher-degree terms. Assume that two more terms added to Eq. (556) are necessary to represent the curve.

$$\Delta i = A^{\mathrm{I}}\Delta e_0 + \frac{A^{\mathrm{II}}}{2}(\Delta e_0)^2 + \frac{A^{\mathrm{III}}}{6}(\Delta e_0)^3 + \frac{A^{\mathrm{IV}}}{24}(\Delta e_0)^4 \tag{556a}$$

where A^{I}, A^{II}, A^{III}, and A^{IV} are the first, second, third, and fourth derivatives of the curve at point Q. Substituting the value of Δe_0 in Eq. (556a)

$$\begin{aligned}\Delta i = {} & \left(\frac{1}{2}A^{\mathrm{II}}(\Delta E_0)^2 + \frac{1}{16}A^{\mathrm{IV}}(\Delta E_0)^4\right) + \left(\sqrt{2}A^{\mathrm{I}}\Delta E_0 + \frac{\sqrt{2}}{4}A^{\mathrm{III}}(\Delta E_0)^3\right) \sin \omega_h t \\ & - \left(\frac{1}{2}A^{\mathrm{II}}(\Delta E_0)^2 + \frac{1}{\sqrt{2}}A^{\mathrm{IV}}(\Delta E_0)^4\right) \cos 2\omega_h t \\ & - \frac{\sqrt{2}}{12}A^{\mathrm{III}}(\Delta E_0)^3 \sin 3\omega_h t + \frac{1}{48}A^{\mathrm{IV}}(\Delta E_0)^4 \cos 4\omega_h t \end{aligned} \tag{556b}$$

Equation (556b) shows that the third derivative does not affect the steady component but does add to the fundamental and adds a third harmonic. The fourth derivative adds slightly to the steady term and to the second harmonic and adds a fourth harmonic of very small amplitude. The size of the coefficients shows the magnitude of the error introduced if the portion of the characteristic used is not exactly parabolic.

The term $\frac{1}{2}\frac{d^2i}{de^2}(\Delta E_0)^2$ is the only term of frequency lower than that of the impressed voltage and is the important term when detection or rectification is considered. This term is a quantity of the second order in magnitude, because it depends on $(\Delta E_0)^2$, and is denoted by $\overline{\Delta^2 I}$. Hence, we write

$$\overline{\Delta^2 I} = \frac{1}{2}\frac{d^2i}{de^2}(\Delta E_0)^2 \tag{558}$$

One cycle of Δi and of the several components of Δi in Eq. (557) is plotted in Fig. 263. The amplitude of the current of fundamental frequency can be obtained from the tangent to the curve at Q whose slope is di/de, Fig. 263. The second harmonic has an amplitude which is equal to $\overline{\Delta^2 I}$.

If a secant is drawn from a to b, and d is the intersection of this secant with the vertical through Q, distance Qd is equal to $2\overline{\Delta^2 I}$. Let $\widehat{\Delta I}'$ and $\widehat{\Delta I}''$, both taken as positive quantities, be the peak values of Δi measured from $\bar{I}$. From Eq. (556)

$$\left.\begin{aligned} \widehat{\Delta I}' &= \sqrt{2}A^{\mathrm{I}}\Delta E_0 + A^{\mathrm{II}}(\Delta E_0)^2 \\ -\widehat{\Delta I}'' &= -\sqrt{2}A^{\mathrm{I}}\Delta E_0 + A^{\mathrm{II}}(\Delta E_0)^2 \end{aligned}\right\} \tag{559}$$

Now, $\overline{Qd} = \dfrac{\widehat{\Delta I}' + \widehat{\Delta I}''}{2} - \widehat{\Delta I}'' = \dfrac{\widehat{\Delta I}' - \widehat{\Delta I}''}{2}$. From Eq. (559)

$$\frac{\widehat{\Delta I}' - \widehat{\Delta I}''}{2} = A^{\mathrm{II}}(\Delta E_0)^2 = 2\overline{\Delta^2 I} \tag{560}$$

From Eqs. (559) and (560) we obtain the important relation

$$\frac{\widehat{\Delta I}' - \widehat{\Delta I}''}{4} = \overline{\Delta^2 I} \tag{561}$$

The importance of Eq. (561) is this: We have found that when a *sinusoidal* e.m.f. is impressed upon a device having a parabolic characteristic curve, the change in steady current, or the rectified current, is given by Eq. (558); and, further, Eq. (561) shows that this same change in steady current is given by one-quarter of the difference of the peak values in the positive and negative directions. This relation will be used as a short cut in later theory and will be extended to mean that, when the nonlinear element has a parabolic characteristic, one-quarter of the difference of the positive and negative peak values of any potential or current gives the value of the steady component of that potential or

current. This is strictly true only if the applied potential is sinusoidal. It is nearly enough true if the amplitudes of the harmonics of the applied potential are small in comparison with the amplitude of the fundamental.

Equation (558) may also be deduced by the use of Taylor's series development and will be given because of the frequent use of this method later. Let the characteristic curve be, as before,

$$i = f(e) \tag{553}$$

The steady polarizing potential $\bar{E}_B$ gives a current $\bar{I}$, where

$$\bar{I} = f(\bar{E}_B) \tag{554}$$

If the applied potential is increased by an amount $\widehat{\Delta E}_0$, then by Taylor's theorem

$$\begin{aligned}\bar{I} + \widehat{\Delta I}' &= f(\bar{E}_B + \widehat{\Delta E}_0) \\ &= f(\bar{E}_B) + \frac{di}{de}\widehat{\Delta E}_0 + \frac{1}{2}\cdot\frac{d^2i}{de^2}(\widehat{\Delta E}_0)^2 + \frac{1}{6}\cdot\frac{d^3i}{de^3}(\widehat{\Delta E}_0)^3 + \cdots\end{aligned} \tag{562}$$

If the applied potential is decreased by an amount $\widehat{\Delta E}_0$,

$$\begin{aligned}\bar{I} - \widehat{\Delta I}'' &= f(\bar{E}_B - \widehat{\Delta E}_0) \\ &= f(\bar{E}_B) - \frac{di}{de}(\widehat{\Delta E}_0) + \frac{1}{2}\cdot\frac{d^2i}{de^2}(\widehat{\Delta E}_0)^2 - \frac{1}{6}\cdot\frac{d^3i}{de^3}(\widehat{\Delta E}_0)^3 + \cdots\end{aligned} \tag{563}$$

Adding Eqs. (562) and (563),

$$\left.\begin{aligned}\frac{\widehat{\Delta I}' - \widehat{\Delta I}''}{4} &= \frac{1}{4}\cdot\frac{d^2i}{de^2}(\widehat{\Delta E}_0)^2 + \cdots \\ &= \frac{1}{2}\cdot\frac{d^2i}{de^2}(\Delta E_0)^2 + \cdots\end{aligned}\right\} \tag{564}$$

If the change, $\widehat{\Delta E}_0$, in potential is the amplitude of a sinusoidal potential, and if terms of order higher than the second in Taylor's development are neglected, then, according to Eq. (561), Eq. (564) gives the change in steady current $\overline{\Delta^2 I}$. Since terms of order higher than the second in Taylor's development are neglected in Eqs. (562) and (563), it is implicitly assumed that the applied potential is sufficiently small to permit of expressing by an equation of degree no higher than the second that portion of the characteristic curve in the neighborhood of $\bar{E}_B$ which is traversed in operation. This does not mean that the characteristic curve taken as a whole can be expressed by a second-

degree equation. This assumption puts no limitation upon the shape of the characteristic curve, except that the first and second derivatives of the curve are finite and continuous in the region of operation.

An interesting and useful conception is brought out by multiplying Eq. (558) by the variational resistance of the detector, *i.e.*, by $r = de/di = 1/k$. The left-hand side of the equation thus obtained is a fictitious voltage $[\overline{\Delta^2 E}]$ which can be considered as a voltage produced by detection, because this voltage acting in the circuit containing resistance r gives the current $\overline{\Delta^2 I}$ by Ohm's law. Hence,

$$[\overline{\Delta^2 E}] = \frac{1}{2} \cdot \frac{d^2 i/de^2}{di/de} (\Delta E_0)^2 \tag{565}$$

where ΔE_0 is the r.m.s. value of the applied high-frequency voltage.

This fictitious voltage is represented by the line fg drawn to the tangent to the curve at Q, Fig. 263. This conception of a fictitious voltage of detection will be useful later.

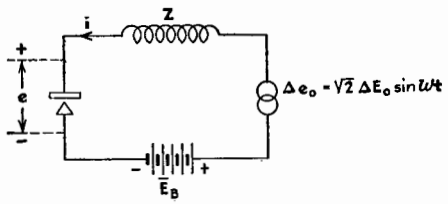


FIG. 264.—Diode detector with load.

197. Two-terminal Detector in Series with Impedance. Unmodulated Signal.—Consider a more general case when, as shown in Fig. 264, an impedance Z is included in the circuit. This impedance represents the sum total of all impedances external to the detector and may have a value $\bar{R}$ for direct currents and a different value Z_h for the high frequency $\omega_h/2\pi$ of the impressed e.m.f. Δe_0. The detector has a resistance of

$$r = \frac{de}{di} \tag{566}$$

for small variations of potential.

Referring to Fig. 264, when Δe_0 is impressed, a current of radio frequency $\omega_h/2\pi$ flows through the circuit and, in addition, the steady current is altered by an amount denoted by $\overline{\Delta^2 I}$. When

no alternating potential is impressed, $\bar{I} = f(\bar{E})$, where $\bar{E}$ is the steady potential across the rectifier and is different from $\bar{E}_B$. Referring now to Fig. 265, the characteristic curve for the rectifier is shown as $i = f(e)$. From point $\bar{E}_B$, a point laid off on the voltage axis corresponding to the steady polarizing-battery potential, a line is drawn making an angle with the vertical whose tangent is equal to $\bar{R}$ the resistance of Z to steady currents.

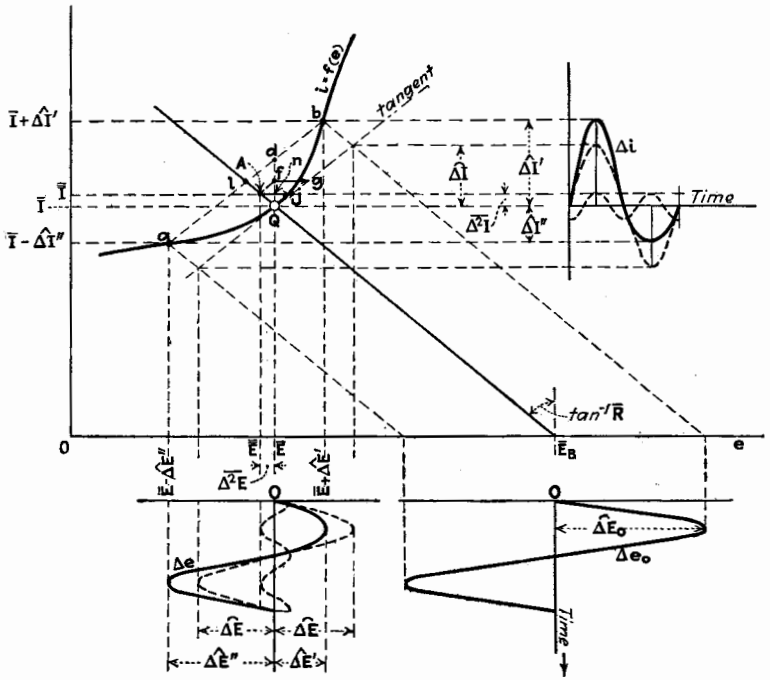


FIG. 265.—Rectification by two-electrode non-linear element in series with a resistance load.

(For simplicity in representation, Z is taken as a pure resistance, the same for steady and for alternating currents, but the mathematical derivation does not require this limitation.) The point Q of intersection of this resistance line and the current curve gives the values of $\bar{I}$ and $\bar{E}$. If $\Delta e_0 = \widehat{\Delta E}_0 \sin \omega_h t$ is superimposed upon $\bar{E}_B$, the new intersection points which are given by lines drawn parallel to the resistance line from points representing the extreme applied voltages $\bar{E}_B + \widehat{\Delta E}_0$ and $\bar{E}_B - \widehat{\Delta E}_0$ give the extreme currents $\bar{I} + \widehat{\Delta I}'$ and $\bar{I} - \widehat{\Delta I}''$, and also give the incre-

ments of voltage on the rectifier, $\widehat{\Delta E'}$ and $\widehat{\Delta E''}$. We may now write

$$\begin{aligned}\bar{I} + \widehat{\Delta I'} &= f(\bar{E} + \widehat{\Delta E'}) \\ &= f(\bar{E}) + \frac{di}{de}\widehat{\Delta E'} + \frac{1}{2}\cdot\frac{d^2i}{de^2}(\widehat{\Delta E'})^2 + \cdots \end{aligned} \tag{567}$$

and

$$\begin{aligned}\bar{I} - \widehat{\Delta I''} &= f(\bar{E} - \widehat{\Delta E''}) \\ &= f(\bar{E}) - \frac{di}{de}\widehat{\Delta E''} + \frac{1}{2}\cdot\frac{d^2i}{de^2}(\widehat{\Delta E''})^2 + \cdots \end{aligned} \tag{568}$$

Adding Eqs. (567) and (568),

$$\widehat{\Delta I'} - \widehat{\Delta I''} = \frac{di}{de}(\widehat{\Delta E'} - \widehat{\Delta E''}) + \frac{d^2i}{de^2}\left(\frac{(\widehat{\Delta E'})^2 + (\widehat{\Delta E''})^2}{2}\right) + \cdots \tag{569}$$

The voltage wave across the detector is not sinusoidal in this case but has peak values $\widehat{\Delta E'}$ and $\widehat{\Delta E''}$, which are different in the positive and negative directions. This voltage wave can be analyzed to a sufficient degree of accuracy into a steady component,

$$\overline{\Delta^2 E} = \frac{\widehat{\Delta E'} - \widehat{\Delta E''}}{4} \tag{570}$$

into a potential of fundamental frequency of amplitude

$$(\widehat{\Delta E})^2 = \frac{(\widehat{\Delta E'})^2 + (\widehat{\Delta E''})^2}{2} \tag{571}$$

and into a second-harmonic potential. The second-harmonic potential has an amplitude of the second order in magnitude and in acting upon the detector produces a change in steady current measured in third-order magnitudes. Therefore, this second-harmonic component is neglected.

Substituting Eqs. (570) and (571) in Eq. (569),

$$\overline{\Delta^2 I} = \frac{di}{de}\overline{\Delta^2 E} + \frac{1}{2}\cdot\frac{d^2i}{de^2}(\Delta E)^2 \tag{572}$$

Since $\overline{\Delta^2 E}$ is the steady voltage across the detector and also the negative of the steady voltage across the impedance Z,

$$\overline{\Delta^2 E} = -\bar{R}\overline{\Delta^2 I} \tag{573}$$

and Eq. (572) can be written

$$\overline{\Delta^2 I} = \frac{r}{r + \bar{R}} \cdot \frac{1}{2} \cdot \frac{d^2 i}{de^2} (\Delta E)^2 \tag{574}$$

or

$$[\overline{\Delta^2 E}] = \frac{1}{2} \cdot \frac{d^2 i/de^2}{di/de} (\Delta E)^2 = \frac{1}{2k} \cdot \frac{dk}{de} (\Delta E)^2 \tag{575}$$

where $k = 1/r$ and ΔE is the r.m.s. high-frequency voltage across the detector.

Expression (575) for the fictitious voltage of detection is the same as Eq. (565), both being in terms of the high-frequency voltage existing across the detector.

If the impedance $(Z)_h$ is negligible, ΔE is the same as ΔE_0. Otherwise ΔE is less than ΔE_0 as shown by

$$\Delta E = \frac{r}{\sqrt{[r + (R)_h]^2 + (X)_h^2}} \Delta E_0 \tag{576}$$

where $(R)_h$ and $(X)_h$ are the resistance and reactance of $(Z)_h$ at the frequency $\omega_h/2\pi$. Although ΔE, in Eqs. (574) and (575), can be replaced by Eq. (576), giving the detection entirely in terms of the applied high-frequency potential ΔE_0, it is simpler to leave the expressions as in Eqs. (574) and (575) and calculate ΔE separately if necessary.

It is advantageous to make ΔE as large as possible in comparison with ΔE_0. Hence $(Z)_h$ should be as small as possible. This condition is usually attained in practice by shunting $(Z)_h$ with a condenser. Since $1/2k \cdot dk/de$ occurs in all the expressions for giving the detection with a two-terminal detector, it will be called the *voltage-detection coefficient for a diode* and will be denoted by

$$(\text{Det.}E)_d = \frac{1}{2k} \cdot \frac{dk}{de} \tag{577}$$

Then,

$$[\overline{\Delta^2 E}] = (\text{Det.}E)_d (\Delta E)^2 \tag{578}$$[2]

[2] Attention is called to the fact that the definition of (Det. E) given here is slightly different from that given by the author and Browning in the article A Theoretical and Experimental Investigation of Detection for Small Signals, *Proc. I.R.E.*, **15**, 113 (1927). This slight change was made in the interest of consistency and simplicity.

A graphical interpretation of some of the quantities can be had by referring to Fig. 265. The distance fg, taken from the mid-point f of the vertical line Qd to the tangent, represents the fictitious voltage $[\Delta^2\overline{E}]$ given by Eq. (575). This fictitious voltage of detection is also given by the distance Aj which is the distance to the tangent from the mid-point of distance Ql on the resistance line. Point A is the average point and gives the average current $\bar{\bar{I}}$ and the average voltage $\bar{\bar{E}}$ across the detector. The voltage, represented by $\overline{Aj}$, is equal to the sum of the increment of voltage across the detector and the increment of voltage across the resistance. The increment of steady voltage across the detector is $\Delta^2\overline{E} = (\Delta^2\overline{I})r$ and is represented by line An; the increment of steady voltage across the resistance is $(\Delta^2\overline{I})\bar{R}$ and is represented by line nj. The current through and the voltage across the detector are shown and also the components of the current and voltage waves are given in Fig. 265. $(R)_h$ is assumed equal to $\bar{R}$, and $(X)_h$ is assumed to be zero. The case in which the two resistances are not the same can easily be pictured.

198. Detection of Modulated Signal. First Method.—In radio communication, the detector usually rectifies a modulated radio-frequency e.m.f., giving rise to a current having a frequency equal to the modulation frequency. Figure 261 represents a sinusoidally modulated radio-frequency e.m.f., which is expressed mathematically by

$$\Delta e_0 = \widehat{\Delta E}_0(1 + m \sin \omega_l t) \sin \omega_h t \qquad (579)$$

In this expression, $\widehat{\Delta E}_0$ is the average amplitude of the alternating e.m.f. or the amplitude of the unmodulated radio-frequency e.m.f. of angular velocity ω_h; ω_l is the angular velocity corresponding to the modulation frequency; m is a factor ranging from 0 for no modulation to 1 for complete modulation and gives the fraction of complete modulation.

In Eq. (579), if ω_h and ω_l are sufficiently different so that there are many cycles of the high-frequency e.m.f. in one cycle of the low-frequency variation, the amplitude of the radio-frequency e.m.f. may be considered as varying relatively slowly as compared to the radio-frequency period, and the $\widehat{\Delta E}_0$ used in Secs. 196 and 197 of this chapter can be replaced by

$$\widehat{\Delta E}_0(1 + m \sin \omega_l t).$$

Similarly, the amplitude of the radio-frequency voltage across the detector can be replaced by $\widehat{\Delta E}(1 + m \sin \omega_l t)$.

With this pulsating amplitude of the applied high-frequency potential across the detector, $\overline{\Delta^2 I}$ of Eq. (572) becomes a pulsating current and may be represented by the symbol $\Delta^2 i$. Similarly, $\overline{\Delta^2 E}$ of Eq. (572) is now pulsating and will be denoted by $\Delta^2 e$. Equation (572) becomes

$$\Delta^2 i = \frac{\Delta^2 e}{r} + \frac{1}{2} \cdot \frac{d^2 i}{de^2}(1 + m \sin \omega_l t)^2 (\Delta E)^2 \tag{580}$$

where ΔE is related to ΔE_0 by Eq. (576) as before. Equation (580) breaks up into three equations, one for each frequency, as follows:

$$\overline{\Delta^2 I} = \frac{\overline{\Delta^2 E}}{r} + \frac{1}{2} \cdot \frac{d^2 i}{de^2}\left(1 + \frac{m^2}{2}\right)(\Delta E)^2 \tag{581}$$

$$(\boldsymbol{\Delta}^2 \mathbf{I})_l = \frac{(\boldsymbol{\Delta}^2 \mathbf{E})_l}{r} + \frac{1}{2} \cdot \frac{d^2 i}{de^2}\sqrt{2} m (\Delta E)^2 \tag{582}$$

$$(\boldsymbol{\Delta}^2 \mathbf{I})_{2l} = \frac{(\boldsymbol{\Delta}^2 \mathbf{E})_{2l}}{r} + \frac{1}{2} \cdot \frac{d^2 i}{de^2} \frac{m^2}{2\sqrt{2}} (\Delta E)^2 \tag{583}$$

where $(\boldsymbol{\Delta}^2 \mathbf{I})_l$ is the complex r.m.s. current of frequency $\omega_l/2\pi$, $(\boldsymbol{\Delta}^2 \mathbf{E})_l$ is the complex r.m.s. voltage of frequency $\omega_l/2\pi$ across the detector, and $(\boldsymbol{\Delta}^2 \mathbf{I})_{2l}$ and $(\boldsymbol{\Delta}^2 \mathbf{E})_{2l}$ are the corresponding quantities for frequency $\omega_{2l}/2\pi$.

Since currents of frequencies $\omega_l/2\pi$ and $\omega_{2l}/2\pi$, as well as steady currents and currents of high frequency, circulate in the circuit, it is well to redefine the various impedances of the circuit. The detector has a resistance r for currents of all frequencies. $\bar{R}$ is the equivalent series resistance to steady currents of the rest of the circuit external to the detector. $(Z)_l$ and $(Z)_{2l}$ are the series impedances of the circuit external to the detector for currents of frequencies $\omega_l/2\pi$ and $\omega_{2l}/2\pi$. Similarly, $(Z)_h$ is the series impedance for the high frequency of the circuit external to the detector, the series e.m.f. Δe_0 having no internal impedance.

Then

$$\overline{\Delta^2 E} = -\bar{R}\, \overline{\Delta^2 I} \tag{584}$$

$$(\boldsymbol{\Delta}^2 \mathbf{E})_l = -(\mathbf{Z})_l (\boldsymbol{\Delta}^2 \mathbf{I})_l \tag{585}$$

$$(\boldsymbol{\Delta}^2 \mathbf{E})_{2l} = -(\mathbf{Z})_{2l} (\boldsymbol{\Delta}^2 \mathbf{I})_{2l} \tag{586}$$

Substituting Eqs. (584), (585), and (586) in Eqs. (581), (582), and (583), respectively,

$$\overline{\Delta^2 I} = \frac{1}{r + \bar{R}} \cdot (\mathrm{Det}.E)_d \left(1 + \frac{m^2}{2}\right)(\Delta E)^2 \tag{587}$$

$$(\Delta^2 \mathbf{I})_l = \frac{1}{r + (\mathbf{Z})_l} \cdot (\text{Det.}E)_d \sqrt{2} m (\Delta E)^2 \tag{588}$$

$$(\Delta^2 \mathbf{I})_{2l} = \frac{1}{r + (\mathbf{Z})_{2l}} \cdot (\text{Det.}E)_d \frac{m^2}{2\sqrt{2}} (\Delta E)^2 \tag{589}$$

Expressed in terms of the three fictitious voltages of detection, these three equations may be written

$$[\overbrace{\Delta^2 E}] = (\text{Det.}E)_d \left(1 + \frac{m^2}{2}\right)(\Delta E)^2 \tag{590}$$

$$[(\Delta^2 E)_l] = (\text{Det.}E)_d \sqrt{2} m (\Delta E)^2 \tag{591}$$

$$[(\Delta^2 E)_{2l}] = (\text{Det.}E)_d \frac{m^2}{2\sqrt{2}} (\Delta E)^2 \tag{592}$$

The voltage $[(\Delta^2 E)_l]$ is in phase with the modulation which is expressed by the term $\sin \omega_l t$ in Eq. (579). In other words, it is in phase with the positive envelope of the modulated e.m.f. in Fig. 261. The voltage $[(\Delta^2 E)_{2l}]$ is related in phase to $[(\Delta^2 E)_l]$ as $-\cos 2\omega_l t$ is related to $\sin \omega_l t$.

Comparing Eq. (591) with Eq. (578), we see that the distance fg in Fig. 265 represents one-half the amplitude of the voltage $[(\Delta^2 E)_l]$ when $m = 1$. In other words, the horizontal distance from d to the tangent represents the amplitude of the fictitious voltage of modulation frequency for complete modulation.

***199. Detection of Modulated Signal. Second Method.**—A second method of deriving Eqs. (590), (591), and (592) will now be given. It is perhaps more rigorous but less simple than the first method.

Let the portion of the characteristic in the neighborhood of the quiescent point Q of Fig. 265 be given by the second-degree equation already used

$$\Delta i = A^{\text{I}} \Delta e + \frac{A^{\text{II}}}{2} (\Delta e)^2 \tag{556}$$

where A^{I} and A^{II} are the first and second derivatives of the characteristic curve at Q.

If we call e' the total instantaneous voltage across the impedance Z of Fig. 264,

$$\Delta e = \Delta e_0 - \Delta e' \tag{593}$$

where

$$\Delta e_0 = \widehat{\Delta E}_0 (1 + m \sin \omega_l t) \sin \omega_h t \tag{579}$$

Substituting Eq. (593) in Eq. (556)

$$\Delta i = A^{\mathrm{I}}(\Delta e_0 - \Delta e') + \frac{A^{\mathrm{II}}}{2}(\Delta e_0 - \Delta e')^2 \tag{594}$$

Since the circuit contains a detector which is a nonlinear element, Δi and Δe are made up of components of many frequencies. Although we do not yet know the amplitude and phase of the several components of Δi, we can express them as quantities whose magnitudes are to be determined later. The most important components of Δi are

$$\begin{aligned}\Delta i = {} & \overline{\Delta^2 I} + (\widehat{\Delta I})_h(1 + m \sin \omega_l t) \sin (\omega_h t - \alpha_h) \\ & + (\widehat{\Delta I})_{2h}(1 + m \sin \omega_l t) \sin (2\omega_h t - \alpha_{2h}) \\ & + (\widehat{\Delta^2 I})_l \sin (\omega_l t - \beta_l) + (\widehat{\Delta^2 I})_{2l} \sin (2\omega_l t - \beta_{2l}) \\ & + \cdots\end{aligned} \tag{595}$$

These components of Δi produce across Z the following components of $\Delta e'$

$$\begin{aligned}\Delta e' = {} & \bar{R}\overline{\Delta^2 I} + (R)_h(\widehat{\Delta I})_h(1 + m \sin \omega_l t) \sin (\omega_h t - \alpha_h) \\ & + (X)_h(\widehat{\Delta I})_h(1 + m \sin \omega_l t) \cos (\omega_h t - \alpha_h) \\ & + (R)_{2h}(\widehat{\Delta I})_{2h}(1 + m \sin \omega_l t) \sin (2\omega_h t - \alpha_{2h}) \\ & + (X)_{2h}(\widehat{\Delta I})_{2h}(1 + m \sin \omega_l t) \cos (2\omega_h t - \alpha_{2h}) \\ & + (R)_l(\widehat{\Delta^2 I})_l \sin (\omega_l t - \beta_l) \\ & \quad + (X)_l(\widehat{\Delta^2 I})_l \cos (\omega_l t - \beta_l) \\ & + (R)_{2l}(\widehat{\Delta^2 I})_{2l} \sin (2\omega_l t - \beta_{2l}) \\ & \quad + (X)_{2l}(\widehat{\Delta^2 I})_{2l} \cos (2\omega_l t - \beta_{2l}) + \cdots\end{aligned} \tag{596}$$

We may now break up Eq. (594) into as many separate equations as there are frequencies. First, pick out all constant terms in Eq. (594), remembering that $A^{\mathrm{I}} = 1/r$.

$$\begin{aligned}\overline{\Delta^2 I} = {} & -\frac{\bar{R}}{r}\overline{\Delta^2 I} + \frac{A^{\mathrm{II}}}{4}\left(1 + \frac{m^2}{2}\right)(\widehat{\Delta E}_0)^2 + \frac{A^{\mathrm{II}}}{4}\left(1 + \frac{m^2}{2}\right)(R)_h^2(\widehat{\Delta I})_h^2 \\ & + \frac{A^{\mathrm{II}}}{4}\left(1 + \frac{m^2}{2}\right)(X)_h^2(\widehat{\Delta I})_h^2 + \frac{A^{\mathrm{II}}}{4}\left(1 + \frac{m^2}{2}\right)(R)_{2h}^2(\widehat{\Delta I})_{2h}^2 \\ & + \frac{A^{\mathrm{II}}}{4}\left(1 + \frac{m^2}{2}\right)(X)_{2h}^2(\widehat{\Delta I})_{2h}^2 - \frac{A^{\mathrm{II}}}{2}\left(1 + \frac{m^2}{2}\right)(R)_h(\widehat{\Delta E}_0)(\widehat{\Delta I})_h \cos \alpha_h \\ & - \frac{A^{\mathrm{II}}}{2}\left(1 + \frac{m^2}{2}\right)(X)_h(\widehat{\Delta E}_0)(\widehat{\Delta I})_h \sin \alpha_h + \cdots\end{aligned} \tag{597}$$

The terms containing $(\widehat{\Delta I})_{2h}^2$ in Eq. (597) may be neglected in comparison with the terms containing $(\widehat{\Delta I})_h^2$. Other terms

of a greater degree of smallness have been omitted. It is assumed in the process carried through above that both $(R)_h$ and $(X)_h$ are the same for the carrier frequencies $\omega_h/2\pi$ as for the side-band frequencies $\frac{\omega_h - \omega_l}{2\pi}$ and $\frac{\omega_h + \omega_l}{2\pi}$. This means that $(Z)_h$ cannot be a sharply tuned circuit.

In Eq. (597) let the value of $\cos \alpha_h = \dfrac{(R)_h + r}{\sqrt{[r + (R)_h]^2 + (X)_h^2}}$, $\sin \alpha_h = \dfrac{(X)_h}{(Z)_h}$, and $(\widehat{\Delta I})_h = \dfrac{\widehat{\Delta E}_0}{(Z)_h}$. Then Eq. (579) reduces to

$$\overline{\Delta^2 I} = \frac{r}{r + \bar{R}} \cdot \frac{r^2}{[r + (R)_h]^2 + (X)_h^2} \cdot \frac{A^{II}}{2}\left(1 + \frac{m^2}{2}\right)(\Delta E_0)^2 \quad (598)$$

This is the equivalent of Eq. (587).

The terms of modulation frequency $\omega_l/2\pi$ are next collected.

$$(\widehat{\Delta^2 I})_l \sin (\omega_l t - \beta_l) = -\frac{(R)_l}{r}(\widehat{\Delta^2 I})_l \sin (\omega_l t - \beta_l) - \frac{(X)_l}{r}(\widehat{\Delta^2 I})_l \cos (\omega_l t - \beta_l)$$
$$+ \frac{A^{II}}{4}[(\widehat{\Delta E}_0)^2 + (R)_h^2(\widehat{\Delta I})_h^2 + (X)_h^2(\Delta I)_h^2 - 2(R)_h(\widehat{\Delta E}_0)(\widehat{\Delta I})_h \cos \alpha_h - 2(X)_h(\widehat{\Delta E}_0)(\widehat{\Delta I})_h \sin \alpha_h] 2m \sin \omega_l t \quad (599)$$

Collecting terms and making substitution as before for $(\widehat{\Delta I})_h$, $\cos \alpha_h$, and $\sin \alpha_h$, we obtain

$$\sqrt{[r + (R)_l]^2 + (X)_l^2}(\widehat{\Delta^2 I})_l \sin\left(\omega_l t - \beta_l + \tan^{-1}\frac{(X)_l}{r + (R)_l}\right) = \frac{A^{II}}{2} r \frac{r^2}{[r + (R)_h]^2 + (X_h)^2} 2m(\widehat{\Delta E}_0)^2 \sin \omega_l t \quad (600)$$

Equation (600) shows that

$$\beta_l = \tan^{-1}\frac{(X)_l}{r + (R)_l} \quad (601)$$

and therefore that the current of modulation frequency lags behind the envelope of the high-frequency potential by angle β_l. Equation (600) becomes Eq. (591); or

$$[(\Delta^2 E)_l] = (\text{Det. } E)_d \sqrt{2m}(\Delta E)^2 \quad (591)$$

The second harmonic component may be deduced by the same process.

200. Detection with More than One Modulation Frequency.— We have thus far considered the action of a detector only when the high-frequency potential is modulated at a single frequency. If the high-frequency potential is modulated at more than one frequency, the amplitude of the high-frequency potential for two frequencies is expressed by

$$(\text{Amplitude})_h = \widehat{\Delta E}(1 + m' \sin \omega_{l'}t + m'' \sin \omega_{l''}t)$$

If the treatment is carried through with this doubly modulated amplitude, the following components of fictitious voltage result:

$$\left.\begin{aligned}
[\overline{\Delta^2 E}] &= (\text{Det. } E)_d\left(1 + \frac{m'^2}{2} + \frac{m''^2}{2}\right)(\Delta E)^2 \\
[(\Delta^2 E)_{l'}] &= (\text{Det. } E)_d\sqrt{2}m'(\Delta E)^2 \\
[(\Delta^2 E)_{l''}] &= (\text{Det. } E)_d\sqrt{2}m''(\Delta E)^2 \\
[(\Delta^2 E)_{2l'}] &= (\text{Det. } E)_d\frac{m'^2}{2\sqrt{2}}(\Delta E)^2 \\
[(\Delta^2 E)_{2l''}] &= (\text{Det. } E)_d\frac{m''^2}{2\sqrt{2}}(\Delta E)^2 \\
[(\Delta^2 E)_{l'-l''}] &= (\text{Det. } E)_d\frac{m'm''}{\sqrt{2}}(\Delta E)^2 \\
[(\Delta^2 E)_{l'+l''}] &= (\text{Det. } E)_d\frac{m'm''}{\sqrt{2}}(\Delta E)^2
\end{aligned}\right\} \quad (602)$$

The expressions in Eq. (602) show that when there is more than one modulation frequency, there are produced by detection the expected fundamental and second-harmonic components for each modulation frequency. In addition there are components having frequencies equal to the sum and difference of the modulation frequencies. These sum and difference components are more serious in producing distortion than the second-harmonic components, not only because they may not be harmoniously related to any of the original modulation frequencies but also because their amplitude is greater than that of the second-harmonic components if the degrees of modulation are about the same for the two modulation frequencies. The only way to reduce these distorting components is to use low modulation, but that necessarily results in weak signals. If small-signal detection is used,

serious distortion is inevitable if the degree of modulation is much more than about 0.5. In a later chapter it will be shown that less distortion results if the detector is properly adjusted, and is operated with a relatively large impressed voltage.

201. Summary of Important Facts Concerning Small-signal Detection.—Equations (590), (591), and (592) show the following points to which special attention is directed:

1. Small-signal detection depends upon the *square* of the impressed high-frequency potential.

2. Small-signal detection of a sinusoidally modulated high-frequency potential gives rise to three components, namely, a steady current $\overline{\Delta^2 I}$, a current of modulation frequency $(\Delta^2 I)_l$, and a current of twice modulation frequency $(\Delta^2 I)_{2l}$.

3. Each of the three components is proportional to the quantity $(\text{Det. } E)_d$.

4. The ratio of the second harmonic of the modulation frequency to the fundamental is $m/4$.

5. When there is more than one frequency of modulation, sum and difference tones are introduced which are proportional to the product of the degrees of modulation and also to $(\text{Det. } E)_d$.

202. Qualitative Physical Picture of the Action of a Nonlinear Circuit Element.—It is sometimes instructive to think of the action of a nonlinear device, such as a detector, in more general terms. Suppose the current-voltage characteristic of the nonlinear element is represented by a power series of the form

$$\bar{\imath} = A^{\text{I}}\bar{e} + \frac{A^{\text{II}}}{2}\bar{e}^2 + \frac{A^{\text{III}}}{6}\bar{e}^3 + \cdots \tag{603}$$

where the bar over the i and e signifies that they are measured from the quiescent point. Suppose that $\bar{e}$ is made up of potential variations of several frequencies n_1, n_2, n_3, etc., and let n_r, n_s, n_t, and n_u be any four frequencies. The frequencies arising from the frequency n_r due to each of the terms of Eq. (603) are given in Table XIV.

As an example, consider small-signal detection where terms in Eq. (603) above the second degree may be neglected. Suppose that the high-frequency potential is modulated at a single frequency $\omega_l/2\pi$. The modulated potential is represented by

$$\bar{e} = \Delta e = \widehat{\Delta E}_0(1 + m \sin \omega_l t) \sin \omega_h t \tag{604}$$

TABLE XIV.—FREQUENCIES DUE TO EACH OF THE TERMS OF EQ. (603)

Degree of Term	Frequencies
1	n_r
2	0 $2n_r$ $n_r \pm n_s$
3	n_r $3n_r$ $2n_r \pm n_s$ $n_r \pm n_s \pm n_t \pm \cdots$ every combination of signs
4	0 $2n_r$ $4n_r$ $n_r \pm n_s$ $3n_r \pm n_s$ $2(n_r \pm n_s)$ $2n_r \pm n_s \pm n_t$ $n_r \pm n_s \pm n_t \pm n_u \pm \cdots$ every combination of signs

which expands into

$$\Delta e = \widehat{\Delta E}_0 \sin \omega_h t - \frac{\widehat{\Delta E}_0 m}{2} \cos (\omega_h + \omega_l)t + \frac{\widehat{\Delta E}_0 m}{2} \cos (\omega_h - \omega_l)t \quad (605)$$

From Eq. (605) the spectrum of the modulated potential consists of three fine lines, shown in Fig. 266*a*. Figure 266*b* indi-

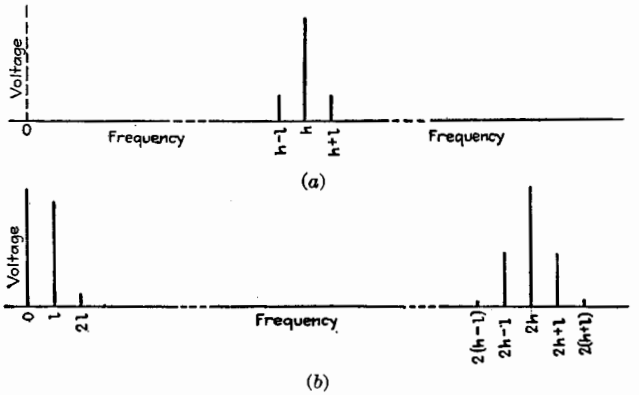


FIG. 266.—(*a*) Frequencies before detection. $m = \frac{1}{2}$. (*b*) Frequencies added by detection.

cates the new frequencies, with an approximate indication of the amplitude at each frequency, which result when the potential of Eq. (604), for $m = \frac{1}{2}$, acts upon a nonlinear element having a parabolic characteristic. These frequencies are given by

Table XIV. The scale of amplitudes is not the same in the three parts of the frequency spectrum of Fig. 266, since the relative amplitudes depend upon the relative values of A^{I} and A^{II} in Eq. (603).

The current of frequency $\omega_l/2\pi$ is the resultant of two currents, one obtained from the carrier potential of frequency $\omega_h/2\pi$ and the upper-side frequency $\frac{\omega_h + \omega_l}{2\pi}$, and the other from the carrier frequency and the lower-side frequency $\frac{\omega_h - \omega_l}{2\pi}$. These two currents combine according to the phase angle between them. If the side-frequency potentials are related to the carrier poten-

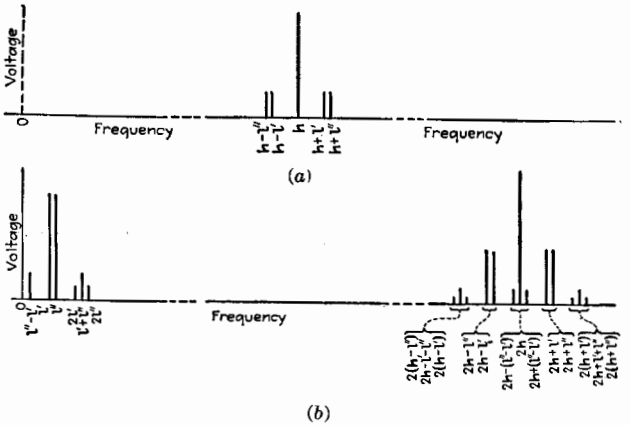


FIG. 267.—(a) Frequencies before detection. $m' = \frac{1}{2}$; $m'' = \frac{1}{2}$. (b) Frequencies added by detection.

tial as in Eq. (605), these two currents are equal and add in the same phase. If some effect shifts the phase of the side frequencies with respect to the carrier wave, these two currents of frequency $\omega_l/2\pi$ are not in phase and may neutralize each other. Such would be the case if the sign of one of the side-frequency terms of Eq. (605) were changed.

Figures 267*a* and *b* are similar to Figs. 266*a* and *b* for the case of two modulation frequencies.

203. Low-frequency Considerations.—Equation (591) gives the fictitious voltage of detection of the component of modulation frequency. Not all of this voltage is available for action upon an output device, such as telephone receivers or an amplifier.

The available voltage depends upon the impedance of Z for the low frequency in comparison with the impedance of the whole circuit. Usually, whatever the output device, it is of advantage to shunt it with a by-pass condenser for the radio-frequency current. By so doing, the radio-frequency voltage ΔE on the detector is a larger fraction of the applied radio-frequency voltage ΔE_0. It would be advantageous to make this shunting condenser large, if it were not for the injurious effect due to the by-passing of some of the currents of the higher audio frequencies by this condenser. The problem of the proper choice of the shunting condenser and its effect upon the ratio of the available voltage to the fictitious voltage demands some study.

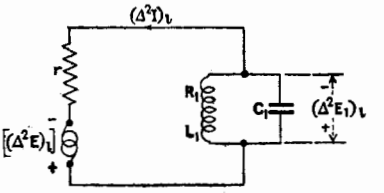


FIG. 268.—Equivalent low-frequency circuit.

Let $(Z)_l$ be made up of the output impedance Z_1, having resistance R_1 and inductance L_1, shunted by a condenser of capacitance C_1. The equivalent circuit for the low-frequency currents is shown in Fig. 268. The output voltage $(\Delta^2 E_1)_l$ is related to the fictitious voltage of detection by

$$\frac{(\Delta^2 E_1)_l}{[(\Delta^2 E)_l]} = \frac{1}{\sqrt{1 + \frac{r(r + 2R_1)}{Z_1^2} + (rC_1\omega_l)^2\left(1 - \frac{2L_1}{C_1 Z_1^2}\right)}} \tag{606}$$

The effect of C_1 upon the effective radio-frequency voltage on the detector will be considered, but Eq. (606) shows that from the point of view of the low-frequency operation, if the term containing C_1 is appreciable compared to unity, the ratio given by Eq. (606) varies with frequency and frequency distortion is present. This indicates that C_1 should be small.

The circuit containing Z_1 is resonant to ω_l when $C_1 Z_1^2 = L_1$. This condition gives the greatest frequency distortion for values of ω_l near the resonant value and gives also the greatest value of the ratio $(\Delta^2 E_1)_l/[(\Delta^2 E)_l]$ as obtained by varying C_1. The last term under the radical in Eq. (606) cannot be more negative than $-(C_1\omega_l r)^2$ and cannot be more positive than $(C_1\omega_l r)^2$. Therefore, if $(C_1\omega_l r)^2$ is small in comparison with unity for the highest audio frequency, the frequency distortion is small.

Hence:

Criterion for small frequency distortion is

$$(rC_1\omega_l)^2 << 1 \tag{607}$$

204. High-frequency Considerations.—The connections shown in Fig. 264 are purely diagrammatic. In this figure, ΔE_0 is the effective series radio-frequency e.m.f. impressed in the system and, as it has no internal impedance, it is assumed to be unaffected by the current flow in the detector system. Z_h is the total series high-frequency impedance external to the detector.

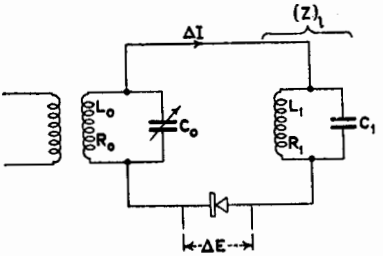


FIG. 269.—High-frequency circuit.

In practice, the usual connections are those shown in Fig. 269, the audio-frequency impedance $(Z)_l$ including a shunting condenser C_1, as explained. A high-frequency e.m.f. $\Delta E_0'$ is either induced in L_0 or otherwise introduced in series with it, L_0

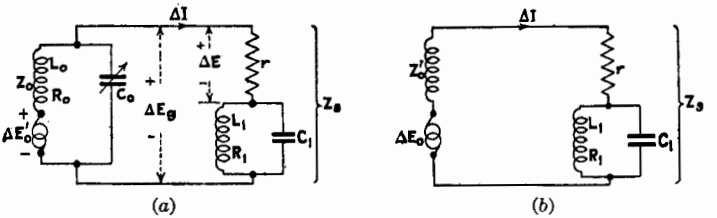


FIG. 270.—(*a*) Equivalent high-frequency circuit. (*b*) Simplified equivalent high-frequency circuit.

being the inductance of the tuned oscillatory circuit. C_0 is the tuning capacitance which is always adjusted to give maximum voltage across the detector. The equivalent circuit for the high-frequency currents is shown in Fig. 270*a*, or in Fig. 270*b* where the series e.m.f. ΔE_0 can be calculated in terms of $\Delta E_0'$, and where $(Z)_h$ is $Z_0' + (Z_1)_h$. $(Z)_h$ can be calculated, but, since the only reason for finding $(Z)_h$ is to determine the ratio of the voltage ΔE across the detector to $\Delta E_0'$, the calculation can be greatly simplified by obtaining this ratio directly as follows: Referring to Fig. 270*a*,

$$\mathbf{\Delta I} = \frac{\mathbf{\Delta E_0'}}{\mathbf{Z_0 + Z_s + jC_0\omega_h Z_0 Z_s}} \tag{608}$$

where Z_s is the impedance of r in series with $(Z_1)_h$.

The value of $C_0\omega_h$ to make ΔI a maximum, and hence to make the voltage across the detector a maximum, is

$$C_0\omega_h = \frac{X_0}{Z_0^2} + \frac{X_s}{Z_s^2} \tag{609}$$

and the maximum value of ΔI is

$$\text{Max. } \Delta I = \frac{\Delta E_0'}{\dfrac{R_0 Z_s}{Z_0} + \dfrac{R_s Z_0}{Z_s}} \tag{610}$$

Since ΔE is equal to $r\Delta I$,

$$\text{Max. } \frac{\Delta E}{\Delta E_0'} = \frac{r}{\dfrac{R_0 Z_s}{Z_0} + \dfrac{R_s Z_0}{Z_s}} \tag{611}$$

If $(Z_1)_h$ is assumed to be $-j/C_1\omega_h$, an assumption which ordinarily introduces very little error since $L_1\omega_h$ is usually very large, Eq. (609) becomes

$$C_0\omega_h = \frac{L_0\omega_h}{Z_0^2} - \frac{C_1\omega_h}{1 + (r_1 C_1\omega_h)^2} \text{ (approx.)} \tag{612}$$

Equation (611) becomes

$$\text{Max. } \left(\frac{\Delta E}{\Delta E_0'}\right)^2 = \frac{1}{\left[\dfrac{R_0}{Z_0} \cdot \dfrac{\sqrt{1 + (rC_1\omega_h)^2}}{rC_1\omega_h} + \dfrac{Z_0}{r} \cdot \dfrac{rC_1\omega_h}{\sqrt{1 + (rC_1\omega_h)^2}}\right]^2} \tag{613}$$

Neglecting R_0^2 in comparison with $L_0^2\omega_h^2$, Eq. (613) becomes

$$\text{Approx. max. } \left(\frac{\Delta E}{\Delta E_0'}\right)^2 = \frac{1}{\left[\eta_0 \dfrac{\sqrt{1 + (rC_1\omega_h)^2}}{rC_1\omega_h} + \dfrac{L_0\omega_h}{r} \dfrac{rC_1\omega_h}{\sqrt{1 + (rC_1\omega_h)^2}}\right]^2} \tag{614}$$

where $\eta_0 = R_0/L_0\omega_h$.

Equation (614) shows the effect of C_1 and of conduction through the detector in reducing the high-frequency voltage ΔE applied to the detector.

Concerning the effect of C_1, it is necessary only that $(rC_1\omega_h)^2$ be large in comparison with unity for C_1 to have a small effect in reducing the high-frequency voltage on the detector. Therefore,

Criterion for C_1 having small effect in reducing high-frequency voltage on the detector is

$$(rC_1\omega_h)^2 >> 1 \tag{615}$$

Therefore C_1 should be chosen considering both the frequency distortion in the low-frequency output and the effect in decreasing the high-frequency voltage.

Examine Eq. (614) for the effect of r upon the ratio max. $\left(\frac{\Delta E}{\Delta E'_0}\right)^2$. If r were infinity or k were zero, the voltage that would then exist across the detector is denoted by $\Delta E_{k=0}$. Then,

$$\left(\frac{\Delta E_{k=0}}{\Delta E'_0}\right)^2 = \frac{1}{\eta_0}.$$

The best way to take account of the effect of r is to obtain the ratio $(\Delta E/\Delta E_{k=0})^2$. This ratio is

$$\left(\frac{\Delta E}{\Delta E_{k=0}}\right)^2 = \frac{1}{\left[\frac{\sqrt{1+(rC_1\omega_h)^2}}{rC_1\omega_h} + \frac{L_0\omega_h}{r\eta_0}\cdot\frac{rC_1\omega_h}{\sqrt{1+(rC_1\omega_h)^2}}\right]^2} \tag{616}$$

If $(rC_1\omega_h)^2$ is large in comparison with unity, Eq. (616) becomes

$$\left(\frac{\Delta E}{\Delta E_{k=0}}\right)^2 = \frac{1}{\left(1+\frac{L_0\omega_h}{r\eta_0}\right)^2} \tag{617}$$

We may now write Eq. (591) in the form

$$[(\Delta^2 E)_l] = (\text{Det. } E)_d \cdot \frac{\sqrt{2}m(\Delta E_{k=0})^2}{\left[\frac{\sqrt{1+(rC_1\omega_h)^2}}{rC_1\omega_h} + \frac{L_0\omega_h}{r\eta_0}\cdot\frac{rC_1\omega_h}{\sqrt{1+(rC_1\omega_h)^2}}\right]^2} \tag{618}$$

$L_0\omega_h/\eta_0$ is approximately the equivalent resistance of the oscillatory circuit at resonance. Equation (618) together with Eq. (616) gives the entire story of the output of a detector at modulation frequency $\omega_l/2\pi$.

*** 205. Comparison of Diode Detectors.**—It is very difficult to compare detectors when their performance depends upon so

many factors. If it is desired to compare detectors all of which are to operate from a given oscillatory circuit, then L_0, R_0, and C_0 are fixed and the approximate factor

$$(\text{Det. } E)_d \frac{1}{\left(1 + \frac{L_0\omega_h}{\eta_0 r}\right)^2} \tag{619}$$

is a fairly satisfactory factor of merit of the detector provided $(Z)_l$ can be so chosen as to make useful a large fraction of the fictitious voltage $[(\Delta^2 E)_l]$.

If, however, the oscillatory circuit is to be designed after a choice of the detector is made, the criterion is to obtain the largest $[(\Delta^2 E)_l]/(\Delta E_0)^2$ possible. This is attained when the factor

$$(\text{Det. } E)_d \frac{1}{(\eta_0 + kL_0\omega_h)^2} \tag{620}$$

is as large as possible, which is true when $kL_0\omega_h$ is small compared with η_0. As the values of η_0 and $L_0\omega_h$ depend so much upon factors of design which cannot be assumed, such as range of frequency, size of wire, type of coil, or size of condenser permissible, it is impossible to give a factor of merit which shall determine the best detector without further inquiry. About the best that can be done is to give the curve of $(\text{Det. } E)_d$ and the curve of k, and then by the use of expression (619) or expression (620), according to circumstances, the best detector can be selected. It is clear that $kL_0\omega_h$ must always be small in comparison with η_0, and, if a reasonable value of η_0 is 0.01 or less, $kL_0\omega_h$ should be at best less than 0.01. This means that if k is very large, $L_0\omega_h$ must be very small and an inconveniently large condenser is necessary.

206. Experimental Determination of Detection Coefficient.—The fictitious voltage $[(\Delta^2 E)_l]$, from which the detection coefficient $(\text{Det. } E)_d$ can be calculated, can be conveniently measured as follows:[3] The method consists briefly in impressing upon the detector a sinusoidally modulated radio-frequency potential of known modulation and in balancing the fictitious voltage of modulation frequency resulting from detection by a potential obtained from a potentiometer which is actuated by the same source of low-frequency potential that produces the modulation.

[3] CHAFFEE, Voltage Detection Coefficient, *Proc. I.R.E.*, **15**, 946 (1927).

The diagram of connections is shown in Fig. 271. The apparatus within the dotted rectangle which indicates an electrostatic shield is a triode oscillator which is modulated at a frequency of 1,000 cycles per second. The triode T_0, which may be a low-power triode such as is used in receiving apparatus, is provided with an oscillatory circuit connected between grid and filament. This circuit contains a small four-terminal resistance R_0, consisting of a short length of fine resistance wire to which are soldered two potential terminals. The resistance R_0 between these potential terminals is accurately known and may be of the order of half an ohm. The oscillatory circuit

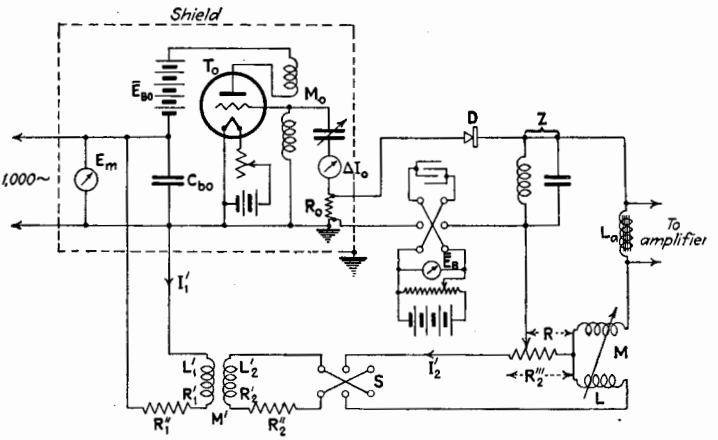


FIG. 271.—Measurement of detection by diode.

also contains a thermoammeter for registering the magnitude ΔI_0 of the oscillatory current, and a variable condenser for setting the frequency of oscillation to any desired value. The plate circuit of the triode T_0 is regeneratively coupled to the oscillatory circuit through the adjustable mutual inductance M_0, and M_0 is made sufficiently large to cause the triode system to oscillate. In order that these oscillations may be modulated, the plate potential of triode T_0 is made pulsating by adding to the steady potential $\bar{E}_{B0}$ an alternating potential E_m of a frequency of 1,000 cycles per second. This alternating potential is derived from a source capable of supplying a potential of from 10 to 20 volts. If the resistance of this source of potential is not low, a low-resistance choke coil should be connected across the leads to allow the steady component of the plate current to pass. C_{bo} is

a condenser having a capacitance sufficient to by-pass the radio-frequency component of the plate current. The alternating potential E_m is measured by a suitable voltmeter.

The detector D under test is connected in a circuit containing R_0, some means of adjusting the polarizing potential $\bar{E}_B$ to any desired value, and an impedance Z which may be given suitable values at both high and low frequencies. The voltmeter $\bar{E}_B$ is preferably shunted by a condenser having a capacitance of several microfarads in order to by-pass currents of both high and low frequencies. The lower terminal of R_0 should be grounded. L_a is a low-resistance choke coil to permit easy passage of the steady component of current in case the amplifier has high resistance.

Detection of the modulated radio-frequency voltage across R_0 produces a voltage of modulation frequency across impedance Z, Fig. 271. This audio-frequency voltage is measured by opposing it, in both phase and magnitude, by a known voltage obtained from an a-c. potentiometer. This potentiometer consists of a variable resistance R and a variable mutual inductance M, through which is passed a known current derived from E_m. This current is fed to the potentiometer through a closely coupled air-core transformer M', and the circuit constants are so adjusted that the current through R and L is in phase with E_m.

The condition under which the potentiometer current is in phase with the modulating voltage E_m is stated below. Consider, apart from the particular circuits of Fig. 271, a primary circuit which has resistance R_1 and inductance L_1, coupled by a mutual inductance M to a secondary circuit having resistance R_2 and inductance L_2. The expression for the secondary current when an e.m.f. of E_1 is impressed in the primary circuit is

$$\begin{aligned}\mathbf{I}_2 &= \frac{jM\omega\mathbf{E}_1}{(R_1 + jL_1\omega)(R_2 + jL_2\omega) + M^2\omega^2} \\ &= \frac{jM\omega\mathbf{E}_1}{(R_1R_2 - L_1L_2\omega^2 + M^2\omega^2) + j(R_1L_2\omega + R_2L_1\omega)}\end{aligned} \tag{621}$$

If the first part of the denominator of the second form of Eq. (621) is made equal to zero, the secondary current is in phase with the e.m.f. E_1 and is given by

$$I_2 = \frac{E_1M}{R_1L_2 + R_2L_1} \tag{622}$$

If this first part equals zero, which is the condition necessary for no phase difference between E_1 and I_2,

$$R_1R_2 = (L_1L_2 - M^2)\omega^2 \tag{623}$$

Applying this theory to the present case, given by the diagram of Fig. 271, the condition to be fulfilled in order that the current I_2' be in phase with E_m is, by Eq. (623),

$$(R_1' + R_1'')(R_2' + R_2'' + R_2''') = [L_1'(L_2' + L) - M'^2]\omega^2 \tag{624}$$

The value of the secondary current I_2' is obtained by applying Eq. (622) to the present case. Then,

$$I_2' = \frac{M'E_m}{(R_1' + R_1'')(L_2' + L) + (R_2' + R_2'' + R_2''')L_1'} \tag{625}$$

Although the proper values of the resistances of the circuit can be determined by calculation from Eq. (624), it is usually

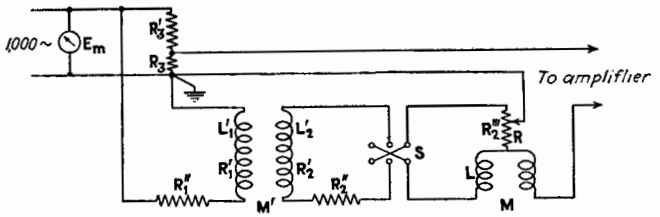


FIG. 272.—Calibration of circuits of Fig. 271.

easier to determine their proper values by a separate experiment as follows: Referring to Fig. 272, two resistances, R_3 and R_3', are temporarily connected in series across E_m and the junction between the two resistances is connected to the terminal of the amplifier. R_3 may be conveniently one one-thousandth of R_3'. Adjust R_2'' and R_1'' so that, with M equal to zero, balance is obtained with some value of R. Let R' represent this value of R. The value of R' gives the number of ohms which introduces a voltage equal to $\frac{E_mR_3}{R_3+R_3'}$ into the balancing circuit. It is usually desirable to make R_2'' about 10,000 ohms, so that the variation of R in normal operation causes only a small percentage variation in the total secondary resistance. R is shown in the diagram as an ordinary resistance box connected as a potenti-

ometer on the high-dial end. In the experiments to be described, R was a three-dial box having dials of tenths, units, and tens of ohms. The connections were changed so that all of the 10-ohm coils were in the secondary circuit, the switch arm being represented by the arrow in the diagram. The total maximum change of resistance due to varying the two lower dials was only 11 ohms.

Before measurements of the detection coefficient can be made, the oscillator T_0 must be adjusted to give sinusoidal modulation. The oscillator is tested by shutting off the source of the alternating voltage E_m and reading ΔI_0 for various values of $\bar{E}_B'$. The value of coupling M_0 and the grid-polarizing potential of the oscillator should be so chosen that the plot of ΔI_0 against $\bar{E}_B'$ is straight over a considerable range of $\bar{E}_B'$, as shown in Fig. 273. A value of $\bar{E}_B'$ is chosen corresponding to a point about midway on the straight portion of the plot. The peak value of E_m, or $\sqrt{2}E_m$, is so chosen as to vary the plate voltage over the straight-line portion of the plot. The degree of modulation m is given by the ratio $\overline{ab}/\overline{ob}$, where $\overline{ob}$, Fig. 273, is equal to ΔI_0, the r.m.s. value of the unmodulated radio current, and $\overline{oa}$ is the minimum value of the r.m.s. radio-frequency current when it is modulated. It is assumed that the amplitude of the radio-frequency current changes, as the plate voltage of the oscillator is varied at 1,000 cycles, according to the plot of Fig. 273 which was obtained by a slow variation of plate voltage.

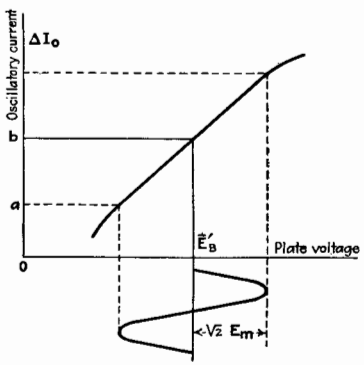


FIG. 273.—Characteristic of oscillator.

The value of ΔE to be used in Eq. (591) is $(\Delta I_0)R_0$, provided the impedance $(Z)_h$ is small in comparison with r of the detector. ΔI_0 is the oscillatory current with no modulation. When E_m is impressed, the oscillatory current increases to $\Delta I_0\sqrt{1 + \frac{m^2}{2}}$.

The voltage, which balances the audio voltage across $(Z)_l$ resulting from detection, is

$$(\Delta^2 E)_l = \frac{E_m R_3}{R_3 + R_3'} \cdot \frac{\sqrt{R^2 + M^2\omega^2}}{R'}. \qquad (626)$$

Its phase with respect to E_m is

$$\theta = \pm \tan^{-1} \frac{M\omega}{R}, \tag{627}$$

the sign of the angle depending upon the position of switch S and the sign of M.

If $(Z)_l$ is made infinite, Eq. (626) gives the fictitious voltage of detection $[(\Delta^2 E)_l]$ of Eq. (591).

Before measurements begin, certain tests should be made to make sure that the system is properly assembled with no serious coupling between circuits. The first test is to turn off

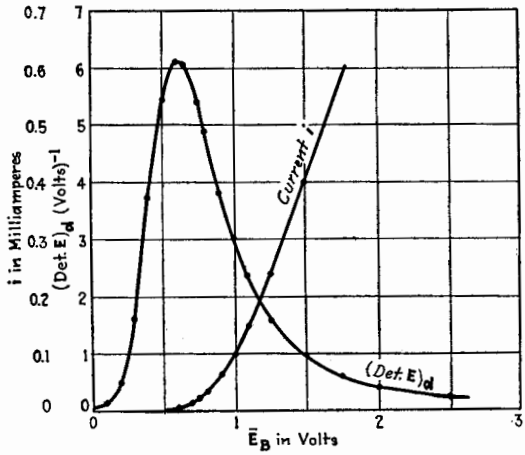


FIG. 274.—Current characteristic and detection coefficient of a carborundum crystal contact. $m = 0.50$; $\Delta E = 0.0388$ volt. Crystal positive, metal point negative.

the filament of the oscillator tube T_0 but to allow E_m to act. Then, with R and M set at zero and Z short-circuited, there should be no sound in the telephone receivers for both positions of S. Sound under these conditions is good evidence of direct coupling between the 1,000-cycle source and the amplifier. When this test is satisfactory, $(Z)_l$ may be made very large and there still should be no 1,000-cycle sound in the telephone receivers. The third test consists in disconnecting from R_0 the upper wire which goes to the detector and connecting it to the lower or grounded wire connecting to R_0. With the modulated high-frequency currents flowing through R_0, and with R and M set at zero, there should be no sound in the telephone

receivers. Sound under these conditions indicates that some radio-frequency voltage is being introduced into the detector circuit owing to insufficient shielding.

207. Tests of Certain Crystal Detectors.—This section describes tests made on certain crystal-contact detectors by the method given in the preceding section. In all of these tests the series polarizing voltage is taken as positive when the crystal is positive with respect to the contacting metallic point.

Figure 274 shows a graph of the current i and the voltage detection coefficient plotted against the impressed polarizing voltage

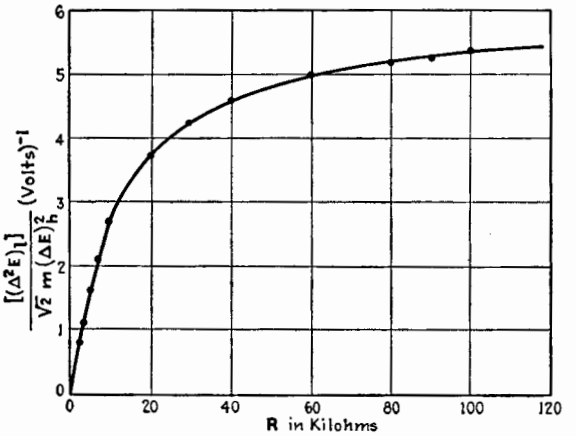


FIG. 275.—Variation of voltage of detection across series resistance R as R is varied. Carborundum detector. $E_B = 0.62$ volt. Crystal positive.

$\bar{E}_B$ for a carborundum crystal. The degree of modulation was 0.50, and the r.m.s. value of the unmodulated high-frequency voltage impressed on the detector was 0.0388 volt.

Figure 275 gives as abscissa the fraction $\frac{(\Delta^2 E)_l}{\sqrt{2}m(\Delta E)_h^2}$ measured across the load resistance R in kilohms plotted when the polarizing potential on the detector was held constant at 0.62 volt, the value which gave a maximum detection in Fig. 274. If R is infinite, the above fraction becomes (Det. $E)_d$. The plot of Fig. 275 shows that the resistance of the detector at this particular polarizing voltage was about 10,000 ohms.

Figure 276 gives $i = f(e)$ and (Det. $E)_d$ for a contact with an iron-pyrites crystal. Maximum detection voltage occurred for

a polarizing voltage of -0.25 volt, and the resistance of the detector was about 3,000 ohms at that point.

Finally, Fig. 277 shows a plot of $i = f(e)$ and (Det. $E)_d$ for a galena detector. The maximum detection coefficient occurred at $\bar{E}_B = 0.05$ volt, at which point the resistance of the detector was about 15,000 ohms.

General References

COLEBROOK: Rectifying Detector, *Exp. Wireless*, **2**, 330, 394, 459 (1925).

CHAPTER XX

SMALL-SIGNAL DETECTION BY TRIODE

The triode can be used as a detector in two ways, according as the nonlinear characteristic used is in the plate circuit or in the grid circuit. When the curvature of the plate-current characteristic is used, the detection is known as *plate-circuit detection*, and when the curvature of the grid-current characteristic is used, the resulting detection is called *grid-circuit detection*. Although the adjustments of the triode may be such that both types of detection take place simultaneously, it is usually best practice to adjust for either one or the other type alone. For this reason the two types of detection will be considered separately.

The general theory to be given follows the first method given in the preceding chapter. The same results may be obtained by the application of the second method of that chapter.

208. General Theory of Detection by Triode. Unmodulated Signal.—Assume that the instantaneous plate current i_p is a function of the instantaneous plate potential e_p and also of the instantaneous grid potential e_g, expressed by

$$i_p = F(e_p, e_g) \tag{628}$$

Similarly, the instantaneous grid current i_g is a function of both e_p and e_g, but, since the dependence of i_g upon e_p is only slight for small values of i_g, we shall assume that i_g is expressible in terms of e_g alone, according to the expression

$$i_g = f(e_g) \tag{629}$$

The diagram of connections is given in Fig. 278 where Z_b and Z_c are the equivalent series impedances of the plate and grid circuits and may have different values at different frequencies.

Assume that the small unmodulated alternating e.m.f. $\Delta e_0 = \widehat{\Delta E}_0 \sin \omega_h t$ is impressed in the grid circuit of the detector of Fig. 278. Because both characteristic graphs, Eqs. (628) and (629), are curves, the changes of grid and plate voltages, due to the impressed e.m.f. Δe_0, are made up of steady rectified

components as well as of variations at frequency $\omega_h/2\pi$. We may designate the increase of e_g and e_p due to $\Delta\widehat{E}_0$ as $\Delta\widehat{E}'_g$ and $\Delta\widehat{E}'_p$, and the decrease in the same quantities due to $-\Delta\widehat{E}_0$ as $\Delta\widehat{E}''_g$ and $\Delta\widehat{E}''_p$. Using Taylor's theorem, we have for the increase of plate current

$$\begin{aligned}\bar{I}_p + \Delta\widehat{I}'_p &= F(\bar{E}_p + \Delta\widehat{E}'_p, \bar{E}_g + \Delta\widehat{E}'_g)\\ &= F(\bar{E}_p, \bar{E}_g) + \frac{\partial i_p}{\partial e_p}\Delta\widehat{E}'_p + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_p^2}(\Delta\widehat{E}'_p)^2 + \cdots\\ &\quad + \frac{\partial i_p}{\partial e_g}\Delta\widehat{E}'_g + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_g^2}(\Delta\widehat{E}'_g)^2 + \cdots\\ &\quad + \frac{\partial^2 i_p}{\partial e_p \partial e_g}\Delta\widehat{E}'_p\Delta\widehat{E}'_g + \cdots \qquad (630)\end{aligned}$$

Similarly, for the decrease in plate current,

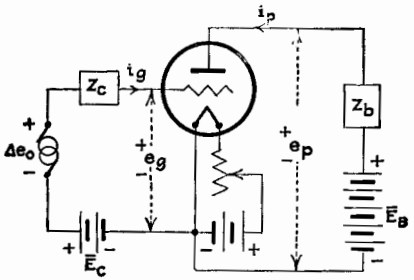


FIG. 278.—The triode as detector.

$$\begin{aligned}\bar{I}_p - \Delta\widehat{I}''_p &= F(\bar{E}_p - \Delta\widehat{E}''_p, \bar{E}_g - \Delta\widehat{E}''_g)\\ &= F(\bar{E}_p, \bar{E}_g) - \frac{\partial i_p}{\partial e_p}\Delta\widehat{E}''_p + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_p^2}(\Delta\widehat{E}''_p)^2 - \cdots\\ &\quad - \frac{\partial i_p}{\partial e_g}\Delta\widehat{E}''_g + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_g^2}(\Delta\widehat{E}''_g)^2 - \cdots\\ &\quad + \frac{\partial^2 i_p}{\partial e_p \partial e_g}\Delta\widehat{E}''_p\Delta\widehat{E}''_g - \cdots \qquad (631)\end{aligned}$$

Adding Eqs. (630) and (631),

$$\begin{aligned}\Delta\widehat{I}'_p - \Delta\widehat{I}''_p &= \frac{\partial i_p}{\partial e_p}(\Delta\widehat{E}'_p - \Delta\widehat{E}''_p) + \frac{\partial i_p}{\partial e_g}(\Delta\widehat{E}'_g - \Delta\widehat{E}''_g)\\ &\quad + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_p^2}[(\Delta\widehat{E}'_p)^2 + (\Delta\widehat{E}''_p)^2] + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_g^2}[(\Delta\widehat{E}'_g)^2 + (\Delta\widehat{E}''_g)^2]\\ &\quad + \frac{\partial^2 i_p}{\partial e_p \partial e_g}[\Delta\widehat{E}'_p\Delta\widehat{E}'_g + \Delta\widehat{E}''_p\Delta\widehat{E}''_g] + \cdots \qquad (632)\end{aligned}$$

In Eqs. (630), (631), and (632), higher-order terms are neglected.

In writing Eqs. (630) and (631), it is implicitly assumed that the increments of e_p and e_g take place simultaneously. This restriction means that Z_b must have no appreciable reactance in comparison with r_p at the frequency $\omega_h/2\pi$. This restriction is not so serious as might at first appear, for usually the plate load is shunted by a condenser which has so low a reactance as to be negligible for the radio-frequency currents. In such a case $(Z_b)_h$ is practically equal to zero.

Equation (632) can be simplified, since the first three terms are expressible as constant values, and the average of the squares of the increments of the plate and grid voltages is approximately equal to the squares of the maximum values of the radio-frequency components of the plate and grid voltages.

Equation (632) then becomes

$$\overline{\Delta^2 I_p} = \frac{\partial i_p}{\partial e_p}\overline{\Delta^2 E_p} + \frac{\partial i_p}{\partial e_g}\overline{\Delta^2 E_g} + \frac{1}{4}\frac{\partial^2 i_p}{\partial e_p^2}(\widehat{\Delta E}_p)^2 + \frac{1}{4}\frac{\partial^2 i_p}{\partial e_g^2}(\widehat{\Delta E}_g)^2 + \frac{1}{2}\frac{\partial^2 i_p}{\partial e_p \partial e_g}\widehat{\Delta E}_p \widehat{\Delta E}_g + \cdots \quad (633)$$

A further simplification of Eq. (633) is afforded by substituting the tube coefficients k_p and s_p for the equivalent derivatives. Rewriting Eq. (633),

$$\overline{\Delta^2 I_p} = \frac{\overline{\Delta^2 E_p}}{r_p} + s_p\overline{\Delta^2 E_g} + \frac{1}{4}\frac{\partial k_p}{\partial e_p}(\widehat{\Delta E}_p)^2 + \frac{1}{4}\frac{\partial s_p}{\partial e_g}(\widehat{\Delta E}_g)^2 + \frac{1}{2}\frac{\partial s_p}{\partial e_p}(\widehat{\Delta E}_p \widehat{\Delta E}_g) + \cdots \quad (634)$$

In Eq. (634), $\widehat{\Delta E}_p$ is the maximum amplitude of the radio-frequency alternating potential between the plate and filament. The plate current of frequency $\omega_h/2\pi$ is

$$\widehat{\Delta I}_p = \frac{u_p \widehat{\Delta E}_g}{r_p + (R_b)_h} \quad (635)$$

where $(R_b)_h$ is the resistance of the plate impedance Z_b at frequency $\omega_h/2\pi$, and $(X_b)_h$ is zero. Multiplying Eq. (635) by $-(R_b)_h$,

$$\widehat{\Delta E}_p = -\frac{u_p (R_b)_h \widehat{\Delta E}_g}{r_p + (R_b)_h} \quad (636)$$

Substituting Eq. (636) in Eq. (634),

$$\overline{\Delta^2 I}_p = \frac{\overline{\Delta^2 E}_p}{r_p} + s_p\overline{\Delta^2 E}_g + \left[\frac{1}{4}\cdot\frac{\partial s_p}{\partial e_g} + \frac{1}{4}\cdot\frac{u_p^2(R_b)_h^2}{[r_p + (R_b)_h]^2}\cdot\frac{\partial k_p}{\partial e_p} - \frac{1}{2}\cdot\frac{u_p(R_b)_h}{r_p + (R_b)_h}\cdot\frac{\partial s_p}{\partial e_p}\right](\Delta E_g)^2 \quad (637)$$

Following the same method for the increments in grid current,

$$\Delta\widehat{I}'_g - \Delta\widehat{I}''_g = \frac{\partial i_g}{\partial e_g}(\Delta\widehat{E}'_g - \Delta\widehat{E}''_g) + \frac{1}{2}\cdot\frac{\partial^2 i_g}{\partial e_g^2}[(\Delta\widehat{E}'_g)^2 + (\Delta\widehat{E}''_g)^2] + \cdots \quad (638)$$

Replacing $\partial i_g/\partial e_g$ by the grid variational conductance $k_g = 1/r_g$, Eq. (638) may be written

$$\overline{\Delta^2 I}_g = \frac{\overline{\Delta^2 E}_g}{r_g} + \frac{1}{4}\cdot\frac{\partial k_g}{\partial e_g}(\Delta\widehat{E}_g)^2 + \cdots \quad (639)$$

Also

$$\overline{\Delta^2 I}_g = -\frac{\overline{\Delta^2 E}_g}{\bar{R}_c} \quad (640)$$

and

$$\overline{\Delta^2 E}_p = -\bar{R}_b\overline{\Delta^2 I}_p \quad (641)$$

Combining Eqs. (637), (639), (640), and (641) to eliminate $\overline{\Delta^2 E}_g$, $\overline{\Delta^2 E}_p$, and $\overline{\Delta^2 I}_g$, we have

$$(r_p + \bar{R}_b)\overline{\Delta^2 I}_p = -\frac{u_p r_g \bar{R}_c}{r_g + \bar{R}_c}\cdot\frac{1}{2}\cdot\frac{\partial k_g}{\partial e_g}(\Delta E_g)^2 + \frac{r_p}{2}\left[\frac{\partial s_p}{\partial e_g} + \frac{u_p^2(R_b)_h^2}{[r_p + (R_b)_h]^2}\cdot\frac{\partial k_p}{\partial e_p} - \frac{2u_p(R_b)_h}{r_p + (R_b)_h}\cdot\frac{\partial s_p}{\partial e_p}\right](\Delta E_g)^2 \quad (642)$$

The left-hand member of Eq. (642) is the *fictitious* voltage of detection in the plate circuit, denoted by $[\overline{\Delta^2 E}_p]$. The first term of the right-hand member of Eq. (642) is the contribution to the fictitious voltage of detection due to detection in the grid circuit. The remainder of the right-hand side of Eq. (642) is the contribution to $\overline{\Delta^2 E}_p$ due to detection in the plate circuit. We may separate the two kinds of detection as follows:

Grid-circuit Detection:

$$[\overline{\Delta^2 E}_p] = -\frac{u_p\bar{R}_c}{r_g + \bar{R}_c}\cdot\frac{1}{2k_g}\cdot\frac{\partial k_g}{\partial e_g}(\Delta E_g)^2 \quad (643)$$

Plate-circuit Detection:

$$[\overline{\Delta^2 E_p}] = \left[\frac{u_p}{2s_p}\cdot\frac{\partial s_p}{\partial e_g} + \frac{u_p^2(R_b)_h^2}{[r_p + (R_b)_h]^2}\cdot\frac{1}{2k_p}\cdot\frac{\partial k_p}{\partial e_p} - \frac{2u_p^2(R_b)_h}{r_p + (R_b)_h}\cdot \frac{1}{2s_p}\cdot\frac{\partial s_p}{\partial e_p}\right](\Delta E_g)^2 \quad (644)$$

In Eqs. (643) and (644), ΔE_g is the r.m.s. value of the high-frequency potential which exists at the grid. This may be less than ΔE_0 by the drop in impedance $(Z_c)_h$, or more than ΔE_0 if the latter is the voltage induced in the coil of an oscillatory circuit connected in the grid circuit of the detector. In any case, the method of calculation of ΔE_g from ΔE_0 is fully given in Chap. XIX and will not be repeated here. It must be remembered, however, that the *equivalent input admittance* of the triode must be used in place of k_g.

209. General Theory of Detection by Triode. Modulated Signal.—We have assumed that the impressed alternating e.m.f. ΔE_0 has a constant amplitude $\Delta\widehat{E}_0$. Following the method adopted in Chap. XIX, suppose that the impressed e.m.f. of frequency $\omega_h/2\pi$ is modulated at a frequency $\omega_l/2\pi$ as indicated by

$$\Delta e_0 = \Delta\widehat{E}_0(1 + m \sin \omega_l t) \sin \omega_h t \quad (645)$$

The derivation just given for no modulation is applicable to the present problem up to and including Eqs. (637) and (639), except that $\Delta\widehat{E}_g$ is to be replaced by the expression for the varying amplitude, *i.e.*, by

$$\Delta\widehat{E}_g(1 + m \sin \omega_l t)$$

When this is substituted in Eqs. (637) and (639), the quantities $\overline{\Delta^2 I_p}$, $\overline{\Delta^2 I_g}$, $\overline{\Delta^2 E_p}$, and $\overline{\Delta^2 E_g}$ become pulsating quantities, each of which is composed of three components, a steady component, a component of modulation frequency, and a component of twice modulation frequency. The new forms of Eqs. (637) and (639) are:

$$\overline{\Delta^2 I_p} = \frac{\overline{\Delta^2 E_p}}{r_p} + s_p\overline{\Delta^2 E_g} + B\left(1 + \frac{m^2}{2}\right)(\Delta\widehat{E}_g)^2 \quad (646)$$

$$(\boldsymbol{\Delta}^2\mathbf{I}_p)_l = \frac{(\boldsymbol{\Delta}^2\mathbf{E}_p)_l}{r_p} + s_p(\boldsymbol{\Delta}^2\mathbf{E}_g)_l + B\sqrt{2}m(\Delta\widehat{E}_g)^2 \quad (647)$$

$$(\boldsymbol{\Delta}^2\mathbf{I}_p)_{2l} = \frac{(\boldsymbol{\Delta}^2\mathbf{E}_p)_{2l}}{r_p} + s_p(\boldsymbol{\Delta}^2\mathbf{E}_g)_{2l} + B\frac{m^2}{2\sqrt{2}}(\Delta\widehat{E}_g)^2 \quad (648)$$

where B stands for the bracket in Eq. (637).

$$\overline{\Delta^2 I_g} = \frac{\overline{\Delta^2 E_g}}{r_g} + \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g}\left(1 + \frac{m^2}{2}\right)(\Delta \widehat{E}_g)^2 \tag{649}$$

$$(\mathbf{\Delta}^2 \mathbf{I}_g)_l = \frac{(\mathbf{\Delta}^2 \mathbf{E}_g)_l}{r_g} + \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g} \sqrt{2} m (\Delta \widehat{E}_g)^2 \tag{650}$$

$$(\mathbf{\Delta}^2 \mathbf{I}_g)_{2l} = \frac{(\mathbf{\Delta}^2 \mathbf{E}_g)_{2l}}{r_g} + \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g} \cdot \frac{m^2}{2\sqrt{2}} (\Delta E_g)^2 \tag{651}$$

In addition to Eqs. (640) and (641), we have

$$(\mathbf{\Delta}^2 \mathbf{I}_g)_l = -\frac{(\mathbf{\Delta}^2 \mathbf{E}_g)_l}{(\mathbf{Z}_c)_l} \tag{652}$$

$$(\mathbf{\Delta}^2 \mathbf{I}_g)_{2l} = -\frac{(\mathbf{\Delta}^2 \mathbf{E}_g)_{2l}}{(\mathbf{Z}_c)_{2l}} \tag{653}$$

$$(\mathbf{\Delta}^2 \mathbf{E}_p)_l = -(\mathbf{Z}_o)_l (\mathbf{\Delta}^2 \mathbf{I}_p)_l \tag{654}$$

$$(\mathbf{\Delta}^2 \mathbf{E}_p)_{2l} = -(\mathbf{Z}_b)_{2l} (\mathbf{\Delta}^2 \mathbf{I}_p)_{2l} \tag{655}$$

Combining, as before, Eqs. (640), (641), (646), and (649); also Eqs. (647), (650), (652), and (654); and (648), (651), (653), and (655), gives the following three expressions for the three components of plate current due to detection:

$$(r_p + \bar{R}_b)\overline{\Delta^2 I}_p = -\frac{u_p r_g \bar{R}_c}{r_g + \bar{R}_c} \cdot \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g}\left(1 + \frac{m^2}{2}\right)(\Delta \widehat{E}_g)^2 + r_p B\left(1 + \frac{m^2}{2}\right)(\Delta \widehat{E}_g)^2 \tag{656}$$

$$[r_p + (\mathbf{Z}_b)_l](\mathbf{\Delta}^2 \mathbf{I}_p)_l = -\frac{u_p r_g (\mathbf{Z}_c)_l}{r_g + (\mathbf{Z}_c)_l} \cdot \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g} \sqrt{2} m (\Delta \widehat{E}_g)^2 + r_p B \sqrt{2} m (\Delta \widehat{E}_g)^2 \tag{657}$$

$$[r_p + (\mathbf{Z}_b)_{2l}](\mathbf{\Delta}^2 \mathbf{I}_p)_{2l} = -\frac{u_p r_g (\mathbf{Z}_c)_{2l}}{r_g + (\mathbf{Z}_c)_{2l}} \cdot \frac{1}{4} \cdot \frac{\partial k_g}{\partial e_g} \cdot \frac{m^2}{2\sqrt{2}} (\Delta \widehat{E}_g)^2 + r_p B \frac{m^2}{2\sqrt{2}} (\Delta \widehat{E}_g)^2 \tag{658}$$

Equations (656), (657), and (658) are similar to Eq. (642) in that the first term of the right-hand side in each equation is the contribution to the fictitious voltage in the plate circuit due to detection in the grid circuit, and the remainder of each equation gives the detection in the plate circuit. The equations can be separated as before into equations giving the separate detection effects in the two circuits. Certain evident rearrange-

ments are made in writing the following final equations for detection in a triode.

Grid-circuit Detection:

$$[\overline{\Delta^2 E_p}] = -\frac{\bar{R}_c}{r_g + \bar{R}_c} \cdot (\text{Det. } E)_g \left(1 + \frac{m^2}{2}\right)(\Delta E_g)_h^2 \quad (659)$$

$$[(\boldsymbol{\Delta}^2 \mathbf{E}_p)_l] = -\frac{(\mathbf{Z}_c)_l}{r_g + (\mathbf{Z}_c)_l} \cdot (\text{Det. } E)_g \sqrt{2} m (\Delta E_g)_h^2 \quad (660)$$

$$[(\boldsymbol{\Delta}^2 \mathbf{E}_p)_{2l}] = -\frac{(\mathbf{Z}_c)_{2l}}{r_g + (\mathbf{Z}_c)_{2l}} \cdot (\text{Det. } E)_g \frac{m^2}{2\sqrt{2}} (\Delta E_g)_h^2 \quad (661)$$

$$\text{where } (\text{Det. } E)_g = \frac{u_p}{2k_g} \cdot \frac{\partial k_g}{\partial e_g} \quad (662)$$

Plate-circuit Detection:

$$[\overline{\Delta^2 E_p}] = (\text{Det. } E)_p \left(1 + \frac{m^2}{2}\right)(\Delta E_g)_h^2 \quad (663)$$

$$[(\Delta^2 E_p)_l] = (\text{Det. } E)_p \sqrt{2} m (\Delta E_g)_h^2 \quad (664)$$

$$[(\Delta^2 E_p)_{2l}] = (\text{Det. } E)_p \frac{m^2}{2\sqrt{2}} (\Delta E_g)_h^2 \quad (665)$$

where

$$(\text{Det. } E)_p = 2r_p B = u_p \left[\frac{1}{2s_p} \cdot \frac{\partial s_p}{\partial e_g} + \frac{u_p (R_b)_h^2}{[r_p + (R_b)_h]^2} \cdot \frac{1}{2k_p} \cdot \frac{\partial k_p}{\partial e_p} - \frac{2u_p (R_b)_h}{r_p + (R_b)_h} \cdot \frac{1}{2s_p} \cdot \frac{\partial s_p}{\partial e_p}\right] \quad (666)$$

Examine the equations for grid-circuit detection. The detection coefficient given in Eq. (662) is similar to that given in Eq. (577), page 489, for a diode. It is the factor which depends only upon the characteristics of the triode. This factor excluding u_p determines the amount of rectification which occurs in the nonlinear grid circuit. The fraction $\frac{(\mathbf{Z}_c)_l}{r_g + (\mathbf{Z}_c)_l}$ in Eq. (660) gives the fraction of the fictitious voltage of detection in the grid circuit which exists from grid to filament. This grid voltage multiplied by u_p, occurring in Eq. (662), gives the effective series fictitious voltage in the plate circuit. This fictitious voltage in the plate circuit is shifted in phase by the reactance of $(\mathbf{Z}_c)_l$. The negative sign means a 180-deg. phase shift with respect to the modulation voltage.

The fictitious voltage in the plate circuit due to plate-circuit detection, given by Eq. (664), is in phase with the modulation

or with the envelope of the modulated high-frequency voltage. The detection coefficient given by Eq. (666), although more complicated than previous coefficients, contains terms each of which is of the same form as the expression for the detection coefficient of a diode. If $(Z_b)_h$ is zero, (Det. $E)_p$ reduces to the simple form

$$(\text{Det. } E)_p = \frac{u_p}{2s_p} \cdot \frac{\partial s_p}{\partial e_g} \text{ (when } (Z_b)_h = 0) \tag{667}$$

The bracket in Eq. (666) can be very much simplified if the grid-polarizing potential is negative, if the triode contains no gas or vapor, and if u_p is approximately constant. We may then assume that

$$i_p = F(e_p + u_p e_g)$$

Hence

$$\frac{u_p}{k_p} \cdot \frac{\partial k_p}{\partial e_p} = \frac{1}{s_p} \cdot \frac{\partial s_p}{\partial e_g} \tag{668}$$

and

$$\frac{u_p}{s_p} \cdot \frac{\partial s_p}{\partial e_p} = \frac{1}{s_p} \cdot \frac{\partial s_p}{\partial e_g} \tag{669}$$

Substituting Eqs. (668) and (669) in Eq. (666) gives, as an approximate value of (Det. $E)_p$ when $\bar{E}_c$ is negative,

$$(\text{Det. } E)_p = u_p\left(1 - \frac{(R_b)_h}{r_p + (R_b)_h}\right)^2 \cdot \frac{1}{2s_p} \cdot \frac{\partial s_p}{\partial e_g} \tag{670}$$

$$= u_p^2\left(\frac{r_p}{r_p + (R_b)_h}\right)^2 \cdot \frac{1}{2k_p} \cdot \frac{\partial k_p}{\partial e_p} \tag{671}$$

when u_p = constant

$i_g = 0$

$i_p = F(e_p + u_p e_g)$

Expression (671) is of interest in that it refers all quantities to the plate circuit. The quantity

$$\left(\frac{r_p}{r_p + (R_b)_h}\right) u_p \Delta E_g = [(\Delta E_r)_h] \tag{672}$$

is the fictitious high-frequency voltage that may be considered as acting across the nonlinear element r_p in the circuit shown in Fig. 279. This conception reduces plate detection to the case

of an equivalent diode, and the voltage of detection may be written

$$[(\Delta^2 E_p)_l] = \frac{1}{2k_p} \cdot \frac{\partial k_p}{\partial e_p} \sqrt{2} m (\Delta E_r)_h^2 \tag{673}$$

210. Grid-circuit Detection. Discussion.—In practice we are usually interested only in those results of detection which have a frequency equal to the modulation frequency $\omega_l/2\pi$. The fictitious voltage of this frequency in the plate circuit, resulting from grid-circuit detection, is given by Eq. (660).

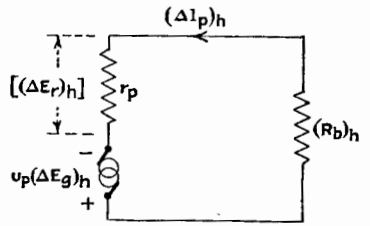


FIG. 279.—Equivalent plate circuit for the high-frequency currents.

The connections generally adopted for grid-circuit detection are shown in Fig. 280, where $(Z_c)_l$ is shown as a resistance R_c, commonly called the *grid leak*, shunted by a condenser C_c.

Any residual plate-circuit detection due to the positive curvature at the lower parts of the plate-current curves is opposite in sign to the detection due to the grid circuit. The only plate-circuit detection which could aid grid-circuit detection is that

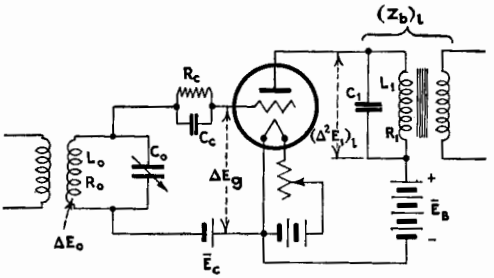


FIG. 280.—Circuits of the triode arranged for grid-circuit detection.

obtained when the plate-current curves have negative curvature, as when saturation is approached. With modern tubes, this negative curvature is very small because of plentiful emission and because saturation is ill defined with oxide-coated and thoriated filaments. Therefore, when using grid-circuit detection, the adjustments are such as practically to eliminate plate-

circuit detection; in other words, the plate voltage $\bar{E}_B$ should be such as to place the operating point on a straight part of the plate-current characteristic.

Grid-circuit detection should be viewed as simple diode detection in the grid circuit, causing an audio-frequency voltage on the grid. This audio-frequency voltage is then amplified through the plate system of the triode. The grid circuit is therefore adjusted for best detection and the plate circuit for best amplification.

Considering the grid circuit as the circuit of a diode detector, C_c corresponds to C_1 of Fig. 269, page 500, Chap. XIX, and R_c to R_1 and L_1 of that figure. The general method of Chap. XIX may be followed in determining the effect of C_c in reducing the radio-frequency voltage on the detector, and in determining its effect in producing frequency distortion within the low-frequency range. Two differences should be noted, however. First, the triode detector has an impedance given by the expressions for input admittance of Chap. XI, whereas the diode of Chap. XIX was assumed to have an impedance equal to r. Second, the diode detector has only one circuit in which frequency distortion may arise, whereas the triode detector has a second circuit, the plate circuit, in which frequency distortion may arise.

Examine the effects of the impedances in the grid and plate circuits upon both the high- and low-frequency operation.

First, consider the plate circuit. The plate load for audio frequency is often shunted by a by-pass condenser, shown as C_1, Fig. 280. This by-pass condenser is used to reduce the radio-frequency voltage on the plate, but, if the plate-circuit operating point is properly fixed on a plane portion of the characteristic surface, this radio-frequency plate voltage does neither harm nor good. However, condenser C_1 may be harmful in by-passing some of the audio currents of the higher frequencies, thereby causing frequency distortion. Equation (606), copied as Eq. (674), with a change of letters to apply to the present case, should be the principal guide in the choice of the size of condenser C_1 for grid-circuit detection.

$$\frac{(\Delta^2 E_1)_l}{[(\Delta^2 E_p)_l]} = \frac{1}{\sqrt{1 + \frac{r_p(r_p + 2R_1)}{Z_1^2} + (r_p C_1 \omega_l)^2 \left(1 - \frac{2X_1}{C_1 \omega_l Z_1^2}\right)}} \tag{674}$$

From Eq. (674), we derive as a criterion for small frequency distortion introduced in the plate circuit,

$$(r_p C_1 \omega_l)^2 << 1 \text{ (for small frequency distortion)} \qquad (675)$$

Considering the grid circuit, the radio-frequency grid voltage ΔE_g is subject to some loss in the capacitance C_c. This capacitance is solely for the purpose of providing a by-pass for the radio-frequency currents. The square of the ratio of the radio-frequency voltage on the grid to that across the oscillatory circuit is

$$\left(\frac{\Delta E_g}{\Delta E_s}\right)^2 = \frac{1 + (R_c C_c \omega_h)^2}{\left(1 + \frac{R_c}{\check{r}_g}\right)^2 + R_c^2 (C_c + \check{C}_g)^2 \omega_h^2} \qquad (676)$$

where $\check{r}_g = 1/g_g$ and $\check{C}_g = -b_g/\omega$ are the equivalent input resistance and capacitance of the tube as given by the formulas of Chap. XI. If the condenser C_1 across the plate load is large in comparison with C_{pg} and C_{pf}, $\check{r}_g$ is very nearly equal to r_g and $\check{C}_g$ is nearly equal to $C_{gf} + C_{pg}$. Generally, r_g is of the order of 250,000 ohms for adjustments for maximum detection with common tubes, so that if C_c is 100$\mu\mu$f or more, $(r_g C_c \omega_h)^2$ is large in comparison with unity. With these approximations Eq. (676) becomes

$$\left(\frac{\Delta E_g}{\Delta E_s}\right)^2 = \left(\frac{C_c}{C_c + \check{C}_g}\right)^2 \text{ (approx.)} \qquad (677)$$

Accordingly, if C_c is ten times $\check{C}_g$, the loss in the voltage of detection is about 17 per cent, which means a loss of about 31 per cent in sound intensity. If C_c is twenty times $\check{C}_g$, there is a loss of 9 per cent in voltage of detection and 17 per cent in sound intensity. If $\check{C}_g$ is of the order of 15$\mu\mu$f, C_c should be not much less than 150$\mu\mu$f. Making C_c large reduces the loss of high-frequency voltage but by-passes some of the currents of higher audio frequencies, thereby introducing frequency distortion, as was explained in Chap. XIX. This effect will be considered for the particular case of grid-circuit detection.

Expression (676) does not give the complete analysis of the effect of r_g and C_c upon the radio-frequency voltage ΔE_g, because ΔE_g, the voltage across the oscillatory circuit, depends greatly upon the values of r_g, $\check{C}_g$, R_c, and C_c. We may determine the square of the ratio of the voltage ΔE_g to the same voltage if

g_g is zero by starting with Eq. (616), page 502, and proceeding as in Chap. XIX. In this case Z_s comprises the input admittance of the tube in series with the combination of R_c shunted by C_c. Figure 281 gives the equivalent high-frequency circuit corresponding to the grid circuit of the detector.

The value of Z_s is

$$Z_s = \sqrt{\left[\frac{\breve{r}_g}{1 + (\breve{r}_g \breve{C}_g \omega_h)^2} + \frac{R_c}{1 + (R_c C_c \omega_h)^2}\right]^2 + \left[\frac{\breve{r}_g^2 \breve{C}_g \omega_h}{1 + (\breve{r}_g \breve{C}_g \omega_h)^2} + \frac{R_c^2 C_c \omega_h}{1 + (R_c C_c \omega_h)^2}\right]^2} \quad (678)$$

Then

$$\left[\frac{\Delta E_s}{(\Delta E_s)_{g_g=0}}\right]^2 = \left\{\frac{\dfrac{R_c Z_0 \breve{C}_g^2 \omega_h^2}{1 + R_c^2(\breve{C}_g + C_c)^2 \omega_h^2} + \dfrac{R_0}{Z_0}}{\dfrac{[\breve{r}_g + R_c + \breve{r}_g R_c (R_c C_c^2 + \breve{r}_g \breve{C}_g^2)\omega_h^2] Z_0}{(\breve{r}_g + R_c)^2 + \breve{r}_g^2 R_c^2 (\breve{C}_g + C_c)^2 \omega_h^2} + \dfrac{R_0}{Z_0}}\right\}^2 \quad (679)$$

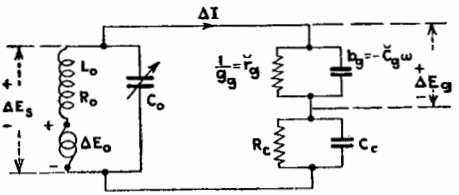


FIG. 281.—Equivalent grid circuit of the triode detector.

If $(\breve{r}_g C_c \omega_h)^2$ is large compared with unity, Eq. (679) reduces to the approximate expression

$$\left[\frac{\Delta E_s}{(\Delta E_s)_{g_g=0}}\right]^2 = \left[\frac{\dfrac{L_0^2 \omega_h^2}{R_0 R_c}\left(\dfrac{\breve{C}_g}{\breve{C}_g + C_c}\right)^2 + 1}{\dfrac{L_0^2 \omega_h^2}{R_0 \breve{r}_g}\left(\dfrac{C_c}{\breve{C}_g + C_c}\right)^2 + 1}\right]^2 \text{(approx.)} \quad (680)$$

Note that $L_0^2 \omega_h^2 / R_0$ is the equivalent resistance of the resonant oscillatory circuit. Since the first term in the numerator of Eq. (680) is generally small, a still further simplification may be made. The following approximate form for Eq. (680) results, and is sufficiently accurate for practical purposes.

$$\left[\frac{\Delta E_s}{(\Delta E_s)_{g_g=0}}\right]^2 = \frac{1}{\left[\dfrac{L_0 \omega_h}{\eta_0 \breve{r}_g}\left(\dfrac{C_c}{\breve{C}_g + C_c}\right)^2 + 1\right]^2} \text{(approx.)} \quad (681)$$

The reduction in voltage ΔE_s due to conduction through the tube, as given by Eq. (681), is a very important factor in the operation of a grid-circuit detector. Assume that C_c is ten times $\breve{C}_g$ and that $r_g = 250{,}000$ ohms. η_0 is usually made as small as possible to give a large value of $(\Delta E_s)_{g_g=0}$ and we may assume $\eta_0 = 0.005$. C_0 in receiving sets for broadcast reception is a variable air condenser and commonly has a capacitance of about $100\mu\mu$f when set to receive a frequency of 1,000 kc. per second, and a value of about $400\mu\mu$f for 550 kc. The value of the ratio of Eq. (681) for these two cases is 0.196 and 0.379.

Consider the frequency distortion arising in the grid circuit owing to the variation of impedance $(Z_c)_l$ with frequency. The conditions in this respect are much the same as in the study of frequency distortion for a diode detector given in Chap. XIX. The factor $\dfrac{(\mathbf{Z}_c)_l}{r_g + (\mathbf{Z}_c)_l}$ in Eq. (660) gives that fraction of the fictitious voltage of detection in the grid circuit which is usefully employed to act upon the grid. This factor is given by Eq. (606), page 499, provided the symbols are changed to fit the present case. Let F represent the numerical value of this factor. Then, by Eq. (606),

$$F = \frac{1}{\sqrt{\left(1 + \dfrac{r_g}{R_c}\right)^2 + (r_g C_c \omega_l)^2}} \tag{682}$$

In arriving at Eq. (682) it has been assumed that the tube capacitance $(\breve{C}_g)_l$ offers a very large reactance in comparison with r_g at the audio frequencies, or, in other words, that

$$[(\breve{r}_g)(\breve{C}_g)_l \omega_l]^2 << 1 \tag{683}$$

To make F as near unity as possible, R_c is made large in comparison with r_g. In order that the last term in Eq. (682) may have little effect and hence that there may be little variation of F with frequency,

$$(r_g C_c \omega_l)^2 << 1 \tag{684}$$

where r_g is of the order of 250,000 ohms. These conditions, together with the limitation placed on C_c by the radio-frequency part of the problem, are usually difficult to attain and a compromise must be made between frequency distortion and loss of sensitivity. It is ordinarily better to suffer loss of sensitivity and to

preserve fidelity of tone. To assist in calculations, values of $(r_g C_c w_l)^2$ are given in the following table for a resistance of $\frac{1}{4}$ megohm and a capacitance of $100\mu\mu$f. From the values given in the table, $(r_g C_c \omega_l)^2$ can be calculated for other values of C_c and r_g.

TABLE XV.—VALUES OF $(r_g C_c \omega_l)^2$ FOR $r_g = 0.25 \cdot 10^6$ OHMS, $C_c = 100\mu\mu$f

n, cycles per second	$(r_g C_c \omega_l)^2$
100	0.000247
500	0.00617
1,000	0.0247
2,000	0.0988
4,000	0.395
6,000	0.890
8,000	1.58
10,000	2.47

To illustrate the effect of C_c upon the operation of a detector, assume a radio frequency of 10^6 cycles per second and a maximum audio frequency of 6,000 cycles per second. Assume also that r_g is 250,000 ohms at the point of maximum detection, which simply means that the electrode capacitances of the triode, and not r_g, are of importance in determining the values of the ratio given in Eq. (674); but r_g is of importance in the calculation of the factor F of Eq. (682) giving the voltage across the oscillatory circuit as compared to the voltage induced in the circuit. If $C_{pg} + C_{gf}$ is $15\mu\mu$f and we choose C_c equal to $300\mu\mu$f, the ratio given in Eq. (674) is 0.91. But the factor F in Eq. (682) has, for an R_c of 2 megohms, the following values.

TABLE XVI.—VARIATION OF FACTOR F WITH FREQUENCY
$R_c = 2$ megohms; $C_c = 300\mu\mu$f

n, cycles per second	F	$F \times 0.91$
0	0.890	0.810
500	0.872	0.793
1,000	0.820	0.746
2,000	0.682	0.621
4,000	0.451	0.411
6,000	0.328	0.298
8,000	0.254	0.231
10,000	0.206	0.184

grid circuit, it is the only factor which is independent of outside circuits. If $(\text{Det. } E)_g$, k_g, C_{gf}, C_{pg}, R_c, C_c, and ω_l are given, the detection can be calculated for any given conditions.

211. Experimental Method of Testing a Grid-circuit Detector. The method of testing a triode detector is very similar to the method given in Chap. XIX for testing a diode detector. The diagram of connections is given in Fig. 282 and is the same as the diagram given in Fig. 271, page 504, except for the substitution of the triode for the diode. The procedure is the same as described in Chap. XIX.

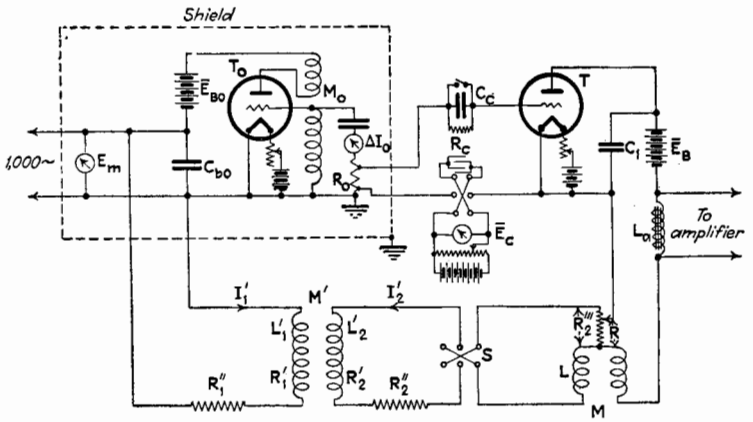


FIG. 282.—Measurement of detection by a triode.

If the output voltage of modulation frequency across a resistance, across a choke coil, or across the secondary coil of a transformer in the plate circuit of the detector is desired instead of the fictitious voltage $[(\Delta^2 E_p)_l]$, the resistance or choke coil can be connected directly across C_1, and if a transformer is used for plate load, the terminals of the balancing circuit, shown connected across C_1, are transferred to the secondary terminals of the transformer.

Since, in the experimental arrangement for measuring $[(\Delta^2 E_p)_l]$, there is no oscillatory circuit, this method does not include the effect of the second factor in Eq. (685), which takes account of the drop in voltage across the oscillatory circuit. Disregarding for the moment this second factor, $[(\Delta^2 E_p)_l]/m(\Delta E_g)_h^2$ is a function of $\bar{E}_c$, $\bar{E}_B$, C_c, R_c, and ω_l. The extent to which C_c, R_c, and ω_l affect the magnitude of the fictitious voltage $[(\Delta^2 E_p)_l]$ is clearly

shown by Eq. (685). The detecting action of the tube is determined by the voltages $\bar{E}_g$ and $\bar{E}_p$. The plate voltage $\bar{E}_p$ is practically equal to $\bar{E}_B$, but $\bar{E}_g$ is less than $\bar{E}_c$ by the voltage drop through R_c. Since the shape of the grid-current curve plotted against $\bar{E}_g$ is but slightly affected by the plate voltage, grid-circuit detection does not vary to any large extent as the plate voltage is changed. The plate voltage should be made such as

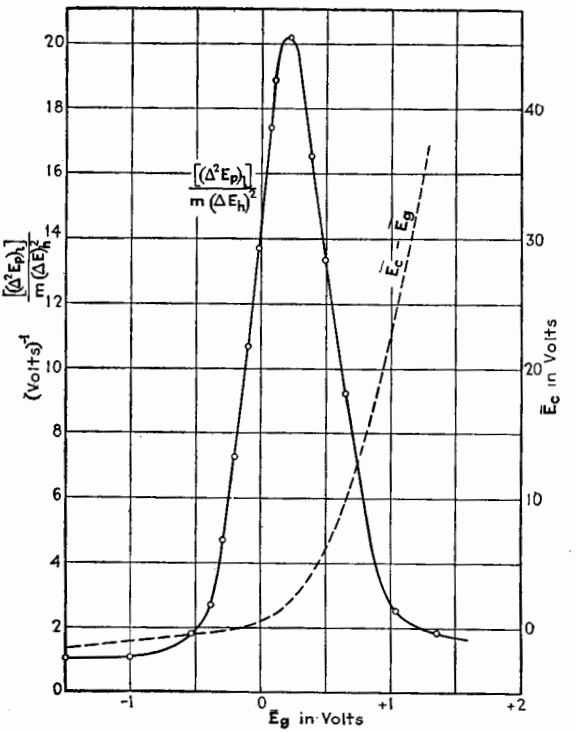


FIG. 283.—Grid-circuit detection. $u_p = 8.5$; $\bar{E}_B = 60$ volts; $R_c = 2.94$ megohms; $C_c = 200\mu\mu f$; $w_l = 2\pi 1{,}000$ radians per second.

to give best amplification, *i.e.*, so that the straight part of the plate-current curve occurs near $e_g = 0$. Grid-circuit detection can be expressed for a particular triode by one curve corresponding to this best value of $\bar{E}_p$.

Figures 283 and 284 give the curves (full-line) for $\frac{[(\Delta^2 E_p)_l]}{m(\Delta E_g)_h^2}$ for two triodes, one having a u_p of 8.5 and the other a u_p of 20. The plate voltages and other constants are given in the figures.

The dotted curves show the relation between the actual grid voltage, plotted as abscissa, and the polarizing-battery voltage, indicated by the scales at the right of the figures.

212. Plate-circuit Detection. Discussion.—Although in the early days of radio communication grid-circuit detection was almost universally used, some advantages of plate-circuit detection were recognized as the art developed.

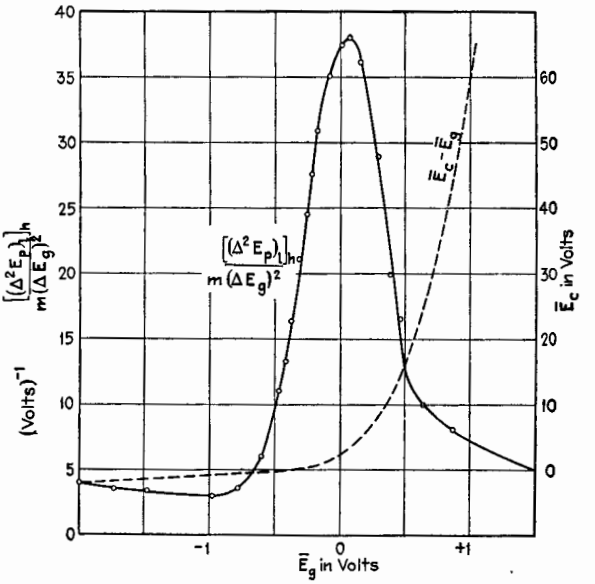


FIG. 284.—Grid-circuit detection. $u_p = 20$; $\bar{E}_B = 60$ volts; $R_c = 2.94$ megohms; $C_c = 200\mu\mu f$; $w_l = 2\pi 1{,}000$ radians per second.

The expression given in Eq. (664) for plate-circuit detection is

$$[(\Delta^2 E_p)_l] = (\text{Det. } E)_p \sqrt{2} m (\Delta E_g)_h^2 \tag{664}$$

where (Det. E)$_p$ is given by the long expression of Eq. (666). If u_p is constant and i_g is zero, (Det. E)$_p$ reduces to the simple expression

$$(\text{Det. } E)_p = u_p \left(1 - \frac{(R_b)_h}{r_p + (R_b)_h}\right)^2 \cdot \frac{1}{2s_p} \cdot \frac{\partial s_p}{\partial e_g} \tag{670}$$

$$= u_p^2 \left(\frac{r_p}{r_p + (R_b)_h}\right)^2 \cdot \frac{1}{2k_p} \cdot \frac{\partial k_p}{\partial e_p} \tag{671}$$

Generally these expressions are sufficiently accurate for all practical purposes when dealing with high-vacuum triodes.

Plate-circuit detection is clearly simpler and much freer from frequency distortion than grid-circuit detection, because Eq. (664) contains no factor which is a function of the modulation frequency $\omega_l/2\pi$. It is true in both types of detection that the fraction of $[(\Delta^2 E_p)_l]$ which is available may be a function of ω_l on account of the shunting capacitance C_1. For this reason, C_1 must be made as small as possible without an unwarranted sacrifice in detection.

From Eqs. (670) and (671), it is advantageous to make $(R_b)_h$ small in comparison with r_p. This is accomplished by the connection of C_1 across the plate load. Sometimes a tickler coil is included in the plate circuit to act regeneratively upon the grid circuit. In this case, it is advantageous to make the inductance of the tickler coil as small as possible so as not to increase $(X_b)_h$ unnecessarily, but the mutual inductance between the tickler coil and the grid inductance must be sufficient to give the required regeneration. Usually, regeneration, by increasing $(\Delta E_g)_h$, more than compensates for the decrease in detection coefficient.

In the use of plate-circuit detection the grid is polarized negatively so that k_g is zero. The presence of the detector connected across the oscillatory circuit in the grid increases the capacitance by approximately $C_{pg} + C_{gf}$ without decreasing the voltage across the circuit. ΔE_g of Eq. (664) is, therefore, the same as $(\Delta E_s)_{g_g=0}$ of Eq. (685).

The most troublesome feature of plate-circuit detection arises from the fact that k_p is usually small where detection is greatest. This small value of k_p, or large value of r_p, makes it more difficult to make available as large a fraction of $[(\Delta^2 E_p)_l]$ as in the case of grid-circuit detection. This is shown by Eq. (674). Therefore, for plate-circuit detection, the plate load should have a higher impedance at audio frequencies than for grid-circuit detection.

In Eq. (674) appears the capacitance C_1 which shunts the plate load Z_1. The amount of frequency distortion produced in the plate circuit depends upon the value of $r_p C_1 \omega_l$. If the frequency distortion is to be no greater than that introduced in the plate circuit of a grid-circuit detector, then, with the larger value of r_p, C_1 must be correspondingly smaller. We may set up the same criterion as in Eq. (675) or

$$(r_p C_1 \omega_l)^2 << 1 \tag{686}$$

The only harmful effect of making C_1 very small is the loss of sensitivity due to the reduction of (Det. $E)_p$ as given by Eq. (666). To obtain an approximate idea of the effect of the size of C_1 on the value of (Det. $E)_p$, we may, without much error, substitute for $(R_b)_h$ in Eq. (670) the impedance of C_1, or $1/jC_1\omega_h$. The approximate value of the detection coefficient becomes

$$(\text{Det. } E)_p = u_p\left(1 - \frac{1}{\sqrt{1 + (r_p C_1 \omega_h)^2}}\right)^2 \cdot \frac{1}{2s_p} \cdot \frac{\partial s_p}{\partial e_g} \quad (687)$$

If

$$(r_p C_1 \omega_h)^2 >> 1 \quad (688)$$

the detection coefficient is little affected by the size of C_1. The choice of C_1 is a compromise between Eqs. (686) and (688).

It is of considerable importance to note that, whereas in plate-circuit detection frequency distortion is introduced only in the plate-circuit, in grid-circuit detection frequency distortion is intro-

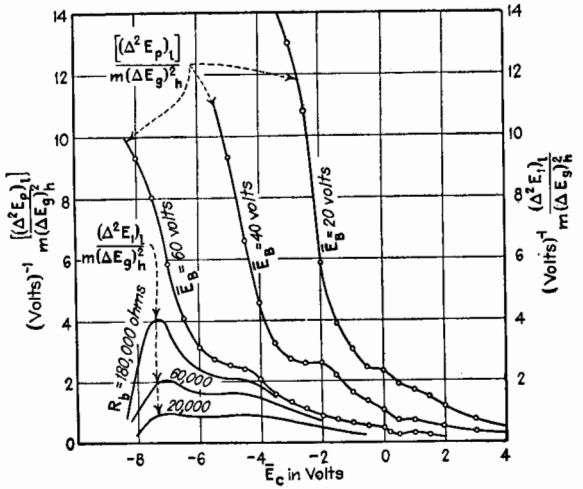


FIG. 285.—Plate-circuit detection. UX201A. $u_p = 8.5$.

duced in both the grid and the plate circuits. As a consequence, grid-circuit detection, although a little more sensitive, usually gives considerably more frequency distortion. Plate-circuit detection is therefore better for high-quality receiving apparatus.

213. Experimental Measure of Plate-circuit Detection.—The apparatus shown in Fig. 282 may be used for measuring the fictitious voltage of detection, $[(\Delta^2 E_p)_l]$, or the output voltage,

$(\Delta^2 E_1)_l$, produced by plate-circuit detection. The only change necessary is to short-circuit C_c.

Curves of $[(\Delta^2 E_p)_l]/m(\Delta E_g)_h^2$ as determined by this method are plotted in Figs. 285 and 286 for various plate voltages $\bar{E}_B$. The curves of Fig. 285 are for a triode having a voltage ratio of 8.5, while those of Fig. 286 are for a triode having a voltage ratio of 20. The curves in these two figures, having notations for certain

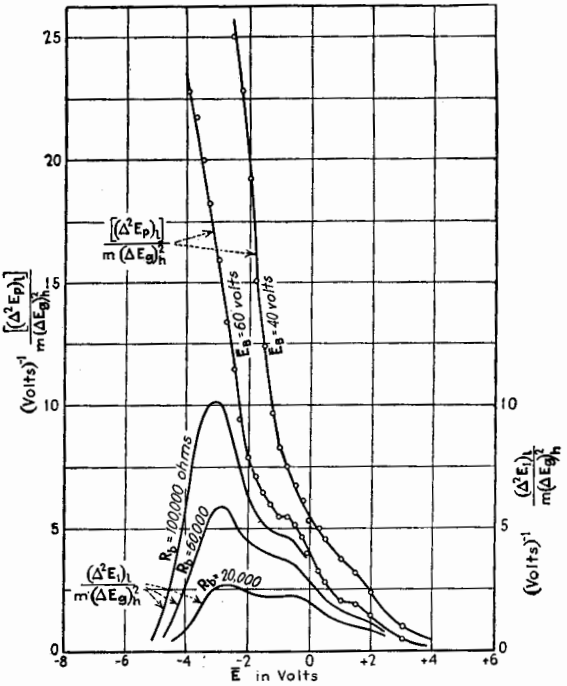


FIG. 286.—Plate-circuit detection. $u_p = 20$.

resistances, give the voltage across these resistances when connected in the plate circuit. These curves are plotted to the scale of $(\Delta^2 E_1)_l/m(\Delta E_g)_h^2$ at the right-hand side of the figures.

The curves of fictitious voltage rise indefinitely as the polarizing voltage increases negatively, because k_p approaches zero.

214. Recapitulation. Summary of Grid- and Plate-circuit Detection.—To assist in comparison of the two types of detectors and for convenient reference, the formulas and criteria for the two kinds of detection are collected.

Grid-circuit Detection:

$$[(\Delta^2 \mathbf{E}_p)_l] = -\frac{(\mathbf{Z}_c)_l}{r_g + (\mathbf{Z}_c)_l}(\text{Det. } E)_g\sqrt{2}m(\Delta E_g)_h^2 \tag{660}$$

where $(\text{Det. } E)_g = \frac{u_p}{2k_g} \cdot \frac{\partial k_g}{\partial e_g}$ (662)

If $(\mathbf{Z}_c)_l$ is R_c shunted by C_c, and if $(\Delta E_g)_h^2$ is expressed in terms of the voltage which would exist across the oscillatory circuit in the grid circuit if g_g were zero, this voltage being denoted by $(\Delta E_s)_{g_g=0}$, the numerical value of $[(\Delta^2 E_p)_l]$ is

$$[(\Delta^2 E_p)_l] = \frac{1}{\sqrt{\left(1 + \frac{r_g}{R_c}\right)^2 + (r_g C_c \omega_l)^2}} \cdot \left\{ \frac{\frac{R_c Z_0 \breve{C}_g^2 \omega_h^2}{1 + R_c^2(\breve{C}_g + C_c)^2\omega_h^2} + \frac{R_0}{Z_0}}{\frac{[\breve{r}_g + R_c + \breve{r}_g R_c(R_c C_c^2 + \breve{r}_g \breve{C}_g^2)\omega_h^2]Z_0}{(\breve{r}_g + R_c)^2 + \breve{r}_g^2 R_c^2(\breve{C}_g + C_c)^2\omega_h^2} + \frac{R_0}{Z_0}} \right\} \cdot \frac{1 + (R_c C_c \omega_h)^2}{\left(1 + \frac{R_c}{\breve{r}_g}\right)^2 + R_c^2(C_g + C_c)^2\omega_h^2} \cdot (\text{Det. } E)_g\sqrt{2}m(\Delta E_s)_{g_g=0}^2 \tag{689}$$

The numerical value of the voltage across the plate load is

$$(\Delta^2 E_1)_l = \frac{[(\Delta^2 E_p)_l]}{\sqrt{1 + \frac{r_p(r_p + 2R_1)}{Z_1^2} + (r_p C_1 \omega_l)^2\left(1 - \frac{2X_1}{C_1 \omega_l Z_1^2}\right)}} \tag{674}$$

For small frequency distortion,

$$(r_g C_c \omega_l)^2 << 1 \tag{684}$$

$$(r_p C_1 \omega_l)^2 << 1 \tag{675}$$

For greatest sensitivity,

$$\frac{C_c}{(\breve{C}_g)_h} >> 1 \tag{690}$$

$$\frac{R_c}{r_g} >> 1 \tag{691}$$

$$\frac{(Z_1)_l}{r_p} >> 1 \tag{692}$$

If $[(\breve{r}_g)_h C_c \omega_h]^2 >> 1$, as is almost always true, Eq. (689) reduces to the more practical form

$$[(\Delta^2 E_p)_l] = \frac{1}{\sqrt{\left(1+\frac{r_g}{R_c}\right)^2 + (r_g C_c \omega_l)^2}} \cdot \frac{1}{\left[1+\frac{L_0\omega_h}{\eta_0 \breve{r}_g}\left(\frac{C_c}{(\breve{C}_g)_h + C_c}\right)^2\right]^2} \cdot \left(\frac{C_c}{(\breve{C}_g)_h + C_c}\right)^2 \cdot (\text{Det. } E)_g \sqrt{2m(\Delta E_s)^2_{g_g=0}} \quad (685)$$

Plate-circuit Detection:

$$[(\Delta^2 E_p)_l] = (\text{Det. } E)_p \sqrt{2m(\Delta E_s)^2_{g_g=0}} \quad (664)$$

where

$$(\text{Det. } E)_p = u_p\left[\frac{1}{2s_p}\frac{\partial s_p}{\partial e_g} + \frac{u_p(R_b)_h^2}{(r_p+(R_b)_h)^2}\cdot\frac{1}{2k_p}\cdot\frac{\partial k_p}{\partial e_p} - \frac{2u_p(R_b)_h}{r_p+(R_b)_h}\cdot\frac{1}{2s_p}\cdot\frac{\partial s_p}{\partial e_p}\right] \quad (666)$$

If $\bar{E}_c$ is negative and u_p is essentially constant, Eq. (666) reduces to the practical form

$$(\text{Det. } E)_p = u_p\left[1-\frac{(R_b)_h}{r_p+(R_b)_h}\right]^2 \cdot \frac{1}{2s_p}\cdot\frac{\partial s_p}{\partial e_g} \quad (670)$$

$$= u_p^2\left[\frac{r_p}{r_p+(R_b)_h}\right]^2 \cdot \frac{1}{2k_p}\cdot\frac{\partial k_p}{\partial e_p} \quad (671)$$

$$(\Delta^2 E_1)_l = \frac{[(\Delta^2 E_p)_l]}{\sqrt{1+\frac{r_p(r_p+2R_1)}{Z_1^2} + (r_p C_1\omega_l)^2\left(1-\frac{2X_1}{C_1\omega_l Z_1^2}\right)}} \quad (674)$$

For small frequency distortion,

$$(r_p C_1 \omega_l)^2 << 1 \quad (675)$$

For greatest sensitivity,

$$(r_p C_1 \omega_h)^2 >> 1 \quad (688)$$

General References

BALLANTINE: Detection with Grid Rectification with the High-vacuum Triode, *Proc. I.R.E.*, **16**, 593 (1928).

COLEBROOK: The Rectification of Small Radio Frequency Potential Differences by Means of Triode Valves, *Exp. Wireless*, **2**, 865, 946 (1925); **3**, 34, 90 (1926).

SMITH: Theory of Detection in a High-vacuum Thermionic Tube, *Proc. I.R.E.*, **14**, 649 (1926).

TERMAN: Some Principles of Grid-leak Grid-condenser Detection, *Proc. I.R.E.*, **16**, 1384, 19 (1928).

CHAPTER XXI

THEORY OF THE OPERATION OF NONLINEAR CIRCUITS WITH LARGE ELECTRICAL VARIATIONS

SPECIAL REFERENCE TO DETECTION WITH LARGE SIGNALS

215. Introduction.—The theory so far presented (excepting the treatment of regeneration with large signals) has been limited to circuits in which the electrical variations have been assumed so small that the first-order effects were calculated neglecting the curvature of the path of operation. When detection was studied, the second-order effects were obtained by assuming the simplest curve for the portion of the path of operation covered by the electrical variations. This assumed curve was a second-degree or parabolic curve.

When the electrical variations cover a substantial portion of the curved characteristic, the theory of the action in nonlinear circuits is very difficult. One method[1] which has been suggested for treating such problems depends upon fitting the curved characteristic by some explicit mathematical function. This method is in general unsatisfactory, because it is usually impossible to find a simple mathematical expression which fits the characteristic with sufficient accuracy. Another method[2] which has been used to some extent makes use of a power-series development to express the curved characteristic of the nonlinear circuit. Since all problems involve operation about some point on the characteristic curve, a Taylor's development of the curve about this quiescent point is usually adopted. This series development expresses the characteristic in terms of the first, second, third, etc., derivatives of the curve *at the quiescent point.* Usually so many terms are required in the development that calculation is very laborious and the solution is extremely complicated. A further objection to this method arises from the difficulty of obtaining accurately the several derivatives of the experimental

[1] BARCLAY, *Exp. Wireless*, **6**, 178 (1929).

[2] LLEWELLYN, *Bell. Tech. J.*, **5**, 433, July, 1926.

curve giving the path of operation, these derivatives being obtained at the quiescent point.

The method of attack here presented expresses the result in terms of either the first or the second derivative of the characteristic curve at *all* points along the path of operation or in terms of the current itself at all points of the path. The method is applicable to any form of characteristic curve even with sharp bends. The solution is expressed as an integral instead of as a series. Although, with this method, calculation of numerical results is sometimes laborious, the form of expression of the results lends itself to easy graphical evaluation as will be explained. Furthermore, the expressions for the final results are accurate and are general in form, since they can be transformed readily to the results derived in the application of any of the other methods referred to.

I. SIMPLE UNDERLYING PRINCIPLE OF THE METHOD

216. Method.—The method to be described is the application of the Fourier analysis to the current wave form. This wave form is first derived from the known dynamic path of operation for the *whole* circuit. The current i in the circuit is plotted against the instantaneous voltage e_0 impressed in the circuit. Let $i = F(e_0)$ be this path of operation, usually determined experimentally, although the analytical form of the function may be known or assumed. There are no restrictions as to the shape of this curve other than that i shall be single valued, which means that i must be the same for ascending and descending values of e_0. If the magnitude and form of the applied voltage variation e_0 are known, and if the polarizing potential $\bar{E}_0$ about which this voltage fluctuation takes place is given, the form of the current wave can be deduced. This current is in general nonsinusoidal, and the magnitude of any sinusoidal component can be obtained by Fourier's theorem.

Before describing the method in detail, there is an important variation of the method which is here noted. Instead of using the path of operation $i = F(e_0)$ to determine the current wave form, the current may be derived from the slope of the path of operation plotted against e_0, or even from the second derivative of the path also plotted against e_0.

In Fig. 287*a*, an illustrative curve is drawn for $i = F(e_0)$, and in Fig. 287*b* the same curve is drawn as expressed in the con-

verse form $e_0 = \psi(i)$. The derivative of i with respect to e_0, *i.e.*, k_0, is also plotted in Fig. 287*a*; and the curve of r_0, *i.e.*, de_0/di, is drawn as a function of i in Fig. 287*b*. Let Q be the quiescent point, and $\bar{I}$ and $\bar{E}_0$ the quiescent current and voltage. If the applied voltage is increased from $\bar{E}_0$ by an amount $\bar{e}_0$, the increase in current above $\bar{I}$, denoted by $\bar{\imath}$, is

$$\bar{\imath} = \int_0^{\bar{e}_0} k_0 de_0 \tag{693}$$

where Q is the origin of coordinates for $\bar{\imath}$ and $\bar{e}_0$. The shaded area in Fig. 287*a* gives the value of $\bar{\imath}$. Similarly,

$$\bar{e}_0 = \int_0^{\bar{\imath}} r_0 di \tag{694}$$

and $\bar{e}_0$ is given by the shaded area in Fig. 287*b*.

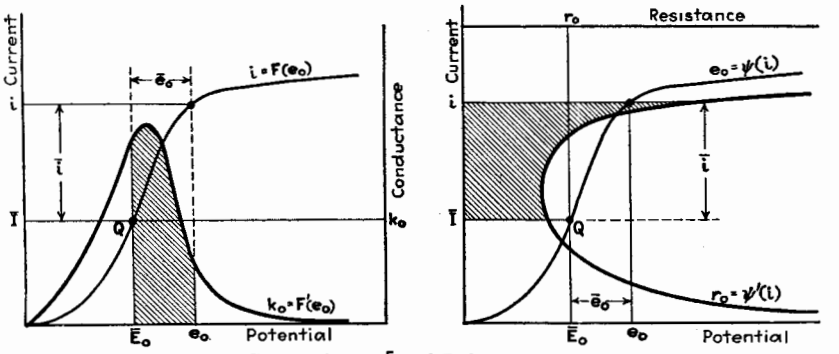


FIG. 287.—Calculation of $\bar{\imath}$ and $\bar{e}_0$ from curves of k_0 and r_0.

The curve of k_0 or of r_0 can be determined directly by the experimental methods described in Chap. IX. Usually the curve of k_0 is preferred because it does not go to infinite values as the curve of r_0 often does. The advantage of using the curve of k_0 rather than the curve of i is a gain in accuracy and in simplicity. The curve of k_0 is usually expressible by an equation of lower degree than the curve of i. If the curve of i against e_0 is a parabola, the curve of k_0 is a straight line.

We may go one step further and make use of the second derivative of the characteristic curve. Methods were described in Chap. XIX for measuring directly the quantity dk_0/de_0 by measuring the current of detection for small signals.

$$k_0 = \int_0^{\bar{e}_0} \left(\frac{dk_0}{de_0}\right) de_0 \tag{695}$$

The value for k_0 given by Eq. (695) can be substituted in Eq. (693) to give i in terms of dk_0/de_0.

II. APPLICATION OF THE THEORY TO TWO-TERMINAL DEVICE

217. Sinusoidal E.m.f. Impressed in Circuit Containing Diode. $\tilde{R} = \bar{R}$.—Study the method of deriving the current that flows in a circuit containing a nonlinear element when a large sinusoidal e.m.f. $\tilde{e}_0$ is impressed, where

$$\tilde{e}_0 = \hat{E}_0 \sin \omega t \tag{696}$$

The circuit is shown in Fig. 288, the resistance R being independent of frequency.

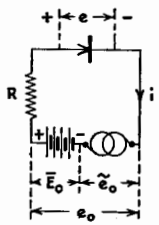


FIG. 288.—Series circuit containing a nonlinear element.

Let the characteristic curve of the whole circuit be represented by

$$i = F(e_0) \tag{697}$$

We shall denote the variational conductance di/de_0 by

$$k_0 = F'(e_0).$$

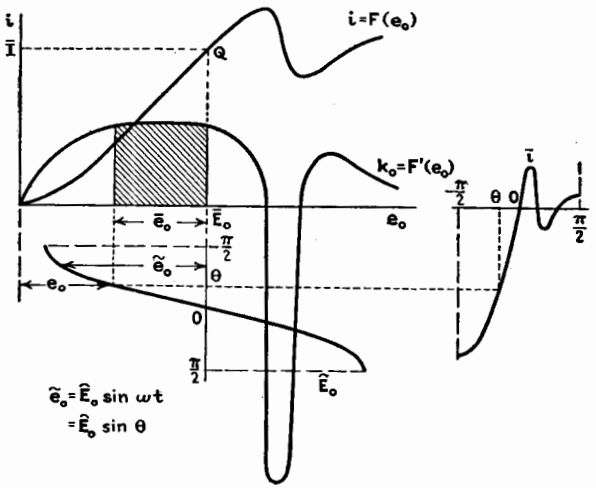


FIG. 289.—Calculation of $\bar{i}$ from k_0 when a sinusoidal e.m.f. is impressed in the circuit.

The curves of i and k_0 are shown in Fig. 289. If the large sinusoidal e.m.f. $\tilde{e}_0$ and the steady potential $\bar{E}_0$ are impressed in

series upon the circuit, the current for one-half of a cycle of the impressed e.m.f. can be obtained directly by projection from the current curve as shown in Fig. 289. On the other hand, the instantaneous current i measured from the quiescent current $\bar{I}_0$, at any time in the cycle corresponding to the impressed voltage $\bar{e}_0$, can be obtained by Eq. (693). The current curve $\bar{\imath} = F_1(e_0)$ is shown in the right-hand part of Fig. 289 as determined from the area under the k_0 curve.

The current i is nonsinusoidal and can be expressed by a Fourier series. Since the origin of time is taken when $\bar{e}_0$ is a maximum, $\bar{\imath}$ can be expressed by

$$\bar{\imath} = \bar{\bar{I}} - \bar{I} + (\hat{I})_1 \sin \omega t + (\hat{I})_2 \cos 2\omega t + (\hat{I})_3 \sin 3\omega t + \cdots \quad (698)$$

The problem is to determine the values of the steady component $\bar{\bar{I}} - \bar{I}$, and the amplitudes $(\hat{I})_1$, $(\hat{I})_2$, etc., of the fundamental and harmonic currents.

The Fourier analysis gives the following formulas for these currents:

$$\bar{\bar{I}} - \bar{I} = \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \bar{\imath}\, d\theta \quad (699)$$

$$(\hat{I})_1 = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \bar{\imath} \sin\theta\, d\theta = \frac{2}{\pi \hat{E}_0}\int_{-\hat{E}_0}^{\hat{E}_0} \bar{\imath}\frac{e_0}{\sqrt{\hat{E}_0^2 - e_0^2}} de_0 \quad (700)$$

$$(\hat{I})_2 = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \bar{\imath} \cos 2\theta\, d\theta \quad (701)$$

$$(\hat{I})_3 = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \bar{\imath} \sin 3\theta\, d\theta \quad (702)$$

where $\theta = \omega t$ and $\bar{\imath} = F_1(e_0)$.
If, in place of $\bar{\imath}$, its value as expressed by Eq. (693) is substituted, the formulas become

$$\bar{\bar{I}} - \bar{I} = \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta \int_0^{\bar{e}_0} k_0\, de_0 \quad (703)$$

$$(\hat{I})_1 = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta\, d\theta \int_0^{\bar{e}_0} k_0\, de_0 \quad (704)$$

$$(\hat{I})_2 = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos 2\theta \, d\theta \int_0^{\bar{e}_0} k_0 \, de_0 \tag{705}$$

The double-integral expressions for the several component currents can be simplified by integration by parts.[3] This reduction of Eqs. (703), (704), etc., gives the following forms:

$$\bar{\bar{I}} - \bar{I} = \frac{1}{2}\int_0^{\hat{E}_0} k_0 \, de_0 + \frac{1}{2}\int_0^{-\hat{E}} k_0 \, de_0 - \frac{1}{\pi}\int_{-\hat{E}_0}^{\hat{E}_0} k_0\theta \, de_0 \tag{706}$$

$$= \frac{1}{2}\int_0^{\hat{E}_0} k_0 \, de_0 + \frac{1}{2}\int_0^{-\hat{E}_0} k_0 \, de_0 - \frac{\hat{E}_0}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0\theta \cos\theta \, d\theta \tag{707}$$

$$(\hat{I})_1 = \frac{2}{\pi}\int_{-\hat{E}_0}^{\hat{E}_0} k_0 \cos\theta \, de_0 \tag{708}$$

$$= \frac{2\hat{E}_0}{\pi}\int_{-1}^{1} k_0\sqrt{1-y^2}\, dy \text{ where } y = \frac{\bar{e}_0}{\hat{E}_0} = \sin\theta \tag{709}$$

$$= \frac{2\hat{E}_0}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \cos^2\theta \, d\theta \tag{710}$$

$$(\hat{I})_2 = -\frac{1}{\pi}\int_{-\hat{E}_0}^{\hat{E}_0} k_0 \sin 2\theta \, de_0 \tag{711}$$

$$= -\frac{2}{\pi}\int_{-1}^{1} k_0 y\sqrt{1-y^2}\, dy \tag{712}$$

$$= -\frac{2\hat{E}_0}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \sin\theta\cos^2\theta \, d\theta \tag{713}$$

[3] The simplification of Eq. (704) will be carried through as an illustration of this method.

In the familiar formula $\int u\, dv = uv - \int v\, du$, let $u = \int_0^{\bar{e}_0} k_0 \, de_0$ and $dv = \sin\theta \, d\theta$. Then $du = \frac{du}{de_0} de_0 = k_0 \, de_0$ and $v = -\cos\theta$

Substituting these values, Eq. (704) becomes

$$(\hat{I})_1 = \frac{2}{\pi}\left[-\cos\theta \int_0^{\bar{e}_0} k_0 de_0\right]\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} - \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (-\cos\theta) k_0 \, de_0$$

$$= \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \cos\theta \, de_0 \text{ (as given by Eq. (708)).}$$

$$(\hat{I})_3 = \frac{2}{3\pi}\int_{-\hat{E}_0}^{\hat{E}_0} k_0 \cos 3\theta \, de_0 \tag{714}$$

$$= \frac{2\hat{E}_0}{3\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \cos 3\theta \cos\theta \, d\theta \qquad \text{etc.} \tag{715}$$

The integrals of k_0 in Eqs. (703), (704), (705), and the first two integrals of Eqs. (706) and (707) are straightforward integrals giving the area under the k_0 curve between two values of the voltage. In the other integrals, there is a definite connection between the variables which does not appear explicitly in the integrals. In the last integral of Eq. (706), each value of k_0 corresponds to a particular value of $\bar{e}_0$ determined by θ and $\hat{E}_0$.

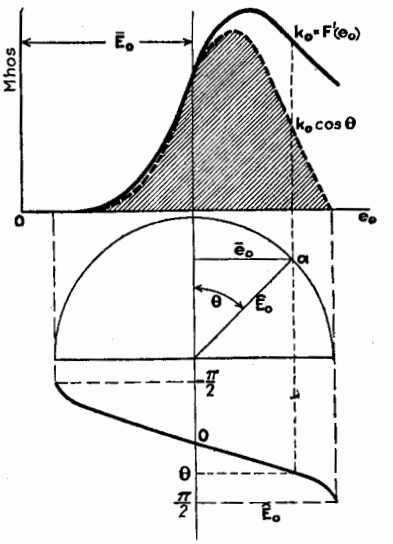


FIG. 290.—Calculation of the amplitude of the fundamental component of current from the curve of k_0.

Several equivalent forms are given for each quantity. These expressions can be evaluated by direct integration if the mathematical expression for k_0 is known. If k_0 is known only by an empirical plot, these expressions can be evaluated graphically, as illustrated for $(\hat{I})_1$ in Fig. 290. In this figure, the curve for k_0 is assumed, and the curve $k_0 \cos\theta$ is derived from it by a method obvious from the diagram. The shaded area multiplied by $2/\pi$ gives the value of $(\hat{I})_1$ as shown by Eq. (708). If the curve of $\bar{\imath}$ against e_0 is known, Eqs. (699), (700), (701), and (702) may be used in place of the expressions of Eqs. (706) through (715). The method of calculation is similar to that already described.

Equation (710) gives the current of fundamental frequency due to the impressed e.m.f. $\hat{E}_0$, and hence the equivalent conductance of the circuit is

$$\text{Equivalent conductance} = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \cos^2\theta \, d\theta \tag{716}$$

The expressions of Eqs. (706) and (707) are useful in studying the action of vacuum-tube voltmeters. Equations (708) through (715) are useful in determining the distortion of amplifiers caused by a nonlinear characteristic and the consequent production of harmonics. Some problems in which these expressions are used will be given in Part IV of this chapter.

218. Modulated E.M.F. Impressed in Circuit Containing Diode. $\tilde{R} = \bar{R}$.—The most natural method of attacking the problem of a modulated impressed e.m.f. is to develop the wave form directly by Fourier's method.

Let the impressed e.m.f. be

$$e_0 = \hat{E}_0(1 + m \cos \omega_l t) \cos \omega_h t, \tag{717}$$

where m is the degree of modulation and ω_l is the angular velocity of the modulation frequency. Since $\bar{\imath} = F_1(e_0)$, the current wave form is obtained by substituting Eq. (717) in this functional equation. In order to simplify the analysis, let it be assumed that $\omega_h = p\omega_l$, where p is a whole number. Then $\bar{\imath}$ is an even function because $F_1(e_0) = F_1(-e_0)$, and hence $\bar{\imath}$ can be developed in a cosine series, the period of the function being T_l.

$$(I)_0 = \frac{1}{\pi}\int_0^\pi \bar{\imath}\, d\phi \quad (\text{where } \phi = \omega_l t) \tag{718}$$

$$(\hat{I})_h = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos p\phi\, d\phi \tag{719}$$

$$(\hat{I})_{nh} = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos np\phi\, d\phi \tag{720}$$

$$(\hat{I})_l = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos \phi\, d\phi \tag{721}$$

$$(\hat{I})_{nl} = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos n\phi\, d\phi \tag{722}$$

These equations can be evaluated if $\bar{\imath} = F_1(e_0)$ is known in mathematical form or as a series development in terms of e_0. In such a development, e_0 is given the value from Eq. (717),

$$e_0 = \hat{E}_0(1 + m \cos \phi) \cos p\phi \tag{723}$$

Alternative forms for Eqs. (720) and (722) are obtained by replacing $\bar{\imath}$ by its value from Eq. (693) and integrating by parts.

The final results are

$$(\hat{I})_{nh} = -\frac{2}{\pi}\int_0^{\pi} k_0 \frac{\sin np\phi}{np} \cdot \frac{de_0}{d\phi}\, d\phi \tag{724}$$

$$= \frac{2\hat{E}_0}{\pi}\int_0^{\pi} k_0 \frac{\sin np\phi}{np}[(1 + m\cos\phi)p \sin p\phi + m \sin\phi \cos p\phi]\, d\phi \tag{725}$$

$$(\hat{I})_{nl} = -\frac{2}{\pi}\int_0^{\pi} k_0 \frac{\sin n\phi}{n} \cdot \frac{de_0}{d\phi}\, d\phi \tag{726}$$

$$= \frac{2\hat{E}_0}{\pi}\int_0^{\pi} k_0 \frac{\sin n\phi}{n}[(1 + m\cos\phi)p \sin p\phi + m \sin\phi \cos p\phi]\, d\phi \tag{727}$$

All expressions from Eq. (718) through Eq. (727) are fairly easy to evaluate if either $\bar{\imath}$ or k_0 is known mathematically. None of them, however, lends itself to easy graphical evaluation.

Consider a second method of attack giving a form which can easily be evaluated either analytically or graphically. Let

$$\tilde{e}_0 = \hat{E}_0(1 + m\cos\omega_l t)\sin\omega_h t \tag{728}$$

We shall limit the treatment here to the condition that the impedance in series with the diode is a pure resistance having the same value at all frequencies, deferring the consideration of other cases.

Since the amplitude of the modulated e.m.f. pulsates as given by Eq. (728), we may substitute for $\hat{E}_0$ in Eqs. (706) through (715) the expression

$$\hat{e}_0 = \hat{E}_0(1 + m\cos\phi) \tag{729}$$

where $\phi = \omega_l t$, and $\hat{e}_0$ represents an amplitude which is a function of time. Substituting Eq. (729) in Eqs. (708), (711), and (714) gives nothing different from Eqs. (724) and (725). Using Eq. (706) or Eq. (707), however, we obtain new expressions for the current of modulation frequency and its harmonics. These expressions will now be developed.

The r.m.s. value of the component of n times modulation frequency is obtained by expressing the Fourier component as follows:

$$(I)_{nl} = \frac{\sqrt{2}}{\pi}\int_0^{\pi} (\bar{\bar{I}} - \bar{I})\cos n\phi\, d\phi \tag{730}$$

where n has the value one for the fundamental, two for the second harmonic, etc. The current $\bar{\bar{I}} - \bar{I}$ is given by either Eq. (699) or Eq. (706), in which $\hat{E}_0$ has the value given by Eq. (729). If Eq. (706) is substituted in Eq. (730), the latter can be considerably simplified by integration by parts,[4] giving as a final expression

$$(I)_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\int_0^{\pi}\frac{\sin\phi\,\sin n\phi}{n}d\phi\int_{-\hat{e}_0}^{\hat{e}_0}k_0\frac{e_0/\hat{e}_0}{\sqrt{1-(e_0/\hat{e}_0)^2}}d\left(\frac{e_0}{\hat{e}_0}\right) \tag{731}$$

[4] The method of simplification follows. When Eq. (706) is substituted in Eq. (730), the resulting equation contains three integrals.

The first integral is

$$\frac{1}{\sqrt{2}\pi}\int_0^{\pi}\cos n\phi\,d\phi\int_0^{\hat{e}_0}k_0\,de_0 \tag{a}$$

In using the canonical expression $\int u\,dv = uv - \int v\,du$, let $\int_0^{\hat{e}_0}k_0\,de_0 = u$, then $du = du/d\hat{e}_0\cdot d\hat{e}_0/d\phi\cdot d\phi = (k_0)_{\hat{e}_0}[-\hat{E}_0 m\sin\phi\,d\phi]$, and if dv is $\cos n\phi\,d\phi$, $v = \dfrac{\sin n\phi}{n}$. Integral (a) then becomes

$$\left|\frac{\sin n\phi}{\sqrt{2}n\pi}\int_0^{\hat{e}_0}k_0\,de_0\right|_0^{\pi} + \frac{\hat{E}_0 m}{\sqrt{2}\pi}\int_0^{\pi}(k_0)_{\hat{e}_0}\frac{\sin\phi\,\sin n\phi}{n}d\phi \tag{b}$$

The first term is zero.

Similarly, the second integral becomes

$$-\frac{\hat{E}_0 m}{\sqrt{2}\pi}\int_0^{\pi}(k_0)_{-\hat{e}_0}\frac{\sin\phi\,\sin n\phi}{n}d\phi \tag{c}$$

The third integral of Eq. (730) is

$$-\frac{\sqrt{2}}{\pi^2}\int_0^{\pi}\cos n\phi\,d\phi\int_{-\hat{e}_0}^{\hat{e}_0}k_0\sin^{-1}\frac{e_0}{\hat{e}_0}\,de_0 \tag{d}$$

In reducing this expression let $u = \int_{-\hat{e}_0}^{\hat{e}_0}k_0\sin^{-1}\dfrac{e_0}{\hat{e}_0}\,de_0$, then

$$du = \left[\frac{\pi}{2}(k_0)_{\hat{e}_0} - \frac{\pi}{2}(k_0)_{-\hat{e}_0} + \int_{-\hat{e}}^{\hat{e}_0}k_0\frac{-e_0/\hat{e}_0}{\sqrt{1-(e_0/\hat{e}_0)^2}}\frac{de_0}{\hat{e}_0}\right]\left[-\hat{E}_0 m\sin\phi\,d\phi\right] \tag{e}$$

Expression (d) becomes

$$\frac{-\hat{E}_0 m}{\sqrt{2}\pi}\int_0^{\pi}(k_0)_{\hat{e}_0}\frac{\sin\phi\,\sin n\phi}{n}d\phi + \frac{\hat{E}_0 m}{\sqrt{2}\pi}\int_0^{\pi}(k_0)_{-\hat{e}_0}\frac{\sin\phi\,\sin n\phi}{n}d\phi + \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\int_0^{\pi}\frac{\sin\phi\,\sin n\phi}{n}d\phi\int_{-\hat{e}_0}^{\hat{e}_0}k_0\frac{e_0/\hat{e}_0}{\sqrt{1-(e_0/\hat{e}_0)^2}}\frac{de_0}{\hat{e}_0} \tag{f}$$

Adding expressions (b), (c), and (f), we obtain Eq. (731).

Equation (731) expressed in terms of angle θ is

$$(I)_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2} \int_0^{\pi} \frac{\sin \phi \sin n\phi}{n} d\phi \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0 \sin \theta \, d\theta \qquad (732)$$

Equations (731) and (732) can be evaluated directly if the mathematical expression for k_0 is known. The method of evaluating Eq. (732), when k_0 is known only in graphical form, is illustrated in Fig. 291. The empirically determined curve for k_0, which is the variational conductance of the *whole* circuit, is

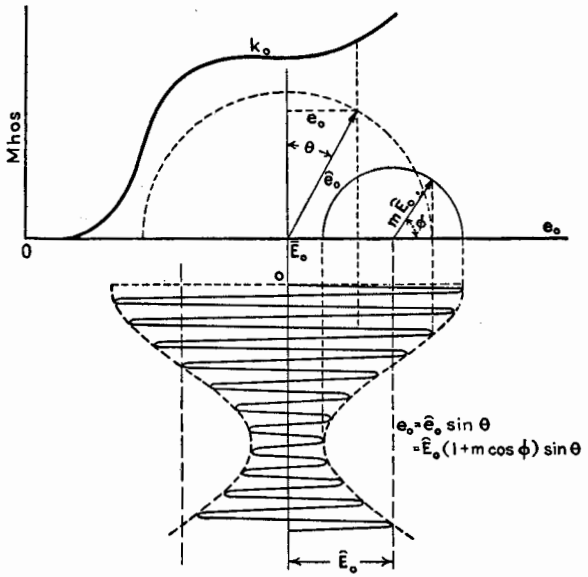


FIG. 291.—Calculation of the magnitude of the current of modulation frequency.

first plotted against the variable polarizing voltage used in the measurement of the conductance. The particular polarizing voltage $\bar{E}_0$, about which the operation takes place, is then marked on the voltage axis. The unmodulated amplitude $\hat{E}_0$ of the impressed voltage is then laid off from $\bar{E}_0$ and a semicircle with radius equal to $m\hat{E}_0$ is drawn as in Fig. 291. The radius of this semicircle makes the angle ϕ with the voltage axis. Choosing some value of ϕ, a second semicircle is drawn about point $\bar{E}_0$ with radius equal to $\hat{E}_0$ $(1 + m \cos \phi)$. The radius of this second semicircle makes the angle θ with the vertical line through $\bar{E}_0$. A curve is now drawn with θ as abscissas and the corresponding

values of $k_0 \sin \theta$, taken from the curve of k_0, as ordinates. Such a curve is illustrated in Fig. 292*a*. The area under this curve for the particular value of ϕ is then multiplied by $\frac{\sin \phi \sin n\phi}{n}$ and plotted vertically against the value of ϕ as abscissa. This process, when repeated for other values of ϕ, yields the curve of Fig. 292*b*. The area under the second curve multiplied by $\sqrt{2}\bar{E}_0 m/\pi^2$ gives the final result. The method is admittedly laborious but there seems to be no further

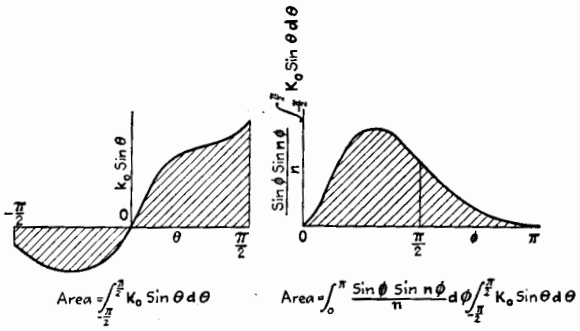


FIG. 292.—(*a*) The θ-integral for one value of ϕ; (*b*) the ϕ-integral for $n = 1$.

simplification possible. The ϕ-curve may usually be obtained by evaluating only a few of the θ-integrals.

Referring to Eq. (730), substitute for $\bar{\bar{I}} - \bar{I}$ the value given by Eq. (699). Then

$$(I)_{nl} = \frac{\sqrt{2}}{\pi^2} \int_0^{\pi} \cos n\phi \, d\phi \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \bar{\imath} \, d\theta \tag{733}$$

Equation (733) can be evaluated from a graph of $\bar{\imath}$ in a manner similar to the method described for evaluating Eq. (732). The labor involved is somewhat less in evaluating Eq. (733) than in calculating according to Eq. (732), especially if the angles for θ be chosen as multiples of $\pi/10$. But Eq. (732) has some advantages in accuracy and simplicity in form of the k_0 curve as compared to the curve for $\bar{\imath}$.

If, in Eq. (732), we substitute the value of k_0 given in Eq. (695) and simplify by integration by parts, an alternative expression is obtained for $(I)_{nl}$.

$$(I)_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\int_0^{\pi}\frac{\sin\phi\,\sin n\phi}{n}d\phi\int_{-\hat{e}_0}^{\hat{e}_0}\frac{dk_0}{de_0}\sqrt{1-\left(\frac{e_0}{\hat{e}_0}\right)^2}de_0 \qquad (734)$$

$$= \frac{\sqrt{2}\hat{E}_0^2 m}{\pi^2}\int_0^{\pi}\frac{(1+m\cos\phi)\,\sin\phi\,\sin n\phi}{n}d\phi\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{dk_0}{de_0}\cos^2\theta\,d\theta \qquad (735)$$

Equations (731) through (735) are useful for calculating the fundamental and harmonics of the modulation frequency produced by detection by a device in a circuit having the variational conductance k_0.

219. Two Sinusoidal Voltages of Nearly the Same Frequency Impressed in a Circuit Containing a Diode, the Amplitude of One Being Small in Comparison with the Amplitude of the Other. $\tilde{R} = \bar{R}$. Heterodyne Detection.—The case described in the heading and now to be examined is applicable mainly to heterodyne detection, which is characterized as follows: The voltage of a signal of single frequency and small amplitude is added to a locally produced sinusoidal e.m.f. of relatively large amplitude, differing only slightly in frequency from that of the signal. These two e.m.f.'s are impressed upon a nonlinear circuit. There is produced in the circuit a current of frequency equal to the difference of the two impressed frequencies. This method of beat reception was first suggested by Fessenden. As used in the superheterodyne receivers, now so common for broadcast reception, the signal comprises many frequencies forming the usual radio spectrum consisting of a carrier and side bands. The locally produced heterodyning e.m.f., with which the various components of the signal beat to produce the difference frequencies, differs in the superheterodyne from the carrier frequency by an amount of the order of 60,000 to 100,000 cycles per second. In the superheterodyne, the effect of heterodyning is to shift the whole radio spectrum without changing the relations among its components. This shift is to a lower frequency region and is produced by subtracting a constant frequency from all of the component frequencies. The theory to be given will be applied to heterodyne detection of a single frequency component.

Let $e_0' = \hat{E}_0' \sin \omega_h t$ be the relatively large e.m.f. produced locally. Let $\Delta e_0'' = \hat{\Delta E}_0'' \sin(\omega_h \pm \Delta\omega_h)t$ be the voltage of the received signal. The e.m.f. which is impressed upon the nonlinear circuit is their sum; or

$$e_0 = e_0' + \Delta e_0'' = (\hat{E}_0' + \widehat{\Delta E}_0'' \cos \Delta\omega_h t) \sin \omega_h t \pm \widehat{\Delta E}_0'' \sin \Delta\omega_h t \cos \omega_h t \tag{736}$$

$$= \sqrt{\hat{E}_0'^2 + (\Delta \hat{E}_0'')^2 + 2\hat{E}_0' \widehat{\Delta E}_0'' \cos \Delta\omega_h t} \sin (\omega_h t + \alpha) \tag{737}$$

where

$$\alpha = \tan^{-1} \frac{\pm \Delta \hat{E}_0'' \sin \Delta\omega_h t}{\hat{E}_0' + \widehat{\Delta E}_0'' \cos \Delta\omega_h t} \tag{738}$$

The phase angle α is a fluctuating angle and is small if $\Delta E_0''$ is small in comparison with $\hat{E}_0''$. Equation (737) shows that the resulting e.m.f. e_0 has a fluctuating amplitude given by the radical of Eq. (737). Calling this amplitude $\hat{E}_0$, we may develop the radical, obtaining the expression

$$\hat{E}_0 = \sqrt{\hat{E}_0'^2 + (\widehat{\Delta E}_0'')^2} + \frac{\hat{E}_0' \widehat{\Delta E}_0'' \cos \Delta\omega_h t}{\sqrt{\hat{E}_0'^2 + (\widehat{\Delta E}_0'')^2}} - \frac{\hat{E}_0'^2 (\widehat{\Delta E}_0'')^2 \cos^2 \Delta\omega_h t}{2(\hat{E}'^2 + (\widehat{\Delta E}_0'')^2)^{3/2}} + \cdots \tag{739}$$

$$= \sqrt{\hat{E}_0'^2 + (\widehat{\Delta E}_0'')^2} \left[1 + \frac{\hat{E}_0' \Delta \hat{E}_0''}{\hat{E}_0'^2 + (\widehat{\Delta E}_0'')^2} \cos \Delta\omega_h t - \frac{\hat{E}_0'^2 (\widehat{\Delta E}_0'')^2}{2(\hat{E}_0'^2 + (\Delta E_0''^2)^2)^2} \cos^2 \Delta\omega_h t + \cdots \right] \tag{740}$$

Equation (740) shows that the amplitude of the resulting modulation is nonsinusoidal, but that, if $(\widehat{\Delta E}_0'')^2$ is small in comparison with $\hat{E}_0'^2$, we may write

$$\hat{E}_0 = \hat{E}_0' \left(1 + \frac{\widehat{\Delta E}_0''}{\hat{E}_0'} \cos \Delta\omega_h t \right) \tag{741}$$

This expression gives the amplitude of a sinusoidally modulated e.m.f. having a degree of modulation equal to $\widehat{\Delta E}_0'' / \hat{E}_0'$. The theory just given for a sinusoidally modulated e.m.f. is applicable to heterodyne detection only when the signal voltage is small. The degree of modulation is given by the ratio of the signal voltage to the heterodyning voltage.

220. Two Sinusoidal Voltages of Widely Different Frequencies Impressed in a Nonlinear Circuit. $\tilde{R} = \bar{R}$. Modulation.— The conditions stated pertain to the process of modulation, the higher of the two frequencies being the carrier frequency and the lower the modulation frequency.

Let $(e_0)_h = (\hat{E}_0)_h \cos \omega_h t$ and $(e_0)_l = (\hat{E}_0)_l \cos \omega_l t$ be the two voltages impressed in series in the circuit which has a current-voltage characteristic given by $i = F(e_0)$. In this problem we are particularly interested in the resulting current of high frequency and in the currents of the sum and difference of frequencies.

The actual current wave form is given by substituting for e_0 the sum of the two voltages. Hence

$$i = F_1[(e_0)_h + (e_0)_l]$$

The analysis is made easy if it is assumed that the low frequency is contained an even number of times in the high frequency, or that $\omega_h = p\omega_l$, where p is a whole number. The amplitude of the carrier current is

$$(\hat{I})_h = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos p\phi \, d\phi \tag{742}$$

where

$$\phi = \omega_l t$$

Equation (742) is the same in form as Eq. (719), but $\bar{\imath}$ is a different function of ϕ.

Other forms of Eq. (742) are given below, including expressions for the higher harmonics of frequency $\omega_h/2\pi$, denoted by n.

$$(\hat{I})_{nh} = -\frac{2}{\pi}\int_0^\pi k_0 \frac{\sin np\phi}{np} \cdot \frac{de_0}{d\phi} d\phi \tag{743}$$

$$= \frac{2}{\pi}\int_0^\pi k_0 \frac{\sin np\phi}{np}[p(\hat{E}_0)_h \sin p\phi + (\hat{E}_0)_l \sin \phi] d\phi \tag{744}$$

The amplitude of the upper and lower side frequencies is given by

$$(\hat{I})_{h \pm l} = \frac{2}{\pi}\int_0^\pi \bar{\imath} \cos (p \pm 1)\phi \, d\phi \tag{745}$$

$$= -\frac{2}{\pi}\int_0^\pi k_0 \frac{\sin (p \pm 1)\phi}{p \pm 1} \cdot \frac{de_0}{d\phi} d\phi \tag{746}$$

$$= \frac{2}{\pi}\int_0^\pi k_0 \frac{\sin (p \pm 1)\phi}{p \pm 1} \cdot [p(\hat{E}_0)_h \sin p\phi + (\hat{E}_0)_l \sin \phi] d\phi \tag{747}$$

***221. When the Characteristic Curve of the Nonlinear Element Alone Is Known.**—The theory thus far presented in this chapter is expressed in terms of the characteristic curve of the

circuit as a whole, being given either as a current-voltage characteristic of the form $i = F(e_0)$, or in terms of the conductance plotted against voltage as expressed by $k_0 = F'(e_0)$. These characteristic curves can be obtained directly by experiment, or they can be derived, as will now be explained, from the characteristic curve of the nonlinear element *alone* if the resistance of the rest of the circuit is known.

Let $i = f(e)$ be the current-voltage characteristic of the nonlinear element alone. Suppose, as in Fig. 288, that this element is in series with a resistance R, which has the same value for

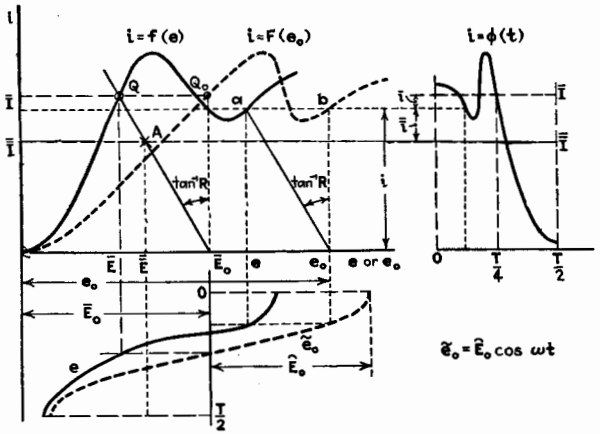


FIG. 293*a*.—Derivation of curve of $i = F(e_0)$ from curve of $i = f(e)$.

steady and for alternating currents of all the frequencies being considered. In Fig. 293*a*, $i = f(e)$ is shown by the full-line curve at the left. It is desired to derive the curve $i = F(e_0)$. Let e_0 be any voltage assumed to be impressed on the whole circuit. A straight line is drawn from point e_0 making an angle whose tangent is equal to R, as in Fig. 293*a*. The intersection of this resistance line with the curve $i = f(e)$, shown at point a in Fig. 293*a*, gives the current i that flows in the nonlinear element and the voltage e across it. The point b obtained by projection from a and e_0 gives a point on the desired curve $i = F(e_0)$. A repetition of this process for a number of assumed voltages e_0 gives the complete curve $i = F(e_0)$.

The quiescent points Q and Q_0 have similarly been determined in Fig. 293*a* for a particular impressed polarizing voltage $\bar{E}_0$. There is also shown an assumed impressed voltage $\tilde{e}_0 = \hat{E}_0 \cos \omega t$.

This particular impressed voltage acting in the circuit would give a current wave form shown by the curve $i = \phi(t)$ at the right in Fig. 293*a*, and also a voltage across the nonlinear element and shown by full-line curve marked e at the lower part of the figure. The average voltage $\bar{\bar{E}}$ and the average current $\bar{\bar{I}}$ are shown, giving the average point A, which must of necessity lie on the steady-current resistance line drawn from the steady voltage $\bar{E}_0$ but does not necessarily lie on the curve $i = F(e_0)$.

We shall now derive the curve of k_0 from the curve of conductance for the nonlinear element alone, *i.e.*, $k = f'(e)$. Referring

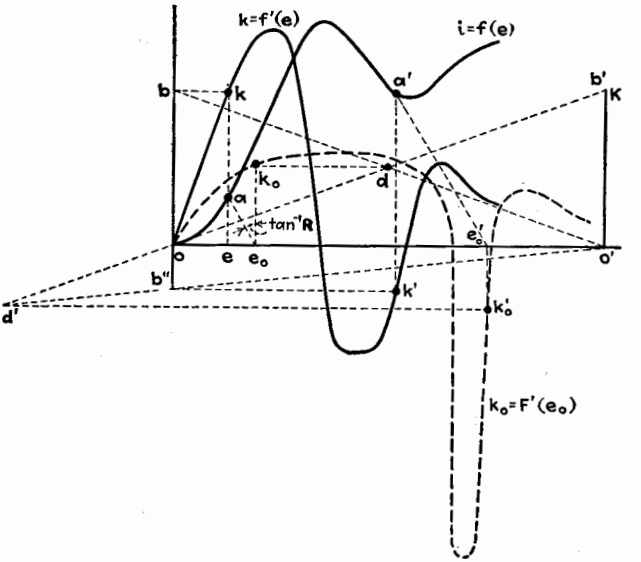


FIG. 293*b*.—Derivation of curve of $k_0 = F'(e_0)$ from curve of $k = f'(e)$.

to Fig. 293*b*, let the curve $k = f'(e)$ represent the slope of the curve $i = f(e)$ for the nonlinear element alone. The curve of k can be determined directly by a bridge method, as described in Chap. IX. From any point on the voltage axis representing some chosen value of e_0, a line is drawn making with the vertical an angle whose tangent is $R = 1/K$, Fig. 293*b*. The intersection with the curve $i = f(e)$ at a gives the voltage across the nonlinear device. Projecting vertically from a to the curve of k gives the corresponding value of k. Projecting horizontally to the left-hand axis determines point b. At any point O' on the horizontal axis,

a line $O'b'$ is erected whose length represents K. Point b' is then connected with O. Similarly, a slant line is drawn connecting O' with b. The intersection of these two oblique lines at d is distant from the horizontal axis by the amount k_0, because $k_0 = \frac{kK}{k+K}$ and the geometrical construction given above satisfies this relation. Point d, projected horizontally to a point over e_0, gives one point on the curve of k_0 as a function of e_0. This process, repeated a sufficient number of times, completes the curve. A second point is determined in Fig. 293*b*, for which k_0 is negative.

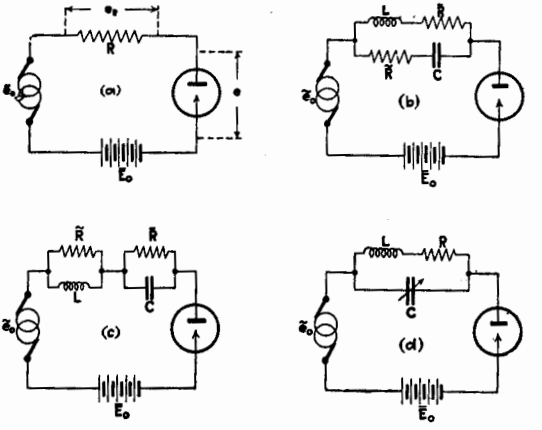


FIG. 294.—Circuits containing a nonlinear element.

222. When the Resistance R Is Different for Steady and Alternating Currents. $\tilde{R} \neq R$.—Up to this point, all of the theory has been given with the limitation that the load resistance R is independent of frequency. We shall now derive the path of operation when the load resistance has a different value for alternating current from that for steady current, as is true for circuits shown in Fig. 294. In the circuits of Fig. 294*b* and *c*, the capacitance C and the inductance L must be assumed to be so large that they have no appreciable reactance to the alternating currents. In Fig. 294*d*, the tuned circuit acts as a pure resistance to the resonance frequency but the harmonics of the resonance frequency must be assumed to be passed by C so that the harmonics meet with negligible reactance and resistance.

Referring to Fig. 294*a*, let e_R be the instantaneous voltage across the element R. Then

$$e_0 = \tilde{e}_0 + \bar{E}_0 = e_R + e \tag{748}$$

The instantaneous current i can be considered as made up of an average value $\bar{\bar{I}}$, and a periodic current $\bar{i}$ measured from the average value; or

$$i = \bar{\bar{I}} + \bar{i} \tag{749}$$

Similarly, e_R has an average value $\bar{\bar{E}}_R$ and a variable component about $\bar{\bar{E}}_R$ denoted by $\bar{e}_R$, so that

$$e_R = \bar{\bar{E}}_R + \bar{e}_R \tag{750}$$

Then

$$e_R = \bar{\bar{I}}\bar{R} + \bar{i}\tilde{R} \tag{751}$$

where $\tilde{R}$ is the resistance of the load to alternating current. Substituting Eq. (751) in Eq. (748) gives

$$e_0 = \bar{\bar{I}}\bar{R} + \bar{i}\tilde{R} + \psi(i) \tag{752}$$

where $e = \psi(i)$ is the inverse of $i = f(e)$. From Eq. (749), $\bar{i} = i - \bar{\bar{I}}$. Substituting this value of $\bar{i}$ in Eq. (752) gives

$$\begin{aligned} e_0 &= \bar{\bar{I}}(\bar{R} - \tilde{R}) + i\tilde{R} + \psi(i) \\ &= \psi_0(i, \bar{\bar{I}}) \end{aligned} \tag{753}$$

From Eq. (753) it is seen that the characteristic curve of the whole circuit is a function of i and also of $\bar{\bar{I}}$ and therefore is no longer the same for all amplitudes of electrical variation. For any given amplitude of $\tilde{e}_0$, the average current $\bar{\bar{I}}$ is constant, and the characteristic curve of the whole circuit is definite and fixed. When the amplitude of $\tilde{e}_0$ changes, $\bar{\bar{I}}$ changes and the characteristic curve of e_0, expressed as a function of i, shifts along the voltage axis by an amount equal to $(\bar{R} - \tilde{R})$ times the change in average current. This is as if the polarizing potential $\bar{E}_0$ were increased by this amount.

When i is expressed as a function of e_0 and $\bar{\bar{I}}$, or $i = F(e_0\bar{\bar{I}})$, the curve shifts, without change of shape, parallel to the voltage axis as $\bar{\bar{I}}$ varies. The amount of this shift, as just given, is $(\bar{\bar{I}} - \bar{I})(\bar{R} - \tilde{R})$. Evidently, therefore, the characteristic curve of the whole circuit remains fixed in position only if either $\bar{\bar{I}} = \bar{I}$ or $\bar{R} = \tilde{R}$.

The shift of the characteristic curve when $\bar{R} \neq \tilde{R}$, and when the curve has a kink so that the shift may better be demonstrated, is shown in Fig. 295. The horizontal distance between the two curves at the right is $(\bar{\bar{I}} - \bar{I})(\bar{R} - \tilde{R})$.

When $\tilde{e}_0$ is zero, the quiescent condition is expressed by the equation

$$\bar{E}_0 = \bar{I}\bar{R} + \bar{E} \tag{754}$$

Subtracting Eq. (754) from Eq. (753),

$$e_0 - \bar{E}_0 = (\bar{\bar{I}} - \bar{I})\bar{R} + (i - \bar{\bar{I}})\tilde{R} + \psi(i) - \bar{E} \tag{755}$$

Equation (755) is useful in suggesting the method of constructing the characteristic curve for the whole circuit for a given value

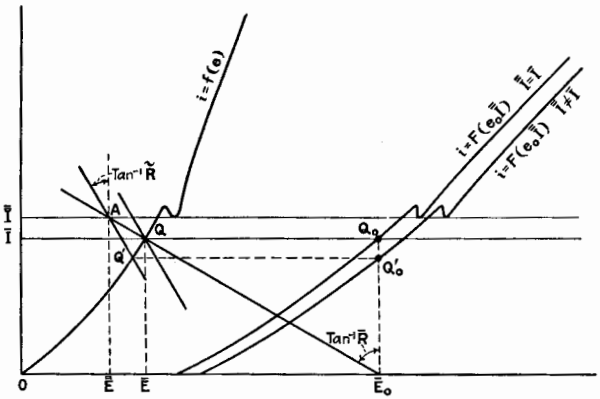


FIG. 295.—Curves of $i = F(e_0,\bar{\bar{I}})$ derived from curve of $i = f(e)$, showing transformation of a kink in the curve.

of $\bar{\bar{I}}$. Referring to Fig. 296, let $i = f(e)$ be the characteristic curve of the diode alone. From a point representing $\bar{E}_0$ on the voltage axis, the resistance line ab is drawn making an angle with the vertical whose tangent is $\bar{R}$. The intersection of this line with the curve $i = f(e)$ gives $\bar{I}$ and $\bar{E}$. Through a point A on the line ab determined by the average current $\bar{\bar{I}}$, a resistance line $a'b'$ is drawn, making an angle with the vertical whose tangent is $\tilde{R}$. Let e_0 be any instantaneous e.m.f. applied to the whole circuit. To find the instantaneous current corresponding to e_0, a line is drawn from e_0, parallel to the resistance line for $\bar{R}$, until it cuts the average current line $\bar{\bar{I}}$ at d. From this intersection point, a line is drawn parallel to the resistance line for $\tilde{R}$

until it intersects the curve $i = f(e)$ at g. This intersection point determines the instantaneous current i, and i and e_0 determine a point on the curve $i = F(e_0\bar{\bar{I}})$. This construction is justified by referring to Eq. (755) and noting that the left-hand side of this equation is equal to $\tilde{e}_0$ and is the same as the voltage corresponding to line $\overline{hg}$; that $(\bar{\bar{I}} - \bar{I})\bar{R}$ is equal to the voltage corresponding to $\overline{jk}$; that $(i - \bar{\bar{I}})\tilde{R}$ is equal to the voltage corresponding to $\overline{hj}$; and that $\psi(i) - \bar{E}$ is equal to the voltage corresponding to kg. The whole curve $i = F(e_0\bar{\bar{I}})$ can be found by a repetition

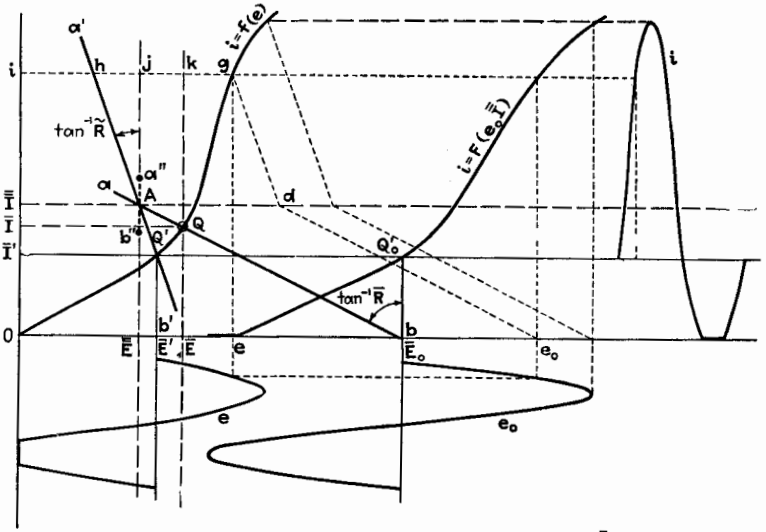


FIG. 296.—Method of the derivation of the curve of $i = F(e_0, \bar{\bar{I}})$ from the curve of $i = f(e)$ when $\bar{R} \neq \tilde{R}$.

of the process just given. The full-line curves show the current i and voltage e plotted against t. The method outlined was used in constructing the curves of Fig. 295.

***223. Special Case of Fig. 294*d*.**—Consider the special case of Fig. 294*d*. Although the parallel tuned circuit acts as a pure resistance at the resonance frequency $\omega_h/2\pi$, it acts as a capacitive reactance of small value at the harmonic frequencies. The voltage across the circuit and the current through the circuit are out of phase for the harmonic components. When plotted as in Fig. 296, the locus is an ellipse, instead of a straight line which it is for the fundamental. As a result, the path of operation $i = F(e_0\bar{\bar{I}})$ is a curve having a different path for ascend-

ing and descending values of i. As explained in the beginning, our theory cannot solve such problems. If, as is usually the case, the resistance and capacitive reactance to the harmonic components are practically negligible in comparison with the resistance $(R)_h$ to the fundamental frequency, the plot of current against voltage for the circuit when only the harmonics are considered is approximately a vertical line, $a''b''$, shown dotted in Fig. 296. Since the impedance of the load to the harmonic currents is assumed to be negligible, the harmonic components of current produce no voltage across the load. Hence, the alternating voltages both across the load and across the tube are sinusoidal.

To find the path of operation for this case, expand $\bar{i}$ of Eq. (749) into $i_h + \bar{i}'$, where i_h represents the sinusoidal current of frequency $\omega_h/2\pi$, and $\bar{i}'$ is used to signify all the other components except i_h, all being measured from the average value. Equation (751) becomes

$$e_R = \bar{I}\bar{R} + i_h(R)_h \tag{756}$$

and Eq. (752) becomes

$$e_0 = \bar{I}\bar{R} + i_h(R)_h + \psi(i) \tag{757}$$

Since

$$i_h = i - \bar{I} - \bar{i}'$$

Eq. (757) becomes

$$\begin{aligned} e_0 &= \bar{I}[\bar{R} - (R)_h] + (i - \bar{i}')(R)_h + \psi(i) \\ &= \psi_0(i, \bar{i}', \bar{I}) \end{aligned} \tag{758}$$

We may subtract Eq. (754) from Eq. (758), giving

$$e_0 - \bar{E}_0 = (\bar{I} - \bar{I})\bar{R} + i_h(R)_h + \psi(i) - \bar{E} \tag{759}$$

The characteristic curve for the whole circuit for this case must be constructed in accordance with Eq. (759) instead of Eq. (755).

The net result of the discussion of this section is that, whenever the load resistance is different for alternating and steady currents, the path of operation is not uniquely determined. It depends in shape and position upon the polarizing potential and upon the amplitude of the impressed voltage. It is then impossible to obtain exact solutions for the component currents when the

amplitude of the impressed voltage varies, as with a modulated voltage. The experimental method to be described in the next chapter gives results for any actual case but the theory presented here may help to interpret these experimental results and to help to an understanding of the factors upon which under these particular conditions the action of nonlinear devices depends.

However, this case is not entirely outside the bounds of approximate theoretical treatment. As stated, and as illustrated by Fig. 296, the path of operation shifts and changes shape with a variation of $\bar{\bar{I}}$. If, however, the curve of $i = f(e)$ has no sharp kinks, and if $\bar{\bar{I}} - \bar{I}$ is relatively small, as is usually the case, the shape of the path $i = F(e_0\bar{\bar{I}})$ traversed during operation changes but little when a variation of $\bar{\bar{I}}$ takes place, and we may deduce an approximate value of the quantity $\bar{\bar{I}} - \bar{I}$ in terms of the shape of this path, *i.e.*, in terms of k_0.

The curve of k_0 cannot easily be obtained experimentally. It can be obtained by the graphical method of Fig. 293*b*, provided the line ae_0 is laid off corresponding to the resistance $\bar{R}$ and the line $o'b'$ is made equal to $\bar{\bar{K}}$.

Referring again to Fig. 295, let $i = F(e_0\bar{\bar{I}})$ be the current curve corresponding to the average current $\bar{\bar{I}}$. Let Q_0' be the quiescent point as determined on the assumption that the curve $i = F(e_0\bar{\bar{I}})$ remains fixed in the position it has when the impressed alternating voltage is zero. If $\bar{I}'$ is the current corresponding to Q_0', the change of current $\bar{\bar{I}} - \bar{I}'$ is the quantity determined by the shape of the curve $i = F(e_0\bar{\bar{I}})$, whereas the actual observed change of current is $\bar{\bar{I}} - \bar{I}$. We wish, therefore, to find $\bar{\bar{I}} - \bar{I}$ in terms of $\bar{\bar{I}} - \bar{I}'$.

Let $\bar{\bar{E}}$, $\bar{E}$, and $\bar{E}'$ be the voltages across the diode corresponding to the points A, Q, and Q'. Then,

$$\bar{E} - \bar{\bar{E}} = (\bar{E} - \bar{E}') + (\bar{E}' - \bar{\bar{E}}) \tag{760}$$

Substituting for these differences in voltage their equals in terms of currents and resistances, Eq. (760) becomes

$$(\bar{\bar{I}} - \bar{I})\bar{R} = (\bar{I} - \bar{I}')r + (\bar{\bar{I}} - \bar{I}')\bar{R} \tag{761}$$

where r is the variational resistance of the diode alone at the quiescent point.

Rearranging the terms of Eq. (761),

$$(\bar{\bar{I}} - \bar{I})(\bar{R} + r) = (\bar{\bar{I}} - \bar{I}')(\bar{R} + r) \tag{762}$$

All preceding equations giving $\bar{\bar{I}} - \bar{I}$, or derived from this change of current, can now be extended to include the case wherein $\bar{R} \neq \tilde{R}$ by noting that the equations really determine the quantity $\bar{\bar{I}} - \bar{I}'$, and that the actual change $\bar{\bar{I}} - \bar{I}$ is obtained by multiplying $\bar{\bar{I}} - \bar{I}'$ by the factor $\frac{\tilde{R} + r}{\bar{R} + r}$. Hence, the more general forms for Eqs. (706), (707), (731), (732), (734), and (735) are given below.

$$\bar{\bar{I}} - \bar{I} = \frac{\tilde{R}+r}{\bar{R}+r}\left[\frac{1}{2}\int_0^{\hat{E}_0} k_0 de_0 + \frac{1}{2}\int_0^{-\hat{E}_0} k_0 de_0 - \frac{1}{\pi}\int_{-\hat{E}_0}^{\hat{E}_0} k_0\theta\, de_0\right] \quad (706)\quad (763)$$

$$= \frac{\tilde{R}+r}{\bar{R}+r}\left[\frac{1}{2}\int_0^{\hat{E}_0} k_0 de_0 + \frac{1}{2}\int_0^{-\hat{E}_0} k_0 de_0 - \frac{\hat{E}_0}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0\theta\cos\theta\, d\theta\right] \quad (707)\quad (764)$$

For modulated voltage

$$(\bar{\bar{I}})_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\cdot\frac{(R)_h + r}{(R)_l + r}\int_0^{\pi}\frac{\sin\phi\sin n\phi}{n}d\phi\int_{-\hat{e}_0}^{\hat{e}_0} k_0\frac{e_0/\hat{e}_0}{\sqrt{1-(e_0/\hat{e}_0)^2}}d\left(\frac{e_0}{\hat{e}_0}\right) \quad (731)\quad (765)$$

$$= \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\cdot\frac{(R)_h + r}{(R)_l + r}\int_0^{\pi}\frac{\sin\phi\sin n\phi}{n}d\phi\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_0\sin\theta\, d\theta \quad (732)\quad (766)$$

or

$$(\hat{I})_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\cdot\frac{(R)_h + r}{(R)_l + r}\int_0^{\pi}\frac{\sin\phi\sin n\phi}{n}d\phi\int_{-\hat{e}_0}^{\hat{e}_0}\frac{dk_0}{de_0}\sqrt{1-\left(\frac{e_0}{\hat{e}_0}\right)^2}\,de_0 \quad (734)\quad (767)$$

$$= \frac{\sqrt{2}\hat{E}_0 m}{\pi^2}\cdot\frac{(R)_h + r}{(R)_l + r}\int_0^{\pi}\frac{\sin\phi\sin n\phi}{n}d\phi\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{dk_0}{de_0}\cos^2\theta\, d\theta \quad (735)\quad (768)$$

In Eqs. (765) through (768), $(R)_h$ is used to denote the resistance for the high frequency, and $(R)_l$ to denote the resistance for the low or modulation frequency and is assumed to be equal to $\bar{R}$.

III. APPLICATION OF THE THEORY TO TRIODES

The theory of the operation of triodes with an impressed voltage of large amplitude is obtained by a simple extension of the theory presented for the diode. It is necessary to keep in mind, as explained in Chap. XX, that the nonlinear characteristic may be the curve of plate current plotted against grid voltage or of grid current plotted against grid voltage. In discussing the operation of the triode, refer to the simple diagram of connections shown in Fig. 297.

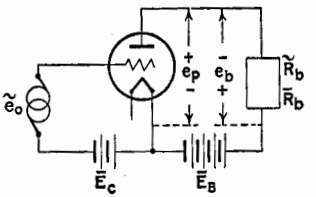


FIG. 297.—A triode as a nonlinear element.

224. Operation of Triode with Nonlinear Plate-current Curve. $\bar{R}_b = \tilde{R}_b$.—Consider the case of a triode having a pure resistance R_b in the plate circuit, the value of which is the same for alternating and steady currents. Assume the triode to be operated so that the nonlinearity of its characteristic is due entirely to the

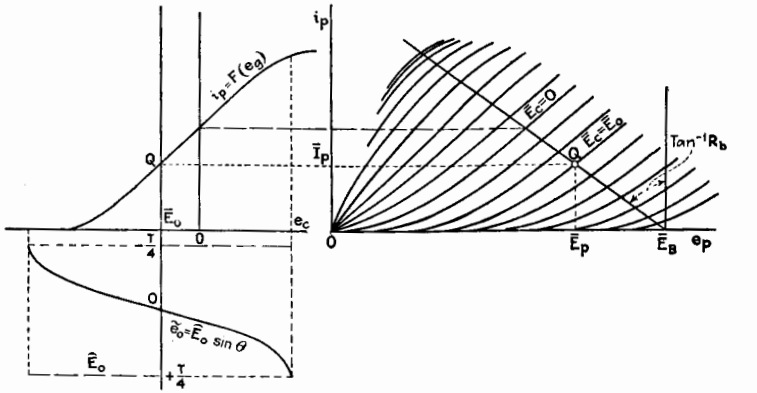


FIG. 298.—The characteristic curve for the plate circuit of a triode having a pure resistance as plate load.

curvature of the plate-current curve. This condition is fulfilled if e_g is always negative. Let the impressed grid voltage e_g which is equal to e_0 be $\tilde{e}_0 + \bar{E}_c$, as indicated in Fig. 297.

The curve of

$$i_p = F(e_g) = F(e_0) \tag{769}$$

for a particular resistance R_b can be obtained by the graphical method indicated in Fig. 298 or directly by experiment. This curve corresponds to the curve $i = F(e_0)$ for the diode.

The curve of the equivalent transconductance, s_{p0}, can readily be obtained only by dynamic measurement in the laboratory and is expressed in the form

$$s_{p0} = F'(e_g) = F'(e_0) \tag{770}$$

If we assume the impressed e.m.f. is

$$\tilde{e}_0 = \hat{E}_0 \sin \omega t = \hat{E}_0 \sin \theta \tag{771}$$

all equations from Eq. (699) through Eq. (716) apply to this case if for $\bar{\imath}$ we substitute $\bar{\imath}_p = F_1(e_0)$ obtained from Eq. (769), and for k_0 we substitute s_{p0} from Eq. (770).

If the impressed grid voltage is modulated in accordance with Eq. (728), all equations from Eq. (718) through Eq. (735) hold, provided $\bar{\imath}_p$ and s_{p0} are substituted for their corresponding quantities.

Similarly, Eqs. (742) through (747) hold with the same substitution.

225. Path of Operation of a Triode When $\bar{R}_b \neq \tilde{R}_b$.—In Sec. 222 the path of operation for a diode was studied, the diode being connected in series with a resistance whose value is different for alternating and steady currents. Explanation was given of the fact that the path in that case is not uniquely determined but depends upon the amplitude of the impressed voltage. Similarly, for a triode having this character of resistance in the plate circuit, the path of operation varies as the amplitude of the impressed grid voltage varies. Therefore, one of the assumptions underlying the theory presented in this chapter is violated, namely, that the current for any particular instantaneous impressed voltage shall be single valued. The theory consequently does not give accurate results for this case but, as with the diode, an approximately correct result can be obtained if the plate-current curve has no sharp kinks and if the difference between the average and quiescent plate currents is relatively small.

At first, we shall derive the path of operation for a particular value of impressed alternating voltage and for the corresponding value of $\bar{\bar{I}}_p$. Referring to Fig. 299, the resistance line ab for $\bar{R}_b$ is first laid off. Through the average point A, determined by the intersection of the average-current line $\bar{\bar{I}}_p$ and the resistance line

for $\bar{R}_b$, the resistance line $a'b'$ for $\tilde{R}_b$ is drawn, as shown in the figure. Then the line $a'b'$ is the path of operation, the ends of which are determined by the grid-voltage variation. At the left in Fig. 299 are shown the path of operation if $\bar{\bar{I}}_p = \bar{I}_p$, the path corresponding to the line $a'b'$, and the path for slow variations of grid potential corresponding to line ab.

The treatment for obtaining the correction factor to be applied in a diode is applicable to the present case. The correction factor is $\dfrac{(R_b)_h + r_p}{(R_b)_l + r_p}$, where r_p is the plate variational resistance at the plate-circuit quiescent point. The formulas of Eq. (763)

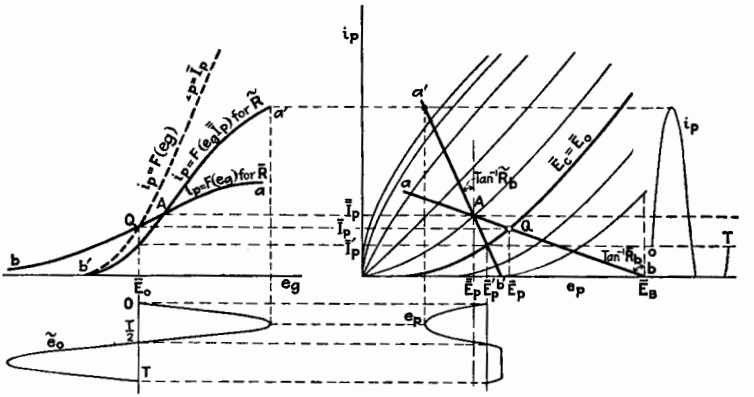


FIG. 299.—Curves for the path of operation for the plate circuit of a triode when $\bar{R}_b \neq \tilde{R}_b$.

through Eq. (768) may now be used with this correction factor in place of $\dfrac{\tilde{R} + r}{\bar{R} + r}$ and with s_{p0} substituted for k_0.

The expression for the current of modulation frequency and its harmonics, produced when a modulated e.m.f. expressed by Eq. (728) is impressed in the grid circuit, is given for reference.

$$(\hat{I}_p)_{nl} = \frac{\sqrt{2}\hat{E}_0 m}{\pi^2} \cdot \frac{(R_b)_h + r_p}{(R_b)_l + r_p} \int_0^{\pi} \frac{\sin\phi \sin n\phi}{n} d\phi \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} s_{p0} \sin\theta \, d\theta \tag{772}$$

226. Operation of a Triode When the Nonlinear Characteristic Is in the Grid Circuit Only.—This corresponds to a triode with a resistance R_c in the grid circuit, by-passed perhaps by a condenser, the polarizing potentials in the grid and plate circuits

having such values that the plate-current curve plotted against grid voltage is essentially straight. We may consider that the action is that of a nonlinear diode, the voltage across the diode being amplified by a linear amplifier.

If a modulated e.m.f., such as is expressed by Eq. (728), is impressed in the grid circuit, the expression for the current of modulation frequency and its harmonics is

$$(\hat{I}_p)_{nl} =$$

$$\frac{\sqrt{2}\hat{E}_0 m}{\pi^2} \cdot \frac{(R_c)_h + r_g}{(R_c)_l + r_g} \cdot \frac{u_p(R_c)_l}{(R_b)_l + r_p} \int_0^{\pi} \frac{\sin \phi \sin n\phi}{n} d\phi \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k_{g0} \sin \theta \, d\theta \tag{773}$$

227. Operation of a Triode When the Nonlinear Characteristics of Both Grid and Plate Circuits Are Effective.—This is obviously the most complicated case of all, especially if the resistances in the grid and plate circuits are different for the high and low frequencies.

First, assume that the external resistances in the grid and plate circuits are independent of frequency. If the plate current is plotted against the voltage impressed in the grid circuit, the effect of all resistances and of nonlinearity in both circuits of the triode is included. The slope of this curve is the over-all transconductance and the formulas for a diode are applicable if this over-all transconductance s_{p0} is substituted for k_0.

If the complicated case arises in which both external resistances are different for the low and high frequencies, an approximate result is obtained by adding the effects as expressed by the sum of Eqs. (772) and (773).

IV. EXAMPLES OF APPLICATION OF THEORY

A few examples are now given in order to illustrate the application of the theory presented. In the first and third examples an assumed empirical equation is used for the nonlinear characteristic curve, while in the second example the graphical method is followed.

228. Example 1.—The first example illustrates the application of the formulas of Eqs. (706) to (715). Assume that some form of two-electrode rectifier is connected in series with a source of sinusoidal e.m.f. and with a pure resistance. Suppose that the current is zero for negative voltages applied to the rectifier, but

that for positive voltages the current curve is parabolic in shape and is given by the expression

$$i = ae_0 + \frac{be_0^2}{2} \text{ (for positive values of } e_0) \tag{774}$$

The conductance is given by the equation

$$k_0 = a + be_0 \text{ (for positive values of } e_0) \tag{775}$$

The impressed voltage is assumed to be

$$\tilde{e}_0 = \hat{E}_0 \sin \omega t \tag{776}$$

which is the same as Eq. (696), page 538.

In order to calculate $\bar{\bar{I}} - \bar{I}$ we may use either Eq. (699) or Eq. (706). If we use Eq. (699),

$$\bar{\bar{I}} - \bar{I} = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left(a\hat{E}_0 \sin\theta + \frac{b}{2}\hat{E}_0^2 \sin^2\theta \right) d\theta \tag{777}$$

giving after integration

$$\bar{\bar{I}} - \bar{I} = \frac{a\hat{E}_0}{\pi} + \frac{b\hat{E}_0^2}{8} \tag{778}$$

If we use Eq. (706),

$$\bar{\bar{I}} - \bar{I} = \frac{1}{2}\int_0^{\hat{E}_0} (a + be_0) de_0 + \frac{1}{2}\int_0^{-\hat{E}_0} 0 de_0 - \frac{\hat{E}_0}{\pi}\int_0^{\frac{\pi}{2}} \theta(a + b\hat{E}_0 \sin\theta)\cos\theta \, d\theta \tag{779}$$

which after integration gives Eq. (778) as before.

The fundamental current is given by Eq. (700) or Eq. (710). Both equations necessarily give the same result, but the integration of Eq. (710) is much simpler.

$$(\hat{I})_1 = \frac{2\hat{E}_0}{\pi}\int_0^{\frac{\pi}{2}} (a + b\hat{E}_0 \sin\theta)\cos^2\theta \, d\theta \tag{780}$$

$$= \frac{\hat{E}_0 a}{2} + \frac{2b\hat{E}_0^2}{3\pi} \tag{781}$$

Equation (713) gives for the second harmonic

$$(\hat{I})_2 = -\frac{2\hat{E}_0}{\pi}\int_0^{\frac{\pi}{2}} (a + b\hat{E}_0 \sin\theta)\sin\theta\cos^2\theta \, d\theta \tag{782}$$

$$= \frac{2a\hat{E}_0}{3\pi} - \frac{\hat{E}_0^2 b}{8} \tag{783}$$

228a. Example 2.—This example is the same as Example 1, except that k_0, instead of being expressed analytically, is given graphically by the full-line curve of Fig. 300. The method of solution has been described in connection with Fig. 290. The results for the fundamental and for the second and third harmonics are given in Fig. 300.

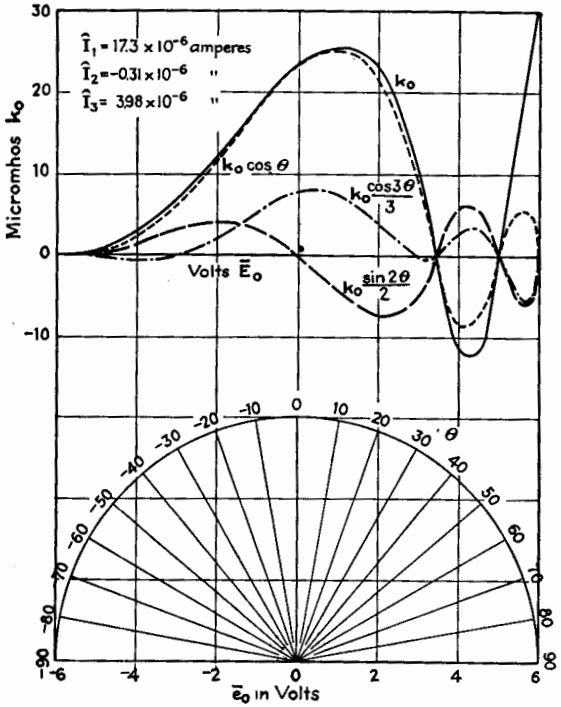


FIG. 300.—Method of calculating the magnitudes of the fundamental and harmonic components of current from the curve of k_0. Example 2.

229. Example 3.—In this example we shall assume that the circuit is the same as in Example 1, and that the current characteristic is of the same form as in the first example. The impressed e.m.f. is assumed to be modulated and given by Eq. (728). The problem is to obtain the current of modulation frequency. The result is given by integrating either Eq. (721) or Eq. (732). Using Eq. (732),

$$(I)_1 = \frac{\hat{E}_0 m}{\sqrt{2}\pi}\left(a + \frac{b\hat{E}_0\pi}{4}\right) + \frac{\sqrt{2}\hat{E}_0^2 m^2 b}{6\pi} \qquad (784)$$

CHAPTER XXII

EXPERIMENTAL TREATMENT OF DETECTION AT HIGH SIGNAL VOLTAGES

When the electrical variations impressed upon a nonlinear element are so large that the portion of the nonlinear characteristic curve traversed by the operating point cannot be represented by a series development containing terms no higher than the second degree, the mathematical theory becomes difficult and unwieldy. The small-signal theory presented in earlier chapters fails completely. The mathematical theory given in Chap. XXI can be applied, but generally the experimental method is much more satisfactory for solving nonlinear circuits with large electrical variations.

In this chapter, the experimental method of analysis will be applied to detection or rectification at high signal voltages. The treatment will be divided into two parts according as the nonlinear characteristic is in the plate circuit or in the grid circuit of a triode.

230. Plate-circuit Detection.—In general terms, detection is a kind of rectification in that it depends upon the change in the average value of a current when a sinusoidal e.m.f. acts upon the detector. Rectification is due to curvature of the characteristic curve of the nonlinear element. The most direct way of determining the amount of detection or of rectification is by measuring experimentally the change in steady current caused by a sinusoidal impressed voltage.

If we consider a triode operating with a negative polarizing voltage in the grid circuit so that there is no grid current, the rectification for high values of impressed voltage can be expressed as the change in steady plate current divided by the change in amplitude of the impressed alternating grid voltage. For a triode, rectification as defined is known as *transrectification.* If there is no appreciable time lag in the detector, and if the external load in the plate circuit has zero impedance at the frequency in question, transrectification is a function of $\bar{E}_c$, $\bar{E}_B$,

and $(\widehat{E}_g)_h$ only, $(\widehat{E}_g)_h$ being the amplitude of the impressed sinusoidal grid voltage. A diagram known as the *transrectification diagram*[1] for any particular triode can be obtained by plotting the average plate current $\bar{I}_p$ against plate-battery voltage $\bar{E}_B$. Several curves may be plotted, each for a constant value of $(\widehat{E}_g)_h$. The potential $(\widehat{E}_g)_h$ used for obtaining the curves may have any frequency, a convenient one being 60 cycles per second.

Figure 301 is a transrectification diagram for a type 201A triode with a grid-polarizing potential of -9 volts. Figure 302 is a similar diagram for a type 171*A* triode when $\bar{E}_c$ is -22.5

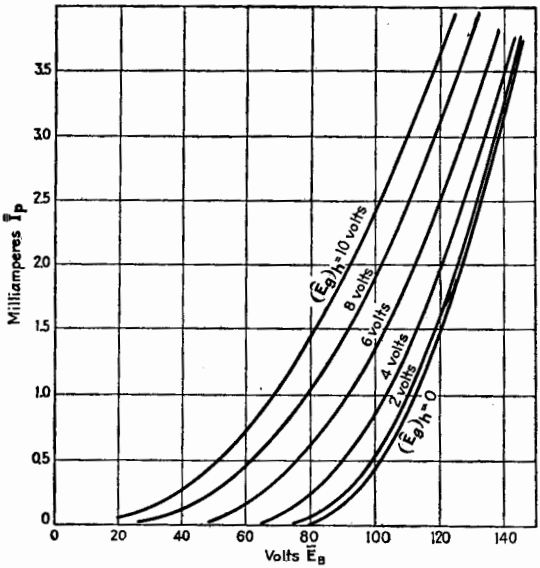


FIG. 301.—Plate-circuit detection characteristics or transrectification diagram. Triode ER201A. $\bar{E}_c = -9$ volts.

volts. The lower curve in each diagram is for $(\widehat{E}_g)_h = 0$ and is, therefore, the static characteristic curve of the triode for the particular value of $\bar{E}_c$ used in taking the curves. The successive curves for values of $(\widehat{E}_g)_h$ which differ by 2 volts are not equally spaced but are closer for small values of $(\widehat{E}_g)_h$. Curves for values of $(\widehat{E}_g)_h$ greater than $\bar{E}_c$ are not shown because grid current would flow for such curves.

[1] The transrectification diagram was suggested by Ballantine and independently by Smith; Ballantine, *Proc. I.R.E.*, **17**, 1153 (1929); Smith, *Proc. I.R.E.*, **15**, 525 (1927).

Referring to Fig. 301, we may plot the change in average plate current caused by various values of $(\bar{E}_g)_h$, for a constant $\bar{E}_B$ of 100 volts, for example. Such a curve is shown in Fig. 303. This curve shows that the change in average plate current increases approximately as the square of $(\bar{E}_g)_h$, for values of $(\bar{E}_g)_h$ from zero to about 6 volts. Thereafter it increases approximately linearly with $(\bar{E}_g)_h$. The first region is that of *square-law detection* where the *small-signal* theory of detection applies.

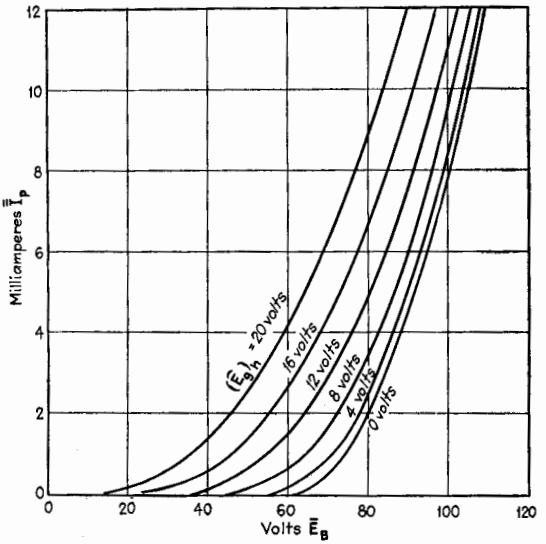


FIG. 302.—Plate-circuit detector characteristics or transrectification diagram. Triode ER171A. $\bar{E}_c = -22.5$ volts.

The slope of the curve of Fig. 303 gives the value of transrectification when $\bar{E}_B$ (which is the same as $\bar{E}_p$ in this case) is constant. The analytical definition of transrectification is

$$\text{Transrectification} = \left(\frac{\partial \bar{\bar{I}}_p}{\partial (\bar{E}_g)_h}\right)_{\bar{E}_p\text{-const.}} \tag{785}$$

The transrectification derived from the curve of change in average plate current in Fig. 303 is also shown plotted to the right-hand scale in the figure. In the region where the transrectification plotted against $(\bar{E}_g)_h$ is a sloping straight line, square-law detection takes place and the characteristic curve of the triode is expressed by a second-degree equation. In this region the small-

signal theory holds. Where the plot of transrectification is horizontal, linear detection takes place. Linear detection exists when the amplitude of the output wave is linearly related to the amplitude of the envelope of the modulated carrier wave. Linear detection is therefore distortionless.

The transrectification diagram, illustrations of which are given in Figs. 301 and 302, can be used to explain the operation of a detector when a modulated signal is impressed on the grid, and when the plate circuit contains a load for the current of modulation frequency. For simplicity, we shall assume at first

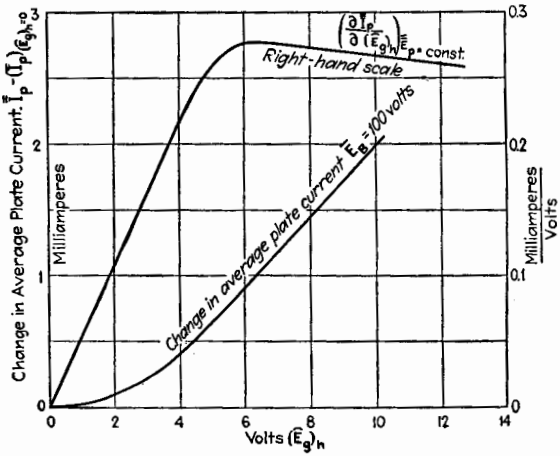


FIG. 303.—Change in average plate current and transrectification derived from Fig. 301.

that this load is a pure resistance for the modulation frequency $\omega_l/2\pi$ but is by-passed for the current of high or carrier frequency. Figure 304 is the same as Fig. 301, with the addition of the resistance line for $(R_b)_l$ drawn from the point on the voltage axis given by the plate-battery voltage. For illustration, the resistance line in Fig. 304 is drawn for a resistance of 40,000 ohms.

If $(\widehat{E}_g)_h$ is zero, the plate current is $\bar{I}_p$ and the plate voltage is $\bar{E}_p$, both obtained by projection from the point Q. If a modulated signal is impressed having an average amplitude of $(\widehat{E}_g)_h$ for $m = 0$, the average value of the average plate current is ${}_{av.}(\bar{\bar{I}}_p)$, and the average value of the average plate voltage is ${}_{av.}(\bar{\bar{E}}_p)$. This change in average plate voltage is due to the

increased voltage drop through $(R_b)_l$. The average quiescent point is $Q_{av.}$. When the amplitude of the carrier grid voltage varies sinusoidally, the operating point which determines average values moves over the resistance line and determines the fluctuations in average plate current. The diagram of Fig. 304, giving average values, is used therefore to describe the operation of the detector in the same way that the usual static-characteristic-curve diagram is used to describe the operation of an amplifier. If the rectification curves are equally spaced on either side of the $Q_{av.}$ point, sinusoidal variation of the envelope

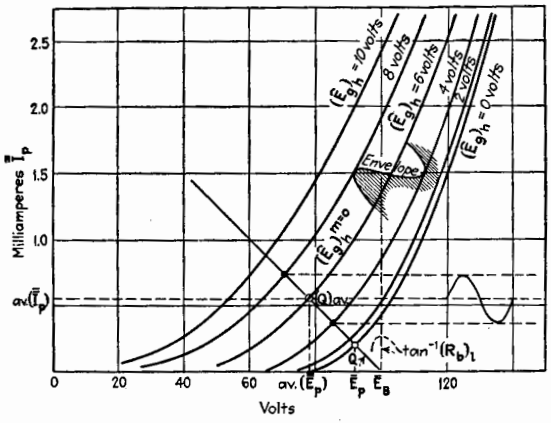


FIG. 304.—Rectification diagram and resistance line.

of the carrier voltage gives sinusoidal variation of the average plate current, and *no* distortion is introduced by detection. If the plate load is reactive for the currents of frequency $\omega_l/2\pi$, the path of operation is an ellipse about the point $Q_{av.}$, provided the rectification lines are straight, parallel, and equidistant.

An analysis of rectification follows which parallels the simple analysis given in Chap. XX for simple increments of currents and voltages. We may write for the rectification characteristics

$$\bar{\bar{I}}_p = f(\bar{\bar{E}}_p(\hat{E}_g)_h) \tag{786}$$

if $\bar{E}_c = \bar{E}_g$ is held constant.

Then

$$d\bar{\bar{I}}_p = \frac{\partial \bar{\bar{I}}_p}{\partial \bar{\bar{E}}_p} d\bar{\bar{E}}_p + \frac{\partial \bar{\bar{I}}_p}{\partial (\hat{E}_g)_h} d(\hat{E}_g)_h \tag{787}$$

The partial derivatives of Eq. (787) may be defined as follows:

$$\left(\frac{\partial \bar{\bar{I}}_p}{\partial \bar{\bar{E}}_p}\right)_{(\bar{\bar{E}}_g)_h\text{-const.}} = \bar{\bar{k}}_p \tag{788}$$

where $\bar{\bar{k}}_p$ means the *plate-circuit variational conductance for average values.* The reciprocal of $\bar{\bar{k}}_p$ is $\bar{\bar{r}}_p$, the *variational resistance for average values.* We may define the change in average plate current divided by the change in radio-frequency amplitude as $\bar{\bar{s}}_p$; or

$$\left(\frac{\partial \bar{\bar{I}}_p}{\partial (\bar{\bar{E}}_g)_h}\right)_{\bar{\bar{E}}_p\text{-const.}} = \bar{\bar{s}}_p \tag{789}$$

The quantity $\bar{\bar{s}}_p$ is analogous to the variational transconductance and may be defined as the *transconductance for rectification.* It is the same as the transrectification defined in Eq. (785).

If $d\bar{\bar{E}}_p$ and $d(\bar{\bar{E}}_g)_h$ vary so as to make $d\bar{\bar{I}}_p$ equal to zero,

$$-\left(\frac{\partial \bar{\bar{E}}_p}{\partial (\bar{\bar{E}}_g)_h}\right)_{\bar{\bar{I}}_p\text{-const.}} = \frac{\bar{\bar{s}}_p}{\bar{\bar{k}}_p} = \bar{\bar{u}}_p \tag{790}$$

By analogy with the fundamental theory of triodes, the quantity defined in Eq. (790) may be called the *voltage ratio for rectification* and is denoted by $\bar{\bar{u}}_p$.

The numerical value of $\bar{\bar{k}}_p$ can be measured dynamically, but for calculation purposes, it is generally sufficiently accurate to find it from the slope of the curve of rectification at the point in question. For illustration, in Fig. 304 at the $Q_{\text{av.}}$ point corresponding to an average plate voltage of 76 volts, the value of $\bar{\bar{k}}_p$ is 28.6 micromhos. The corresponding value of $\bar{\bar{r}}_p$ is 35,000 ohms. The values of $\bar{\bar{k}}_p$ and $\bar{\bar{r}}_p$ are essentially constant along the resistance line shown in Fig. 304 from $(\bar{\bar{E}}_g)_h = 4$ volts to $(\bar{\bar{E}}_g)_h = 8$ volts.

The numerical value of $\bar{\bar{s}}_p$ at any point on the rectification diagram can be found by scaling off on the diagram the proper increments. The value of $\bar{\bar{s}}_p$ at point $Q_{\text{av.}}$ of Fig. 304 is 196 micromhos.

The numerical value of $\bar{\bar{u}}_p$ can be found by scaling off the proper increments on the rectification diagram, or it can be calculated from the relation $\bar{\bar{s}}_p = \bar{\bar{u}}_p \bar{\bar{k}}_p$. The value of $\bar{\bar{u}}_p$ is 6.85 at point $Q_{\text{av.}}$ of Fig. 304.

If the amplitude of the carrier voltage varies sinusoidally about the value corresponding to the point $Q_{\text{av.}}$ $[(\bar{\bar{E}}_g)_h = 6$ volts in

Fig. 304], and if the variation is such that the path of operation is on an essentially plane portion of the characteristic surface constructed with $\bar{\bar{I}}_p$, $\bar{\bar{E}}_p$, and $(\hat{E}_g)_h$ as coordinates, the average plate current varies sinsusoidally. The r.m.s. value of the plate current can be calculated as explained below.

Without going through the steps of the analysis, it is obvious that we can set up an e-p-c. theorem for rectification analogous to the e-p-c. theorem for small increments. Its mathematical formulation is

$$(\Delta\bar{\bar{\mathbf{I}}}_p)_l = \frac{\bar{\bar{u}}_p(\Delta(\hat{\mathbf{E}}_g)_h)_l}{\bar{\bar{r}}_p + (\mathbf{Z}_b)_l} \tag{791}$$

In Eq. (791), $(\Delta\bar{\bar{\mathbf{I}}}_p)_l$ is the r.m.s. value of the variations of average plate current measured from the average point $Q_{\text{av.}}$. In this equation and in those that follow in this chapter, the Δ does not necessarily imply small variations but implies merely that the quantity is measured from $Q_{\text{av.}}$ as origin. The voltage $(\Delta(\hat{\mathbf{E}}_g)_h)_l$ indicates the r.m.s. value of the variation of $(\hat{E}_g)_h$ about the value which determines $Q_{\text{av.}}$.

If, owing to modulation of degree m, $(\hat{E}_g)_h$ varies sinusoidally, then $(\Delta(\hat{\mathbf{E}}_g)_h)_l$ is equal to m times the r.m.s. value of the *unmodulated* carrier amplitude, which is indicated as $((E_g)_h)_{m=0}$. Equation (791) may then be written

$$(\Delta\bar{\bar{\mathbf{I}}}_p)_l = \frac{m\bar{\bar{u}}_p((E_g)_h)_{m=0}}{\bar{\bar{r}}_p + (\mathbf{Z}_b)_l} \tag{792}$$

The quantity $m\bar{\bar{u}}_p\ ((E_g)_h)_{m=0}$ is the *fictitious voltage of detection* introduced into the plate circuit, the resistance of the plate circuit being $\bar{\bar{r}}_p$.

Referring to Fig. 304, it is clear that the path of operation can be placed in the region where the graphs of constant $(\hat{E}_g)_h$ are straight and parallel only by the application of very high values of $\bar{E}_B$ and by the use of a low value of $(R_b)_l$. A low value of $(R_b)_l$ gives a small output voltage. In the particular case of the triode characteristics shown in Fig. 304, a value of $\bar{E}_B$ equal to 150 volts and the value of $(R_b)_l$ of 40,000 ohms would be an improvement over the use of an $\bar{E}_B$ of 100 volts shown in the diagram. However, if the intersections of the curves for equal increments of $(\hat{E}_g)_h$ with the resistance line are equally spaced over a region, no distortion results even though the graphs of constant $(\hat{E}_g)_h$ are curved, and in such a case, $\bar{\bar{u}}_p$

and $\bar{r}_p$ in Eq. (792) are not constant, but one varies over the path of operation in such a manner as to compensate for the variation of the other. While such a condition exists and distortion is practically absent when a pure resistance load is used, distortion is not absent when an inductive load is used. This is true because an inductive load causes the path of operation to be an ellipse, and the variations of $\bar{u}_p$ could not compensate for the variation of $\bar{r}_p$ over the entire elliptic path.

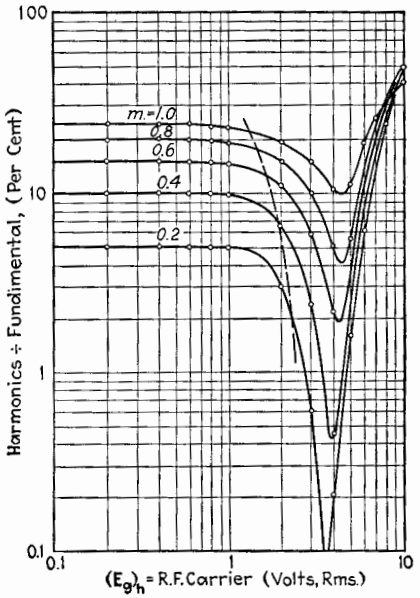


FIG. 305.—Distortion in plate-circuit detection. (*Ballantine.*) Triode 201A. $\bar{E}_B = 90$ volts; $\bar{E}_c = -4$ volts; $R_b = 100{,}000$ ohms; $w_1/2\pi = 400$ cycles per second.

From these considerations, better quality of detection can usually be obtained with a resistance plate load.

Since harmonics are always introduced to some degree at least, it is convenient to have some measure of the distortion. Ballantine[1] gives the following definition of distortion:

$$\text{Distortion} = \sqrt{\frac{(E)_2^2 + (E)_3^2 + (E)_4^2 + \cdots}{(E)_1^2}} \tag{793}$$

In the definition of Eq. (793), each subscript denotes the particular harmonic. Distortion is thus defined as the ratio of the total r.m.s. value of all harmonic components to the r.m.s. value

[1] See note on p. 566.

of the fundamental component. The curves of Fig. 305, from Ballantine, show the distortion as a function of carrier voltage (r.m.s.) and of degree of modulation m. These curves show, as do the rectification curves of Fig. 301, that distortion is large where the carrier voltage is low, and is also large when the degree of modulation is high for all values of carrier voltage. Corresponding to each combination of $\bar{E}_B$ and $\bar{E}_c$, there is a carrier voltage for which distortion is a minimum.

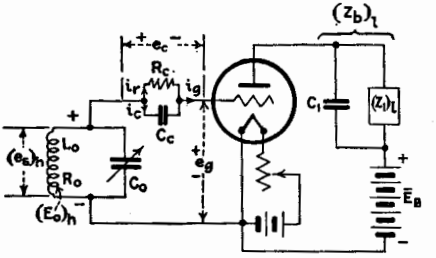


FIG. 306.—Circuits of the triode used for grid-circuit detection.

231. Grid-circuit Detection.—Under certain conditions detection may be caused to take place entirely in the grid circuit of a triode, in which case we have pure grid-circuit detection. Often, however, when the circuits are such as to permit grid-circuit detection to take place, plate-circuit detection is also present to some degree. The circuit connections for grid-circuit detection are given in Fig. 306. The connections shown in this figure are the same as those in the circuit diagram given in Fig. 280 of Chap. XX for small-signal detection, except that the polarizing battery in the grid circuit is omitted in Fig. 306.

The mechanism of detection in grid-circuit detection is quite different from that of plate-circuit detection. The mechanism can best be described by referring to Fig. 307. In this figure, the grid current i_g is shown plotted to e_g. In the lower part of the diagram, the instantaneous grid potential is shown plotted to a time axis extending downward. The instantaneous grid potential e_g is equal to the radio-frequency voltage $(e_s)_h$ across the oscillatory circuit less the instantaneous voltage e_c across the combination of R_c and C_c in parallel.

Assume that $(e_g)_h$ is sinusoidal and of constant amplitude $(\hat{E}_g)_h$. The grid condenser C_c is charged to some average potential $\bar{\bar{E}}_c$ which is maintained by the part of the grid current that flows into the condenser C_c during a small fraction of the

positive peak of $(e_g)_h$ when the grid potential swings slightly positive. The charge, which is communicated to C_c during the time when i_g flows, leaks off through the resistance R_c while there is no grid current. Figure 308 shows graphically some

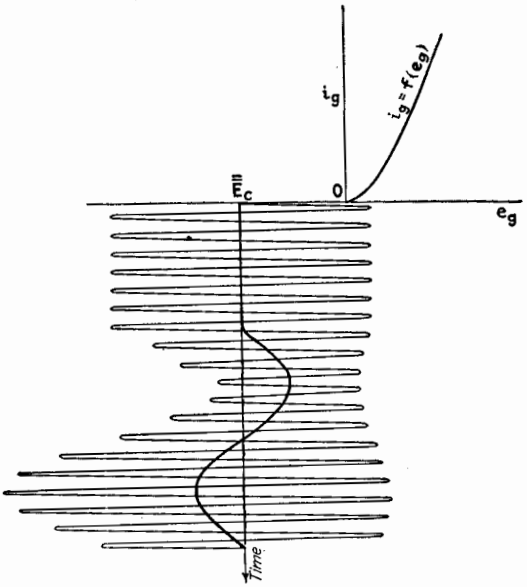


FIG. 307.—Grid-circuit detection with large-signal voltage.

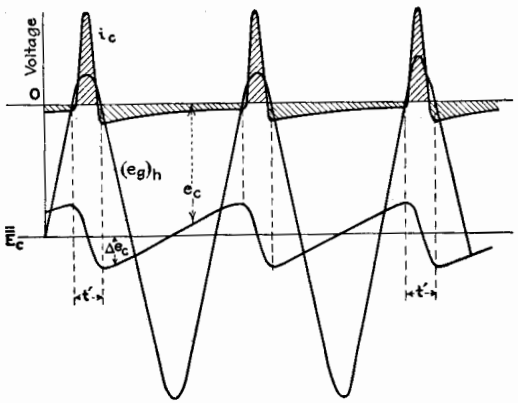


FIG. 308.—Voltages and grid current in grid-circuit detection.

of the potentials and currents in the grid circuit. The carrier voltage $(e_g)_h$ is shown drawn about the average voltage $\bar{\bar{E}}_c$ as axis. When voltage $(e_g)_h - \bar{\bar{E}}_c$ is positive, grid current flows.

Most of the grid current flows into condenser C_c, as shown by the positive loop of the i_c curve in Fig. 308. While $(e_g)_h - \bar{\bar{E}}_c$ is negative, the condenser C_c discharges through resistance R_c, as indicated by the rest of the i_c curve of the figure. If the average voltage $\bar{\bar{E}}_c$ of the condenser remains constant, as much charge flows into C_c during a cycle as flows out. Hence, the areas under the positive and the negative loops of i_c are equal. The instantaneous voltage of the condenser C_c varies approximately as shown by the curve e_c. Strictly speaking, the instantaneous grid voltage is $(e_s)_h - \bar{\bar{E}}_c - \Delta e_c$, where Δe_c is the fluctuation of condenser voltage due to the current i_c, this current being shown shaded in Fig. 308. If $(e_s)_h$ is assumed sinusoidal, the grid voltage $(e_g)_h$ is actually slightly distorted, as shown in somewhat exaggerated fashion at the right in Fig. 308. The voltage Δe_c will be shown to be very small under most conditions, so that the distortion of $(e_g)_h$ is neglected in this analysis.

The value of $\bar{\bar{E}}_c$ which any given $(\hat{E}_s)_h$ will maintain for a certain triode and for a certain value of R_c can be calculated approximately by the following analysis. Let t' be the time during which i_g flows. If the carrier grid voltage is expressed as

$$(e_g)_h = (\hat{E}_g)_h \cos \omega_h t \tag{794}$$

then

$$\bar{\bar{E}}_c = (\hat{E}_g)_h \cos \frac{\omega_h t'}{2} \tag{795}$$

During the time that i_g flows, the condenser C_c receives from i_g an increment of charge equal to

$$\Delta q' = \int_{-\frac{t'}{2}}^{\frac{t'}{2}} i_g dt \tag{796}$$

The condenser C_c discharges constantly through R_c and the decrease in charge during a cycle of duration T_h is

$$\Delta q'' = \frac{\bar{\bar{E}}_c}{R_c} \cdot T_h \tag{797}$$

The right-hand sides of Eqs. (796) and (797) may be equated. Before reducing the expression, some assumption must be made as to the shape of the curve of i_g.

We shall assume that the grid current for small values of grid voltage can be represented with sufficient accuracy by a parabola of the form

$$i_g = be_g + \frac{c}{2}e_g^2 \tag{798}$$

where b and c are the first and second derivatives of i_g with respect to e_g at $e_g = 0$.

From Eqs. (796), (797), and (798),

$$\int_{-\frac{t'}{2}}^{\frac{t'}{2}} \left(be_g + \frac{c}{2}e_g^2\right)dt = \frac{\bar{\bar{E}}_c}{R_c}T_h \tag{799}$$

where

$$e_g = (\hat{E}_g)_h \cos \omega t - \bar{\bar{E}}_c \tag{800}$$

Integrating Eq. (799) and eliminating t' by Eq. (795),

$$R_c = \frac{2\pi \dfrac{\bar{\bar{E}}_c}{(\hat{E}_g)_h}}{\left[\dfrac{c}{2}(\hat{E}_g)_h + \dfrac{\bar{\bar{E}}_c}{(\hat{E}_g)_h}(c\bar{\bar{E}}_c - 2b)\right]\cos^{-1}\dfrac{\bar{\bar{E}}_c}{(\hat{E}_g)_h} + \left(2b - \dfrac{3}{2}c\bar{\bar{E}}_c\right)\sqrt{1 - \dfrac{\bar{\bar{E}}_c^2}{(\hat{E}_g)_h^2}}} \tag{801}$$

If the k_g-curve passes through the origin, $b = 0$ and Eq. (801) reduces to

$$R_c(\hat{E}_g)_h = \frac{2\pi\beta/c}{(\tfrac{1}{2} + \beta^2)\cos^{-1}\beta - \tfrac{3}{2}\beta\sqrt{1 - \beta^2}} \tag{802}$$

where

$$\beta = \frac{\bar{\bar{E}}_c}{(\hat{E}_g)_h} \tag{803}$$

Equation (802) is given in Fig. 309 for several values of c expressed as micromhos per volt. The curves of Fig. 309 are useful in showing the variation of β as $(\hat{E}_g)_h$ varies, for any given constant values of c and R_c. For example, if R_c is 1 megohm and c is 10 micromhos per volt, the curve so marked gives directly the variation of β against $(\hat{E}_g)_h$, the numerical values of $(\hat{E}_g)_h$ being given directly by the scale of abscissas since R_c is unity. In this example, β would vary rapidly for values of $(\hat{E}_g)_h$ less than about 3 volts but would be fairly constant for larger values of $(\hat{E}_g)_h$.

The fluctuation of voltage across the condenser C_c can be found from

$$\Delta e_c = \frac{\overline{\overline{E}}_c}{R_c} \cdot \frac{(T_h - t')}{C_c} \tag{804}$$

provided Δe_c is small compared with $\overline{\overline{E}}_c$. Equation (804) reduces to

$$\Delta e_c = \frac{(\hat{E}_g)_h}{R_c C_c \omega_h} \cdot 2\beta(\pi - \sqrt{1 - \beta^2}) \tag{805}$$

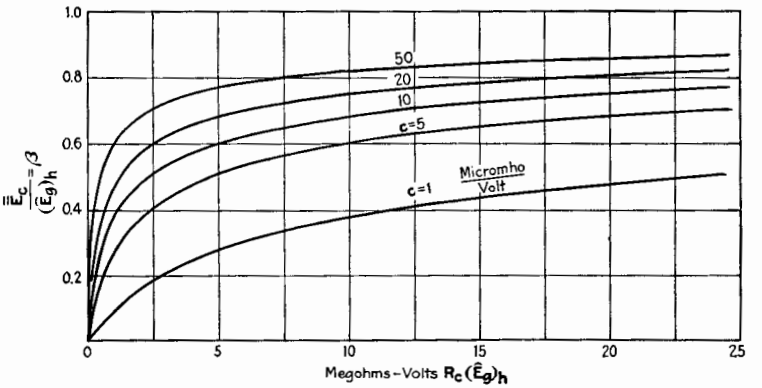


FIG. 309.—Values of β plotted against $R_c(\hat{E}_g)_h$.

The analysis just presented gives the value of β for any constant value of $(\hat{E}_g)_h$. When the radio-frequency carrier voltage is modulated, as shown in the lower part of Fig. 307, $(\hat{E}_g)_h$ varies according to the relation

$$\text{Amplitude} = (\hat{E}_g)_h(1 + m \sin \omega_l t) \tag{806}$$

As $(\hat{E}_g)_h$ varies sinusoidally, $\overline{\overline{E}}_c$ varies sinusoidally if β is constant but contains harmonics if β is not constant.

Another consideration of importance in determining the best values of the circuit constants for grid-circuit detection is whether or not the condenser C_c can lose its charge rapidly enough to permit $\overline{\overline{E}}_c$ to follow the decreasing amplitude of the carrier wave during modulation. If C_c and R_c are too large, the average grid voltage $\overline{\overline{E}}_c$ lags behind the envelope of the signal voltage, and the variation of $\overline{\overline{E}}_c$ is much reduced compared to the variation of the envelope. If C_c and R_c are smaller than necessary, the

sensitivity of the detector is reduced, and the equivalent input resistance of the grid is decreased to such a degree as to load excessively the oscillatory circuit L_0C_0 of Fig. 306. The average condenser voltage can follow the *ideal* curve of $\bar{\bar{E}}_c$ shown in Fig. 307, where the ideal curve of $\bar{\bar{E}}_c$ is β times the envelope, if for every instant the rate of decrease of condenser voltage due to leakage through R_c is equal to or greater than the rate of change of the ideal curve of $\bar{\bar{E}}_c$. Expressed mathematically, this is

$$\frac{\bar{\bar{E}}_c}{R_cC_c} \geqq \frac{d}{dt}(\hat{E}_g)_h(1 + m \sin \omega_l t) \tag{807}$$

or

$$\frac{1}{R_cC_c\omega_l} \geqq \frac{m \cos \omega_l t}{\beta(1 + m \sin \omega_l t)} \tag{808}$$

The maximum value of the right-hand side of Eq. (808), assuming β constant, occurs when $\sin \omega_l t = -m$, and the maximum value is $\dfrac{m}{\beta\sqrt{1 - m^2}}$.

Therefore

$$\frac{1}{R_cC_c\omega_l} \geqq \frac{m}{\beta\sqrt{1 - m^2}} \tag{809*}$$

Combining Eq. (809) with Eq. (805) gives the fluctuation of condenser voltage during a radio-frequency cycle, when R_cC_c is given the largest value desirable for small distortion of the modulation envelope. The result is

$$\frac{\Delta e_c}{(\hat{E}_g)_h} = \frac{2m}{\sqrt{1 - m^2}}(\pi - \sqrt{1 - \beta^2}) \cdot \frac{\omega_l}{\omega_h} \tag{810}$$

Assuming $\beta = 0.70$, $m = 0.50$, and $\omega_l/\omega_h = 0.01$, Eq. (810) gives $0.028\ (\hat{E}_g)_h$ as the order of magnitude of the fluctuation of Δe_c. The assumption made in the analysis just given, that Δe_c may be neglected, is therefore justified except for large values of m.

Consider some practical values of the various quantities in the analysis and the general method of their choice. If the input capacitance of the triode is of the order of $15\mu\mu$f, C_c should be several times this value in order that a fair fraction of the

* Expression (809) without the β was derived by Terman, *Proc. I.R.E.*, **18**, 2172 (1930).

radio-frequency voltage shall exist from grid to cathode. The capacitance C_c should be perhaps from five to ten times as great as the input capacitance. Let C_c be $100\mu\mu$f in this example. If we wish good fidelity of reproduction for frequencies up to 5,000 cycles per second and for degrees of modulation up to 0.6, Eq. (809) gives the maximum value of R_c as 1.07 megohms. In this calculation, β is taken as 0.7. Referring to Fig. 309, it is clear that if c is 10 or more, β is approximately constant for values of $(\bar{E}_g)_h$ above 3 volts. If the amplitude of the unmodulated carrier is 6 volts, fair fidelity is obtained for modulation from 0.50 to 0.60. The values of c for commercial triodes range from about 7 to 50 micromhos per volt. Frequently a value of R_c lower than that deduced in this example is used, but in such a case the other constants are adjusted accordingly.

The fictitious voltage of detection in the plate circuit can now be expressed. The r.m.s. value of the variation of $\bar{\bar{E}}_c$ is $\beta m((E_g)_h)_{m=0}$. The fictitious voltage in the plate circuit is

$$[(E_p)_l] = -u_p\beta m((E_g)_h)_{m=0} \tag{811}$$

The value of $\bar{\bar{r}}_p$ through which this fictitious voltage acts is not much different from r_p. More will be given on this point later.

A factor of considerable importance is the reduction of the voltage across the oscillatory circuit L_0C_0 of Fig. 306, due to the input conductance of the triode. Some idea of this reduction can be had by the use of Eq. (681) of Chap. XX, in which $\check{r}_g$ is replaced by $\bar{\bar{r}}_g$. This value of $\bar{\bar{r}}_g$, or its reciprocal $\bar{\bar{k}}_g$, can be calculated from Eq. (716) of Chap. XXI.

Grid-circuit detection can also be treated in a manner similar to the treatment of plate-circuit detection in Sec. 230 of this chapter. Having chosen a triode and the proper value of R_c, the choice being governed by the considerations just given, a rectification diagram can be obtained, using low frequencies for convenience. Capacitance C_c should be a large condenser during the test, since there is no modulation envelope for the condenser voltage to follow. Two such rectification diagrams are shown in Figs. 310 and 311. At low plate voltages, the rectification lines approach each other and in some cases cross, owing to simultaneous plate-circuit rectification which is opposite to grid-circuit rectification in its effect on the plate current.

A resistance line can be drawn, as indicated in Fig. 310, to intersect the rectification lines so that the points of intersection

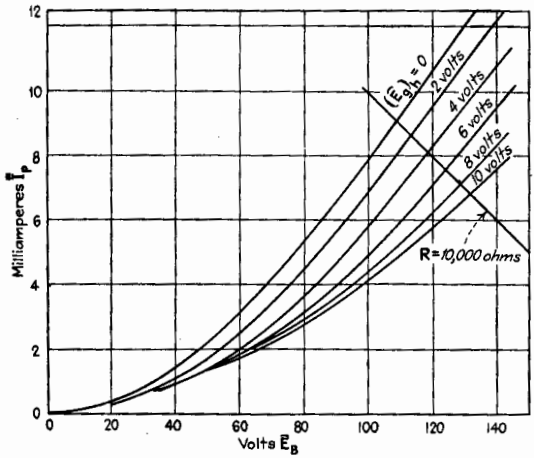


FIG. 310.—Grid-circuit detection characteristics or transrectification diagram. Triode ER201A.

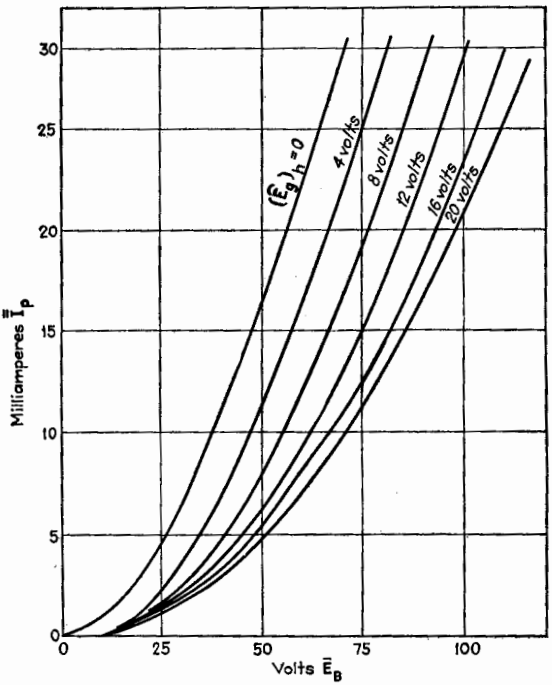


FIG. 311.—Grid-circuit detection characteristics or transrectification diagram. Triode ER171A.

are approximately equally spaced along the resistance line. In order to avoid plate-circuit detection, a high plate-battery voltage must be used.

Equations (786) through (792) hold in this case. The variational conductance for average values, $\bar{\bar{k}}_p$, given by Eq. (788), is the slope of the lines of the rectification diagram. The transconductance for rectification, $\bar{\bar{s}}_p$, given by Eq. (789), is a negative quantity in the case of grid-circuit rectification. The voltage ratio for rectification, defined in Eq. (790), is also a negative quantity and, as shown by Eq. (811), is related to β by

$$\bar{\bar{u}}_p = -u_p\beta \tag{812}$$

Referring to Fig. 310, $\bar{\bar{u}}_p$, as measured from the diagram at $(\hat{E}_g)_h = 4$ volts and $\bar{E}_B = 120$ volts, is -4.8. The value of u_p for this particular tube is about 8. Hence, from Eq. (812), β is 0.60. The rectification diagram of Fig. 310 was taken for $R_c = 1$ megohm. The approximate value of c for the type of tube used in taking the curves of Fig. 310 is about 15 micromhos per volt. The value of β, taken from Fig. 309 for this value of c, and for $R_c(\hat{E}_g)_h$ equal to 4, is 0.624, which checks the other value of β as well as can be expected.

232. Comparison of Plate-circuit and Grid-circuit Detection.—The values of $\bar{\bar{u}}_p$ for the two types of detection are not very different. For a 201A tube, $\bar{\bar{u}}_p$ for plate rectification is about 6 and for grid rectification about 5. The variational resistance for average values, $\bar{\bar{r}}_p$, differs considerably, being of the order of five times greater for plate-circuit rectification than for grid-circuit rectification. Therefore, for the same value of carrier grid voltage, grid-circuit detection gives greater sensitivity in the ratio of about four to one, but this apparent superiority is offset considerably by the decrease in carrier grid voltage in grid-circuit detection, owing to the input conductance of the grid circuit.

Grid-circuit detection has an advantage over plate-circuit detection in that the distortion in the former is less critical to carrier voltage. In plate-circuit detection, minimum distortion occurs at a particular value of $((\hat{E}_g)_h)_{m=0}$, and the distortion increases greatly as $((\hat{E}_g)_h)_{m=0}$ approaches zero. In grid-circuit detection, the rectification lines are more nearly equally spaced as the line for $(\hat{E}_g)_h = 0$ is approached, and hence distortion is not so great at low carrier voltages.

There are other practical considerations which have already been given and which may influence the choice of the type of detection used. For example, plate-circuit detection demands a much higher resistance or impedance in the plate circuit and, because of the higher resistance of the tube, draws much less current from the B-power source. The question of the selectivity of the tuned circuit in the input may be important and in this respect plate-circuit detection has the advantage.

233. Vacuum-tube Voltmeters.—The vacuum-tube voltmeter is of inestimable value for the measurement of alternating voltages of all frequencies. The ordinary electromagnetic type of voltmeter for low-frequency voltages has a much lower resistance than a d-c. voltmeter for the same voltage. For radio frequencies, the only voltmeters aside from the vacuum-tube voltmeter are either thermal instruments, consisting of a sensitive thermocouple and a d-c. meter, or electrostatic voltmeters. All of these voltmeters draw current, which often affects the voltage being measured. The principal advantage of the vacuum-tube type of voltmeter is the very small current taken by the instrument, this current being generally only the charging current into the input capacitance of a few micromicrofarads.

The vacuum-tube voltmeter is essentially a detector or rectifier. The alternating voltage being measured causes a change in the average plate current of the vacuum tube, and this change in plate current is indicated on a d-c. instrument. The principle of action is very similar to that which has been discussed in the foregoing sections for a detector upon which is impressed an unmodulated carrier voltage. Since the type of detector in which rectification takes place in the grid circuit draws more current from the voltage source being measured than a detector with rectification in the plate circuit, vacuum-tube voltmeters are generally of the plate-circuit type.

The simplest form of vacuum-tube voltmeter is shown in Fig. 312*a*. The plate circuit contains only a d-c. instrument and a plate-circuit battery, both shunted with a by-pass condenser. The rectification diagram of Fig. 304 is applicable to this case, the path of operation being along a resistance line corresponding to the resistance of the instrument. The resistance line for the instrument is usually almost vertical.

If there is a conduction path through the potential difference being measured, the grid circuit may be as shown in Fig. 312*a*. If,

however, the circuit to which the voltmeter is connected is open for steady currents, the grid circuit of Fig. 312*b* is used. The input circuit in Fig. 312*a* is preferable to that of Fig. 312*b*, because in the former the input impedance is only the capacitive reactance of the triode, and this is often reduced to a minimum by debasing the tube. The input impedance of the second form of grid circuit, Fig. 312*b*, can be reduced by making R_c very large, as, for example, 10 megohms or more.

The polarizing voltage $\bar{E}_c$ is adjusted to give some predetermined plate current when no alternating voltage acts on the grid.

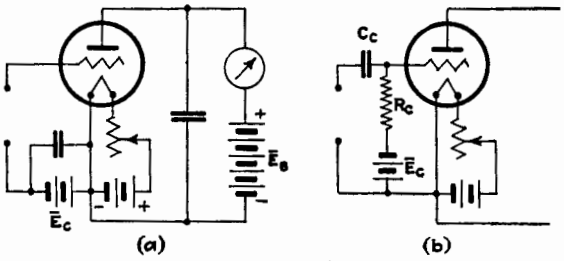


FIG. 312.—Simple vacuum-tube voltmeter.

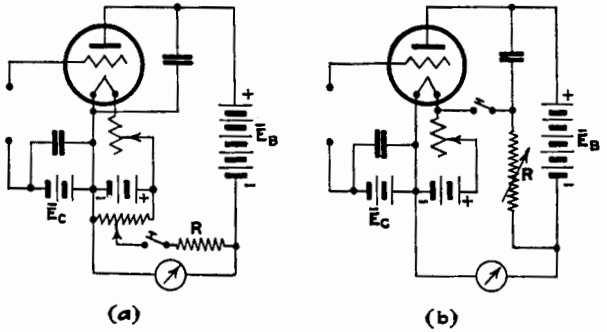


FIG. 313.—Vacuum-tube voltmeter with zero adjustment.

When the alternating voltage is impressed, the *change in plate current* generally increases as the square of the a-c. grid voltage for low voltages and then approaches a linear relation for high grid voltage, as shown in Fig. 303. The calibration curve is always obtained experimentally, but a study of the rectification diagram aids in choosing the proper voltages.

The form of voltmeter shown in Fig. 312 is limited in its sensitivity because the quiescent plate current, when no a-c. voltage is impressed, deflects the d-c. instrument. Smaller

changes of plate current can be measured if this quiescent current is balanced out of the instrument thus permitting a more sensitive d-c. microammeter to be used. Two ways of balancing out the quiescent current are indicated in Fig. 313*a* and *b*. In both cases R is made as large as possible, in order to take only a small fraction of the current increments from the instrument. In the use of a voltmeter with a balancing circuit, the plate current is adjusted to some standard value by varying $\bar{E}_c$ when the switch in series with R is open. The switch is then closed and the balancing circuit is adjusted to give zero current through the instrument.

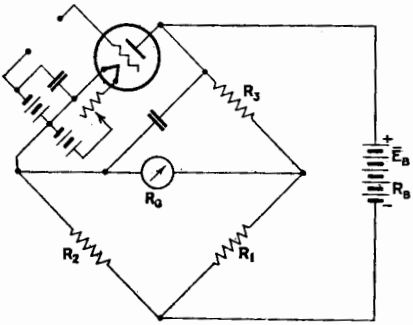


FIG. 314.—Bridge arrangement of vacuum-tube voltmeter.

Another method of balancing out the quiescent plate current is by the use of the bridge principle, as shown in Fig. 314. The condition of balance is

$$\frac{\bar{r}_p}{R_3} = \frac{R_2}{R_1} \tag{813}$$

and the deflections of the instrument can be considered to be due to the change of $\bar{r}_p$. The resistance $\bar{r}_p$ for the quiescent point becomes $\bar{\bar{r}}_p$ when an a-c. voltage is impressed on the grid. The treatment of such bridge systems may follow the usual method, the plate-load resistance being

$$R_b = \frac{R_2R_B(R_3+R_G)+R_1(R_3+R_G)(R_2+R_B)+R_3R_G(R_2+R_B)}{R_1(R_2+R_3+R_B+R_G)+(R_2+R_G)(R_3+R_B)} \tag{814}$$

where R_B and R_G are the resistances of the battery and galvanometer.

The fraction of the change in plate current, due to an impressed a-c. voltage on the grid, which passes through the galvanometer or d-c. instrument is

$$\Delta\bar{\bar{I}}_G = \Delta\bar{\bar{I}}_p \frac{R_1(R_2 + R_B) + R_2(R_3 + R_B)}{R_1(R_2 + R_3 + R_B + R_G) + (R_2 + R_G)(R_3 + R_B)} \tag{815}$$

Generally R_2 is made about equal to the resistance $\bar{r}_p$ of the triode, and R_1 and R_3 are small and about equal to each other. The sensitivity of the arrangement depends somewhat upon the choice of resistances but is in general about the same as the sensitivity of the arrangements of vacuum-tube voltmeters previously given. The bridge arrangement has the advantage that variations in the plate-battery voltage do not upset the zero-voltage balance.

Often a second triode is substituted for R_3, with a slight rearrangement of the bridge, as in Fig. 315. The sensitivity is reduced to about one-half that of the previous arrangements of vacuum-tube voltmeters, but the balance of the steady com-

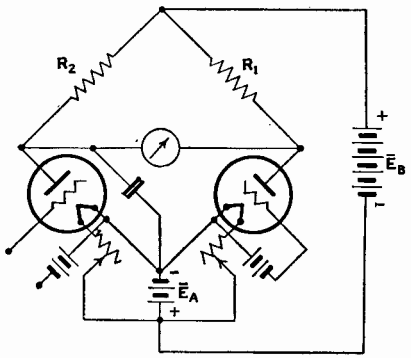


FIG. 315.—Bridge-type vacuum-tube voltmeter with balancing triode.

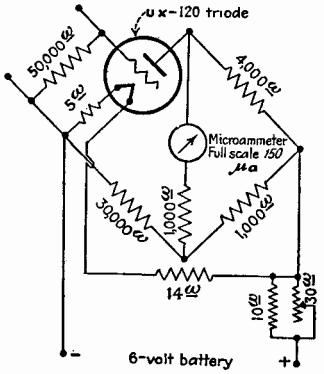


FIG. 316.—Portable bridge-type thermionic voltmeter. (*General Electric Company.*)

ponent of the current is now practically independent of variations in filament voltage as well as variations in plate-battery voltage.

Another bridge arrangement in which the plate voltage is supplied by the filament battery is shown in Fig. 316. This arrangement, although less sensitive than the other arrangements, is convenient as applied to a portable instrument. Full-scale reading is generally of the order of 5 volts.

The range of the vacuum-tube voltmeter shown in Figs. 312 and 313 can be greatly extended by a connection known as *reflexing*. A resistance, one end of which is connected to the filament, is made common to both plate and grid circuits. The voltage drop through this resistance, due to the plate current,

increases the negative grid-polarizing voltage as the a-c. grid voltage increases. The resulting calibration of the voltmeter is more nearly linear the larger the reflexing resistance. The full-scale reading of the instrument may be made to have almost any value, such as 50 or 100 volts, or more.

The connections for reflexing added to the simple connections of Fig. 312 are shown in Fig. 317. A convenient combination of the circuits for reflexing and balancing out $\bar{I}_p$, used by the author, is shown in Fig. 318. Switch S_2, when closed, cuts out the reflexing resistance. The filament voltage is first adjusted

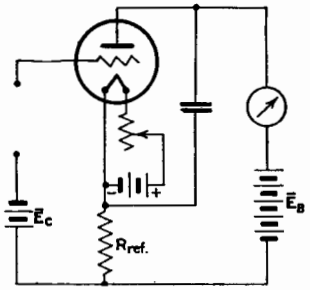


FIG. 317.—Reflexed vacuum-tube voltmeter.

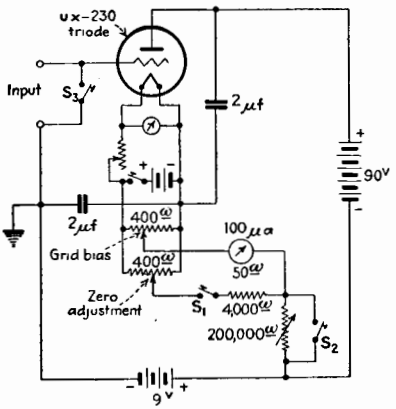


FIG. 318.—Vacuum-tube voltmeter $u_p = 9.3$; $C_{gf} = 3.7\mu\mu f$; $C_{pg} = 6.4\mu\mu f$.

to 2 volts. With switch S_1 open, the grid-bias potential divider is adjusted to give 300 microamp. with the microammeter reduced in sensitivity. Switch S_1 is then closed and the zero-adjustment potential divider is adjusted to give zero deflection. When the microammeter is used to give full deflection for 100 microamp., full-scale reading corresponds to an r.m.s. value of 1.22 volts impressed on the grid. With the microammeter shunted to give full-scale reading for 1 milliamp., full-scale reading corresponds to an r.m.s. value of 4.6 volts. With the reflexing resistance in circuit, the grid bias is readjusted to give no deflection. Then full-scale reading of the microammeter can be made to correspond to any r.m.s. value of voltage from 1.22 to 150 volts, according to the magnitude of the reflexing resistance.

234. Errors in Indication of Vacuum-tube Voltmeter Due to Variation of Wave Form.—Most vacuum-tube voltmeters

give correct indications of the r.m.s. value of a voltage *only* when the voltage is sinusoidal. The errors due to presence of harmonics may be very great and may vary with the phase of the harmonics with respect to the fundamental voltage. For example, the reflexed type of voltmeter gives deflections which are dependent only upon the *positive* loop of potential and are often dependent only upon the *peak* of this loop of potential. Evidently such a voltmeter is useless unless the voltage is sinusoidal.

The only vacuum-tube voltmeter which gives correct indications independent of wave form is one whose operating characteristic curve can be expressed by a power series containing no terms of degree higher than the third. The proof of this statement follows.

Let

$$\bar{\imath}_p = b\bar{e}_g + \frac{c}{2}\bar{e}_g^2 + \frac{d}{6}\bar{e}_g^3 + \frac{f}{24}\bar{e}_g^4 + \cdots \tag{816}$$

be the operating characteristic curve, where $\bar{\imath}_p$ and $\bar{e}_g$ are measured from the quiescent values. If $\bar{e}_g = \sqrt{2}E_{g_1} \sin \omega t + \sqrt{2}E_{g_2} \sin (2\omega t + \phi) + \ldots$, the steady increment obtained when this value of $\bar{e}_g$ is substituted in Eq. (816) is

$$\Delta(\bar{\bar{I}}_p) = \frac{c}{2}(E_{g_1}^2 + E_{g_2}^2 + \cdots) + \frac{f}{24}(E_{g_1}^4 + E_{g_2}^4 + 4E_{g_1}^2 E_{g_2}^2 + \cdots)$$

$$= \frac{c}{2}(E_g)_{\text{r.m.s.}}^2 + \frac{f}{24}(E_{g_1}^4 + E_{g_2}^4 + 4E_{g_1}^2 E_{g_2}^2 + \cdots) \tag{817}$$

The third-degree term in Eq. (816) does not contribute to the value of $\Delta(\bar{\bar{I}}_p)$, but the fourth-degree term causes an error, given by Eq. (817). If f is zero, the change in average plate current is proportional to the square of the r.m.s. value of the a-c. grid voltage.

General References

BALLANTINE: Detection of High Signal Voltages, *Proc. I.R.E.*, **17**, 1153 (1929).

MEDLAM and OSCHWALD: The Thermionic Voltmeter, *Exp. Wireless*, **3**, 589 (1926); (conclusion) **3**, 664 (1926); Further Notes on the Reflex Voltmeter, *Exp. Wireless*, **5**, 56 (1928).

PANFILOV: Errors of Measurement by Means of Electron Tube Voltmeters, *Westnik. Elektrotechniki*, Nr.1, Sekt. I, p. 38 (1931).

TERMAN and MORGAN: Some Properties of Grid-leak Power Detection, *Proc. I.R.E.*, **18**, 2172 (1930).

CHAPTER XXIII

TETRODES AND PENTODES

The terms *tetrode* and *pentode* are generally applied to vacuum tubes having two and three grids, respectively. It is conceivable that the term tetrode might be applied to a full-wave rectifier tube which comprises two filaments, each having its own plate, but such use of the term is not common.

Certain advantages result from the use of more than one grid, depending upon the connections used external to the tube and on the voltages impressed upon the several electrodes. Before describing these advantages, a general theory will be given which is applicable to all multielectrode tubes.

I. GENERAL THEORY OF MULTIELECTRODE TUBES

In the general theory, the voltages of the electrodes are always the voltages between the electrodes and filament or cathode. The voltage to any electrode has a positive sign when the electrode is positive with respect to the cathode. When there are several grids, the first grid denotes the one nearest the cathode. The conventional direction of voltages and currents is shown for a pentode in Fig. 319.

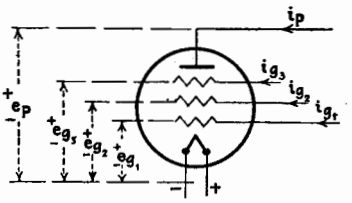


FIG. 319.—Conventions for multielectrode tube.

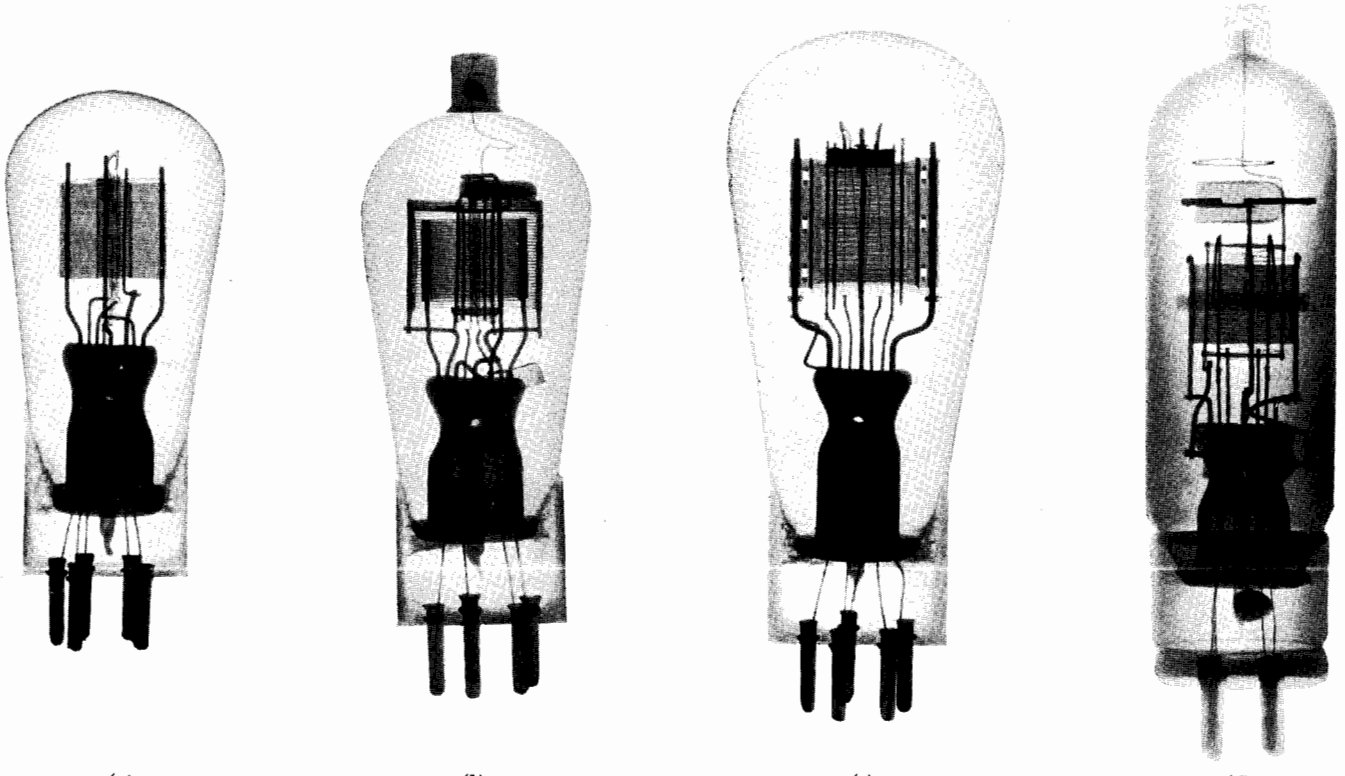
235. General Equations for Multielectrode Tubes.—The plate current is a function of all electrode voltages or

$$i_p = F(e_p e_{g_1} e_{g_2} e_{g_3} + \cdots) \tag{818}$$

If only first-order terms are considered,

$$di_p = \frac{\partial i_p}{\partial e_p} de_p + \frac{\partial i_p}{\partial e_{g_1}} de_{g_1} + \frac{\partial i_p}{\partial e_{g_2}} de_{g_2} + \cdots \tag{819}$$

$$= k_{pp} de_p + s_{pg_1} de_{g_1} + s_{pg_2} de_{g_2} + \cdots \tag{820}$$



(*a*) (*b*) (*c*) (*d*)

PLATE V.—X-ray photographs of vacuum tubes. (*a*) Triode, type 227; (*b*) screen-grid tetrode, type 224; (*c*) tetrode, type 46; (*d*) tetrode, type FP54.

(*Facing page* 588)

Equations (819) and (820), taken together, give the definitions of the variational plate conductance k_{pp}, and the several variational plate transconductances s_{pg_1}, s_{pg_2}, etc.

If all voltages except e_p and e_{g_1} are constant, and if e_p and e_{g_1} vary so as to maintain i_p constant,

$$-\left(\frac{\partial e_p}{\partial e_{g_1}}\right)_{i_p=\text{const.}} = \frac{s_{pg_1}}{k_{pp}} \tag{821}$$

The left-hand member of Eq. (821) is one of the voltage ratios (sometimes called *amplification constants*) and is denoted by u_{pg_1}. The other voltage ratios, u_{pg_2}, u_{pg_3}, etc., may be similarly defined.

The definitions of the tube coefficients which directly concern the plate current are given below.

Plate Coefficients:

$$k_{pp} = \frac{\partial i_p}{\partial e_p} = \frac{1}{r_{pp}} \tag{822}$$

$$s_{pg_1} = \frac{\partial i_p}{\partial e_{g_1}} \tag{823}$$

$$s_{pg_2} = \frac{\partial i_p}{\partial e_{g_2}} \qquad \text{etc.} \tag{824}$$

$$u_{pg_1} = -\left(\frac{\partial e_p}{\partial e_{g_1}}\right)_{i_p=\text{const}} \tag{825}$$

$$u_{pg_2} = -\left(\frac{\partial e_p}{\partial e_{g_2}}\right)_{i_p=\text{const.}} \qquad \text{etc.} \tag{826}$$

$$s_{pg_1} = u_{pg_1} k_{pp} \tag{827}$$

$$s_{pg_2} = u_{pg_2} k_{pp} \qquad \text{etc.} \tag{828}$$

The definitions of tube coefficients concerned with the current to the first grid are defined in a similar manner. The first-grid current is

$$i_{g_1} = f_1(e_p, e_{g_1}, e_{g_2}, e_{g_3} \cdots) \tag{829}$$

and

$$di_{g_1} = \frac{\partial i_{g_1}}{\partial e_p} de_p + \frac{\partial i_{g_1}}{\partial e_{g_1}} de_{g_1} + \frac{\partial i_{g_1}}{\partial e_{g_2}} de_{g_2} + \cdots \tag{830}$$

$$= s_{g_1 p} de_p + k_{g_1 g_1} de_{g_1} + s_{g_1 g_2} de_{g_2} + \cdots \tag{831}$$

If only e_p and e_{g_1} vary, and vary so as to maintain i_{g_1} constant,

$$-\left(\frac{\partial e_{g_1}}{\partial e_p}\right)_{i_{g_1}=\text{const.}} = \frac{s_{g_1p}}{k_{g_1g_1}} \tag{832}$$

The left-hand member of Eq. (832) is defined as the *first-grid reflex factor* and is denoted by u_{g_1p}. The definitions of the coefficients concerned with the current to the first grid are given below.

First-grid Coefficients:

$$k_{g_1g_1} = \frac{\partial i_{g_1}}{\partial e_{g_1}} = \frac{1}{r_{g_1g_1}} \tag{833}$$

$$s_{g_1p} = \frac{\partial i_{g_1}}{\partial e_p} \tag{834}$$

$$s_{g_1g_2} = \frac{\partial i_{g_1}}{\partial e_{g_2}} \qquad \text{etc.} \tag{835}$$

$$u_{g_1p} = -\left(\frac{\partial e_{g_1}}{\partial e_p}\right)_{i_{g_1}=\text{const.}} \tag{836}$$

$$u_{g_1g_2} = -\left(\frac{\partial e_{g_1}}{\partial e_{g_2}}\right)_{i_{g_1}=\text{const.}} \qquad \text{etc.} \tag{837}$$

$$\left.\begin{aligned} s_{g_1p} &= u_{g_1p}k_{g_1g_1} \\ s_{g_1g_2} &= u_{g_1g_2}k_{g_1g_1} \end{aligned}\right\} \tag{838}$$

Similar definitions go with each grid of the tube. Further definitions are not given because the scheme of subscripts is sufficiently well established by the equations already given.

When a multielectrode tube is connected in circuit, the plate is connected to the output circuit. One of the grids is connected to the input circuit and is called the *control* grid. Generally, each of the other grids is connected, perhaps through a battery, to one of the three main electrodes, *i.e.*, the plate, the control grid, and the cathode. When this is the case, the tube is virtually a triode as far as the external circuit is concerned. The characteristics of the equivalent triode depend greatly upon how the auxiliary grids are connected.

236. The Tetrode.—Consider, for illustration, the various ways in which a tetrode may be connected. In Fig. 320*a*, the first grid is used as control grid. The second grid is connected

through a battery to the cathode. If the impedance of the battery is negligible, de_{g_2} in Eq. (820) is zero, and

$$di_p = k_{pp}de_p + s_{pg_1}de_{g_1} \tag{839}$$

If we denote by k_p, s_p, and u_p the plate-current coefficients of the triode which is equivalent to the tetrode, we have from Eq. (839)

$$\left.\begin{aligned} k_p &= k_{pp} \\ s_p &= s_{pg_1} = u_{pg_1}k_{pp} \\ u_p &= \frac{s_p}{k_p} = u_{pg_1} \end{aligned}\right\} \text{Equivalent triode for Fig. 320a} \tag{840}$$

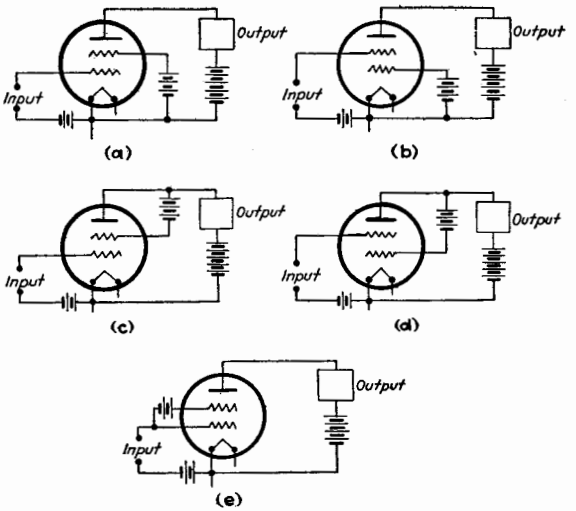


FIG. 320.—Connections of tetrodes.

If the connections are as shown in Fig. 320*b*, de_{g_1} in Eq. (820) is zero. Then

$$di_p = k_{pp}de_p + s_{pg_2}de_{g_2} \tag{841}$$

The equivalent triode coefficients for this method of connection of the tetrode are, therefore,

$$\left.\begin{aligned} k_p &= k_{pp} \\ s_p &= s_{pg_2} = u_{pg_2}k_{pp} \\ u_p &= \frac{s_{pg_2}}{k_{pp}} = u_{pg_2} \end{aligned}\right\} \text{Equivalent triode for Fig. 320b} \tag{842}$$

A third method of connecting a tetrode is shown in Fig. 320*c*, in which de_{g_2} is the same as de_p. The total plate current of the equivalent triode is, in this case,

$$di_p + di_{g_2} = (k_{pp} + s_{pg_2} + s_{g_2p} + k_{g_2g_1})de_p + (s_{pg_1} + s_{g_2g_1})de_{g_1} \quad (843)$$

The equivalent triode coefficients are

$$\left.\begin{aligned} k_p &= k_{pp}(1 + u_{pg_2}) + k_{g_2g_2}(1 + u_{g_2p}) \\ s_p &= s_{pg_1} + s_{g_2g_1} = u_{pg_1}k_{pp} + u_{g_2g_1}k_{g_2g_2} \\ u_p &= \frac{s_p}{k_p} \end{aligned}\right\} \text{Equivalent triode for Fig. 320}c \quad (844)$$

For the connections shown in Fig. 320*d*, $de_{g_1} = de_p$ and the total plate current of the equivalent triode is

$$di_p + di_{g_1} = (k_{pp} + s_{pg_1} + s_{g_1p} + k_{g_1g_1})de_p + (s_{pg_2} + s_{g_1g_2})de_{g_2} \quad (845)$$

The equivalent triode coefficients are

$$\left.\begin{aligned} k_p &= k_{pp}(1 + u_{pg_1}) + k_{g_1g_1}(1 + u_{g_1p}) \\ s_p &= s_{pg_2} + s_{g_1g_2} = u_{pg_2}k_{pp} + u_{g_1g_2}k_{g_1g_1} \\ u_p &= \frac{s_p}{k_p} \end{aligned}\right\} \text{Equivalent triode for Fig. 320}d \quad (846)$$

Finally, a fifth way of connecting a tetrode is shown in Fig. 320*e*. In this case, $de_{g_1} = de_{g_2}$ and Eq. (820) becomes

$$di_p = k_{pp}de_p + (s_{pg_1} + s_{pg_2})de_g \quad (847)$$

The equivalent triode coefficients are

$$\left.\begin{aligned} k_p &= k_{pp} \\ s_p &= s_{pg_1} + s_{pg_2} = k_{pp}(u_{pg_1} + u_{pg_2}) \\ u_p &= \frac{s_{pg_1} + s_{pg_2}}{k_{pp}} = u_{pg_1} + u_{pg_2} \end{aligned}\right\} \text{Equivalent triode for Fig. 320}e \quad (848)$$

The discussion of these five methods of connection of a tetrode shows that, except for the connections shown in Fig. 320*c* and *d*, where a grid current is added to the plate current, the coefficients for the plate circuit of the equivalent triode can be specified in terms of *three* coefficients, *i.e.*, k_{pp}, u_{pg_1}, and u_{pg_2}. Equation (820), the fundamental equation of the tetrode, can be written

$$di_p = k_{pp}(de_p + u_{pg_1}de_{g_1} + u_{pg_2}de_{g_2}) \quad (849)$$

in which only these three coefficients appear.

If Eq. (849) be written in the form

$$di_p = k_{pp}u_{pg_1}\left(\frac{1}{u_{pg_1}}de_p + \frac{u_{pg_2}}{u_{pg_1}}de_{g_2} + de_{g_1}\right) \tag{850}$$

the magnitude of the coefficients can more easily be visualized in terms of shielding factors or Durchgriffs, explained for the triode in Chap. VII. The shielding factor occasioned by the presence of the first grid is u_{pg_2}/u_{pg_1} and may be denoted by D_1. Then, D_1de, impressed on the second grid, causes the same intensity of field at the cathode as de impressed on the first grid. Similarly, $1/u_{pg_1} = D_2$ is the shielding factor for the two grids and D_2de, impressed on the plate, produces the same field at the cathode as de on the first grid. Obviously, D_2 is much smaller than D_1, or u_{pg_1} is much greater than u_{pg_2}.

Again, Eq. (849) may be written

$$di_p = k_{pp}\left[de_p + u_{pg_2}\left(de_{g_2} + \frac{u_{pg_1}}{u_{pg_2}}de_{g_1}\right)\right] \tag{851}$$

In this equation, u_{pg_1}/u_{pg_2} is the voltage ratio for the combination of the cathode, the first grid, and the second grid, considered as a triode. If de_p is zero, the triode, in which the second grid is considered to be the plate, has a plate conductance of $k_{pp}u_{pg_2}$. If de_{g_1} is zero, u_{pg_2} is the voltage ratio of the equivalent triode comprising the plate, the second grid as control grid, and the cathode. Under these conditions, the first grid may be considered part of the cathode structure.

The plate-circuit characteristics of the equivalent triode can be obtained from a knowledge of the three coefficients k_{pp}, u_{pg_1}, and u_{pg_2}. Figures 321, 322, and 323 show graphs of these coefficients for a constant plate voltage of 100 volts. These curves are for a certain French two-grid tube, the grid mesh being approximately the same for both grids. It is seen that u_{pg_1} is large compared with u_{pg_2}, the reason for this having just been explained.

If a tetrode is used in accordance with any of the connections shown in Fig. 320, the tetrode is equivalent to a triode, and all of the theory applies which has been developed for a triode. The characteristics of the equivalent triode are, however, generally different from the characteristics of a simple triode.

If impedances are included in the circuits of *all* electrodes, the tetrode is *not* equivalent to a triode. For this case, a special e-p-c. theorem can be developed. The method of analysis is

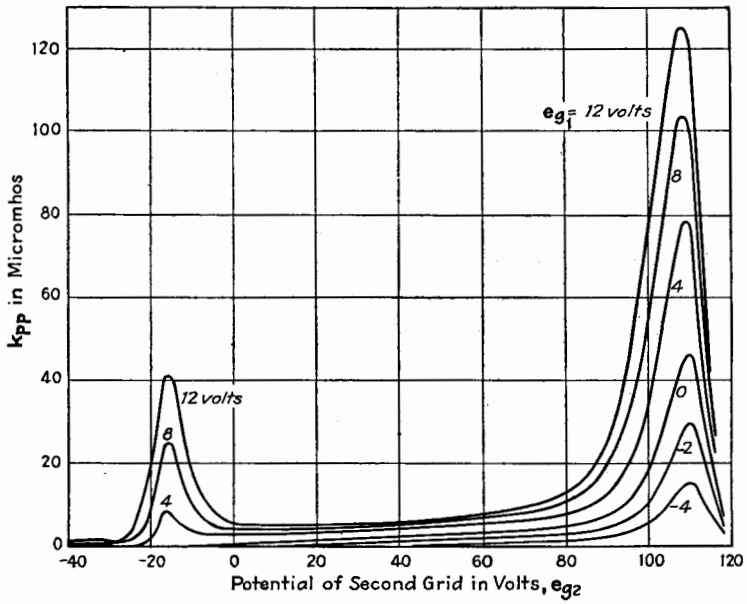


FIG. 321.—Curves of k_{pp} *vs.* e_{g_2} for a tetrode. (French Bigril tube.)

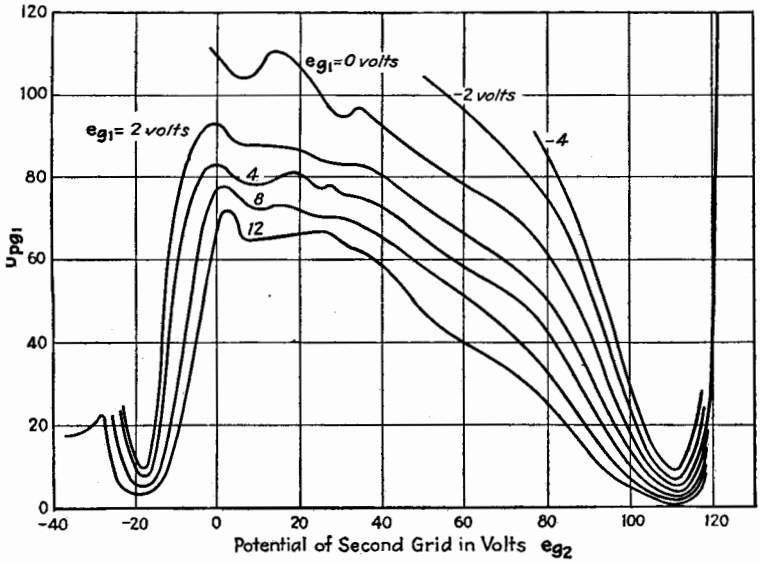


FIG. 322.—Curves of u_{pg_1} *vs.* e_{g_2} for a tetrode. (French Bigril tube.)

similar to that used in developing the e-p-c. and e-g-c. theorems for a triode and will not be given in detail for a tetrode. The final expressions for the equivalent-circuit theorems for a tetrode, applying to alternating currents, are

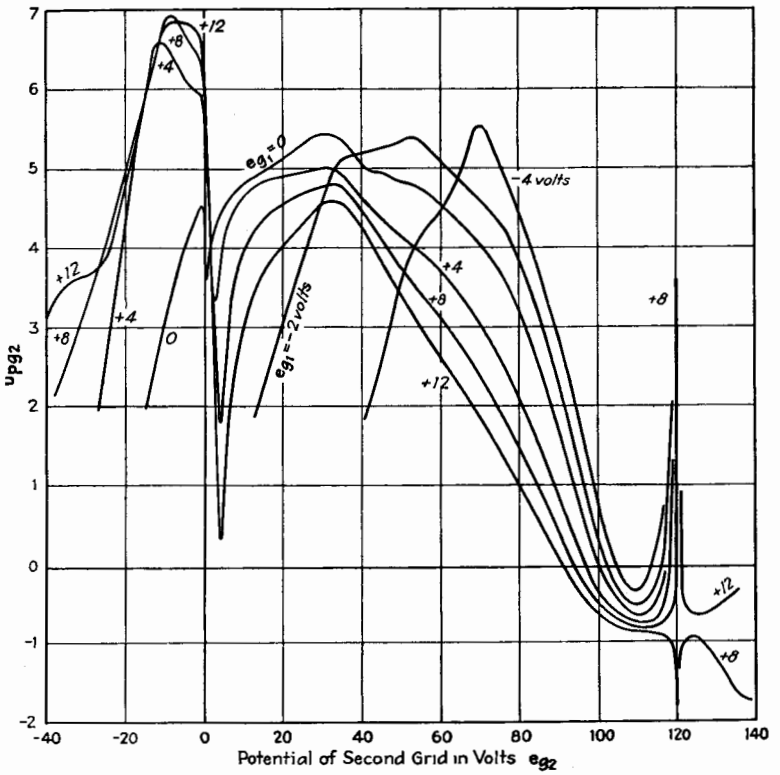


FIG. 323.—Curves of u_{pg_2} *vs.* e_{g_2} for a tetrode. (French Bigrīl tube.)

$$\Delta \mathbf{I}_p = \frac{u_{pg_1}\Delta \mathbf{E}_{g_1} + u_{pg_2}\Delta \mathbf{E}_{g_2}}{r_{pp} + \mathbf{Z}_b} \tag{852}$$

$$\Delta \mathbf{I}_{g_1} = \frac{u_{g_1p}\Delta \mathbf{E}_p + u_{g_1g_2}\Delta \mathbf{E}_{g_2}}{r_{g_1g_1} + \mathbf{Z}_c} \tag{853}$$

$$\Delta \mathbf{I}_{g_2} = \frac{u_{g_2p}\Delta \mathbf{E}_p + u_{g_2g_1}\Delta \mathbf{E}_{g_1}}{r_{g_2g_2} + \mathbf{Z}_{c_2}} \tag{854}$$

where $\mathbf{Z}_b$, $\mathbf{Z}_{c_1}$, and $\mathbf{Z}_{c_2}$ are the impedances in the plate, first-grid, and second-grid circuits. These three equations are solved simultaneously according to the conditions of the particu-

lar problem. For example, if the first grid is the control grid, and $\Delta \mathbf{E}_{g_2}$ is due to the voltage drop through $\mathbf{Z}_{c_2}$,

$$\Delta \mathbf{E}_p = -\mathbf{Z}_b \Delta \mathbf{I}_p$$
$$\Delta \mathbf{E}_{g_2} = -\mathbf{Z}_{c_2} \Delta \mathbf{I}_{g_2}$$

Using these two equations with Eqs. (852) and (854) gives

$$\Delta \mathbf{I}_p = \frac{u_{pg_1}(r_{g_2g_2} + \mathbf{Z}_{c_2}) - u_{pg_2}u_{g_2g_1}\mathbf{Z}_{c_2}}{(r_{pp} + \mathbf{Z}_b)(r_{g_2g_2} + \mathbf{Z}_{c_2}) - u_{pg_2}u_{g_2p}\mathbf{Z}_b\mathbf{Z}_{c_2}} \Delta \mathbf{E}_{g_1} \qquad (855)$$

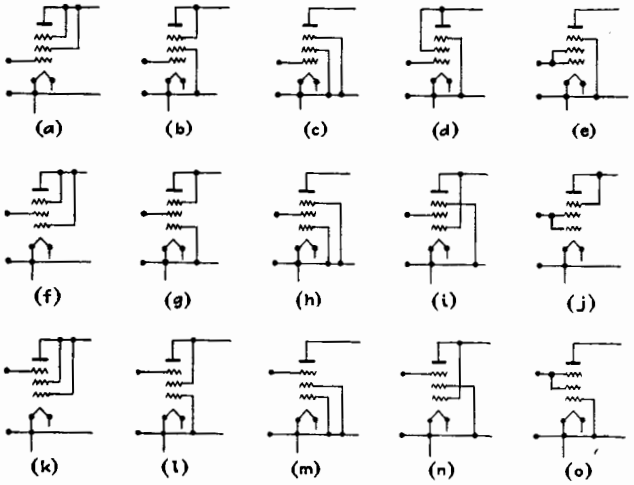


FIG. 324.—Connections of pentodes.

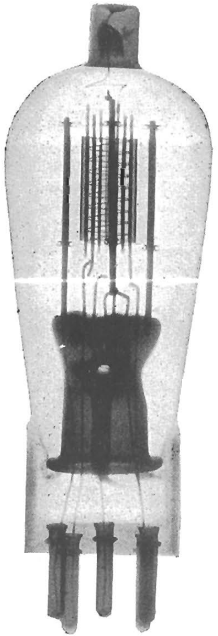
237. The Pentode.—The general theory applied to the pentode is so similar to that just given for the tetrode as to require little additional explanation.

The plate-current equation for a pentode, corresponding to Eq. (849) for the tetrode, is

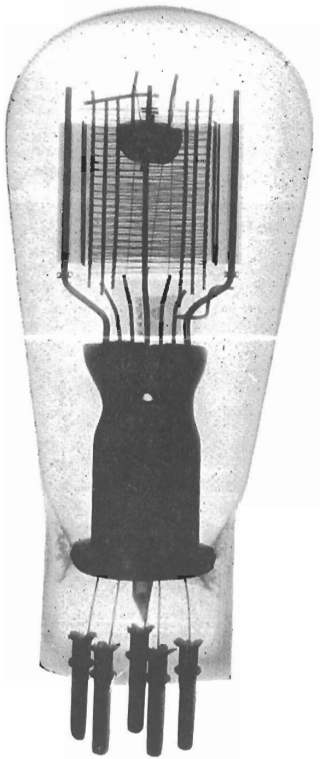
$$di_p = k_{pp}(de_p + u_{pg_1}de_{g_1} + u_{pg_2}de_{g_2} + u_{pg_3}de_{g_3}) \qquad (856)$$

Corresponding equations give the several grid currents.

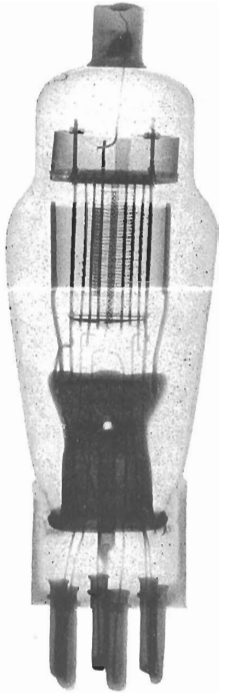
The large number of electrodes in a pentode permits of many connections, some of which are shown in Fig. 324. Batteries are omitted in the drawing but it is to be understood that batteries may be included in circuit with any or all of the electrodes, so that the steady potentials of the electrodes may have any values desired. Many of the connections of Fig. 324 are



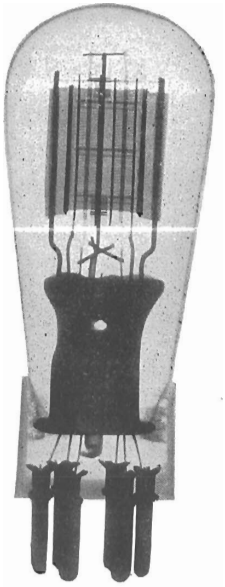
(*a*)



(*b*)



(*c*)



(*d*)

PLATE VI.—X-ray photographs of vacuum tubes. (*a*) Pentode, type 38; (*b*) pentode, type 47; (*c*) pentode, type 57; (*d*) double-cathode tube, type 29.

(*Facing page* 596)

of academic interest only, but all are such that the pentode is reduced to an equivalent triode. The coefficients for the plate circuit of the equivalent triodes can be calculated from the *four* coefficients given in Eq. (856) for the connections shown in Fig. 324*c*, *e*, *h*, *m*, and *o*. The equivalent triode coefficients for the other connections involve the coefficients associated with the grids of the pentode. Consider the connection shown in Fig. 324*c*. For this connection,

$$di_p = k_{pp}(de_p + u_{pg_1}de_{g_1}) \tag{857}$$

and

$$\left.\begin{aligned} k_p &= k_{pp} \\ s_p &= k_{pp}u_{pg_1} \\ u_p &= u_{pg_1} \end{aligned}\right\} \text{Equivalent triode for Fig. 324}c \tag{858}$$

Equivalent-circuit theorems, similar to those for the tetrode, can be developed for the pentode when the pentode is not used as an equivalent triode. Since the pentode, like the tetrode, is generally used as an equivalent triode, these extended circuit theorems are of little practical use but can readily be developed when desired.

238. Internal Capacitances of Multielectrode Tubes.—Each electrode in a multielectrode tube has, theoretically, capacitance with every other electrode in the tube. Following the same notation as used with triodes, the capacitance between a certain pair of electrodes is indicated by the two subscripts corresponding to the two electrodes. For example, the capacitance between the plate and the second grid of a tetrode is C_{pg_2}. The order of the subscripts is immaterial.

If the tube contains n electrodes, the number of internal capacitances is given by the series

$$\text{Number of capacitances} = 1 + 2 + 3 + \cdots (n - 1)$$

Therefore, in a tetrode there are six, and in a pentode ten internal capacitances.

II. TETRODES

The general theory of tetrodes has been given. The characteristics of the equivalent triode corresponding to certain of the connections of Fig. 320 will now be considered in detail and some special applications of the tetrode will be described.

239. Tetrode with Space-charge Grid.—Consider the connections of Fig. 320*b*. The first grid is connected to the cathode through a battery which is directed so as to make the first grid positive with respect to the cathode. This grid attracts electrons from the cathode. Some of the electrons are projected through the grid and are aided thereby in their passage to the plate. The opposing action of the space charge is overcome to some extent by this first grid. The first grid, thus connected, is called a *space-charge grid* because of its effect as just described.

The advantage of the space-charge grid is that the equivalent triode has a higher transconductance at a given plate voltage than a triode without the space-charge grid. The space-charge grid also permits a tetrode to operate satisfactorily on a lower plate voltage than is possible with a triode. This advantage is important when the plate supply is a dry battery, and tetrodes thus used were common in Europe some years ago.

240. Tetrode with Screen Grid.—The complications, in both theory and operation of triodes, due to the capacitance C_{pg} between the input and output circuits, have been described. Schottky[1] in 1919 first suggested the use of an electrostatic screen, interposed between the control grid and the plate, to reduce the capacitance between these electrodes. He suggested the term *screen grid* for this added electrode, and tetrodes possessing a grid of this kind are known as *screen-grid tubes*. Further development of the screen-grid tetrode was made in 1926 by Hull and Williams.[2] A. W. Hull[3] also described several applications of the new tetrode.

The connections of a screen-grid tetrode are shown in Fig. 320*a*. The second or screen grid is constructed so as to prevent, as far as possible, lines of electrostatic force from passing directly between plate and control grid. If this screening is perfectly accomplished, the capacitance C_{pg_1} is zero, and the plate circuit and the grid circuit have no coupling due to capacitance within the tube. Some of the other causes, listed in Chap. XVIII, for energy interchange between the input and output circuit may still exist, however.

In order to obtain a large screening effect, the screen grid is generally made of fine mesh and is formed so as to surround

[1] SCHOTTKY, *Arch. Elektrotech.*, **8**, 299 (1919).
[2] HULL and WILLIAMS, *Phys. Rev.*, **27**, 432 (1926).
[3] HULL, *Phys. Rev.*, **27**, 439 (1926).

the plate. The structure of the screen is shown in the X-ray picture of tube *b*, Plate V, facing page 588 and in the photograph of parts of the screen-grid tube in Plate I, page 14. Not only is the plate itself screened, but the lead wire to the control grid is brought out at the top of the tube in such a way that it is screened from the plate. The capacitance C_{pg_1} is much less than the usual value of from 5 to 10$\mu\mu$f for an ordinary triode. Generally C_{pg_1} for the screen-grid tube is less than 0.01$\mu\mu$f. The capacitance between the control grid and the cathode of the equivalent triode is equal to the sum of C_{g_1k} and C_{g_2k} and is of the order of 5$\mu\mu$f; the plate to cathode capacitance of the equivalent triode is $C_{pk} + C_{pg_2}$ and is of the order of 10$\mu\mu$f. Since the screen-grid tube is usually constructed with a separately heated cathode, the subscript k, denoting cathode, is used instead of f in the symbols for capacitances.

The electrons are aided in their passage through the screen grid by making the potential of the screen grid positive with respect to the cathode but negative with respect to the plate. Generally the screen-grid potential is of the order of one-half of the plate potential.

241. Characteristic Curves of Screen-grid Tetrode.—Typical static characteristic curves of the screen-grid tetrode are shown in Figs. 325, 326, and 327.[4] Figure 327 is the most interesting and instructive of the three figures. Consider the portion to the left of the line AA. As the plate voltage increases, the plate current first increases, then decreases and becomes negative, and then increases again. The screen-grid current passes through changes opposite in direction to those of the plate current. The negative slope of the plate-current graph is due to secondary emission from the plate, the secondary electrons leaving the plate and passing to the more positive screen grid. When the plate current is negative, each primary or bombarding electron liberates more than one secondary electron, so that more electrons leave the plate than arrive at the plate. The tetrode, when adjusted to operate in the region of the characteristic curves just described, has useful applications which will be considered later.

[4] Figures 325, 326, and 327, and several that follow are taken from a "Handbook" of tube characteristics published by the Commercial Engineering Department of the R. C. A. Radiotron Company, Inc., Harrison, N. J.

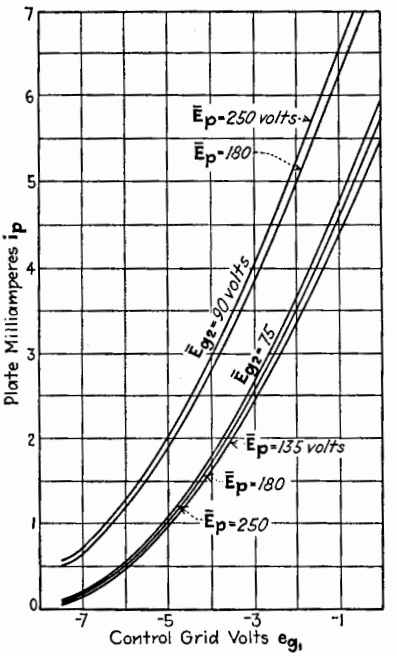


FIG. 325.—Curves of i_p *vs.* e_{g_1} for tetrode UY224. (*R.C.A. Radiotron Handbook*).

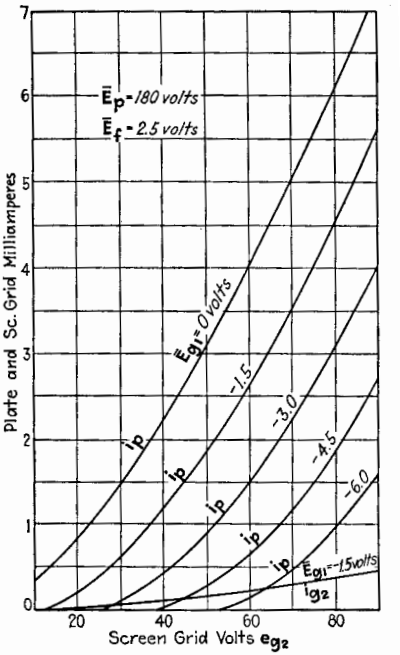


FIG. 326.—Curves of i_p *vs.* e_{g_2} for tetrode UY224. (*R.C.A. Radiotron Handbook.*)

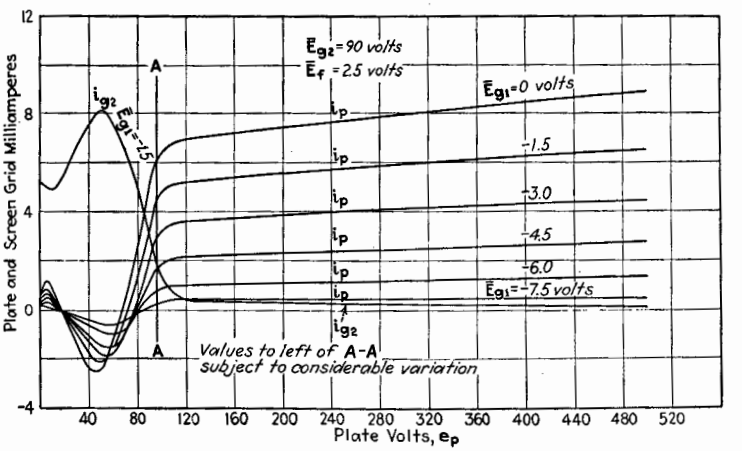


FIG. 327.—Curves of i_p *vs.* e_p and i_{g_2} *vs.* e_p for tetrode UY224. (*R.C.A. Radiotron Handbook.*)

The portion of the curves which is useful for most purposes lies to the right of the vertical line AA, Fig. 327. In this region, the plate-current lines are practically straight and the plate current is nearly independent of plate voltage. Evidently r_p, the reciprocal of the slope of these lines, is very large. For the particular tetrode, if $\bar{E}_p = 250$ volts, $\bar{E}_{g_2} = 90$ volts, and $\bar{E}_{g_1} = -3$ volts, r_p is 600,000 ohms.

Inspection of the plate-current graphs at the right of line AA in Fig. 327 shows that u_{pg_1}, or u_p of the equivalent triode, is

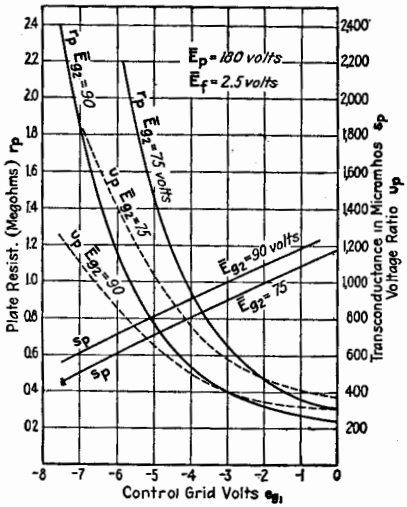


FIG. 328.—Equivalent triode coefficients of tetrode UY224 plotted against e_{g_1}. (*R.C.A. Radiotron Handbook.*)

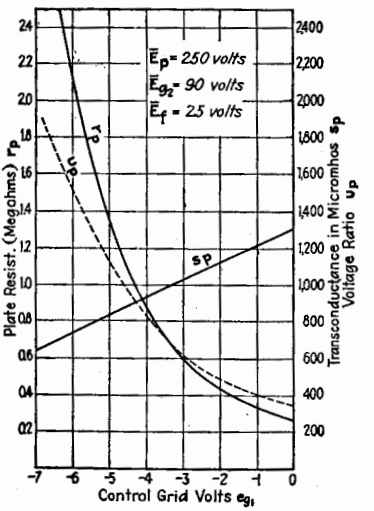


FIG. 329.—Equivalent triode coefficients of tetrode UY224 plotted against e_{g_1}. (*R.C.A. Radiotron Handbook.*)

very large. Under the particular conditions given in the last paragraph, u_p is 615. Although both r_p and u_p are very large compared to the values for simple triodes, s_p is not much different from that for a triode of the same size. The transconductance for the values of r_p and u_p just given is 1,025 micromhos.

The variational coefficients, for the equivalent triode corresponding to the particular screen-grid tetrode for which the curves of Figs. 325, 326, and 327 were taken, are shown in Figs. 328, 329, 330, and 331. It is noteworthy that u_p is not so independent of electrode voltages as in a simple triode. Furthermore, u_p

decreases as r_p *decreases*, as the electrode voltages vary, whereas in a simple triode, u_p always increases as r_p decreases. As was pointed out in Chap. VII, if u_p and r_p vary in the same direction when the electrode voltages vary, the changes in plate current are more nearly linearly related to the changes in control-grid voltage than when u_p and r_p vary as in a simple triode.

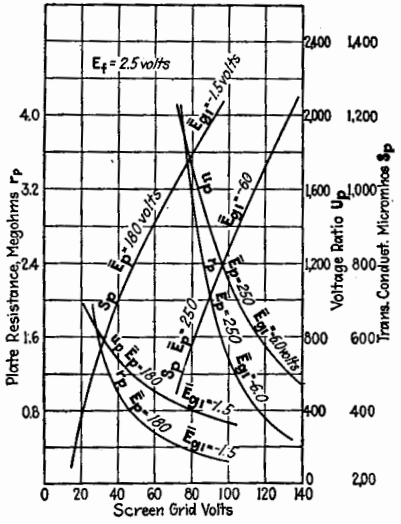


FIG. 330.—Equivalent triode coefficients of tetrode UY224, plotted against e_{g_2}. (*R.C.A. Radiotron Handbook.*)

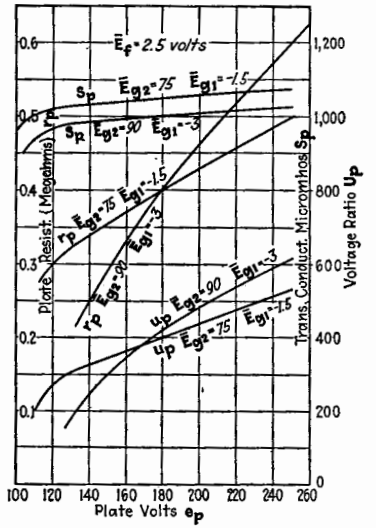


FIG. 331.—Equivalent triode coefficients of tetrode UY224 plotted against e_p. (*R.C.A. Radiotron Handbook.*)

242. Screen-grid Tetrode as a Negative Resistance.—When a plate-current graph has a negative slope, as in Fig. 327 to the right of line AA, k_p is negative. The tetrode is often used under this condition as a negative-resistance device. It is then called a *dynatron*. This negative conductance between plate and cathode may be connected across an oscillatory circuit consisting of an inductance L and a capacitance C. If $k_p + \frac{CR}{L}$ is zero, the circuit oscillates. In the expression just given, R is the a-c. resistance of L and C in series. Unfortunately, the negativity of k_p depends upon the amount of secondary emission, which in turn depends greatly upon the variable condition of the surface of the plate. A very small film of the "getter" material or of

barium and strontium from the cathode, deposited on the plate, changes greatly the amount of secondary emission. These thin films of foreign materials on the surface of the plate vary with time. Consequently, the dynatron characteristics of the tetrode are not constant.

243. Screen-grid Tetrode as a Class A Amplifier.—It was pointed out in Chap. XII that the merit of a triode, when used as an amplifier, is measured by $\sqrt{s_p u_p}$. This figure of merit is eight to ten times greater for a screen-grid tetrode than for a simple triode of the same size.

The high amplification factor u_p of the screen-grid tetrode would lead one to expect a correspondingly high amplification

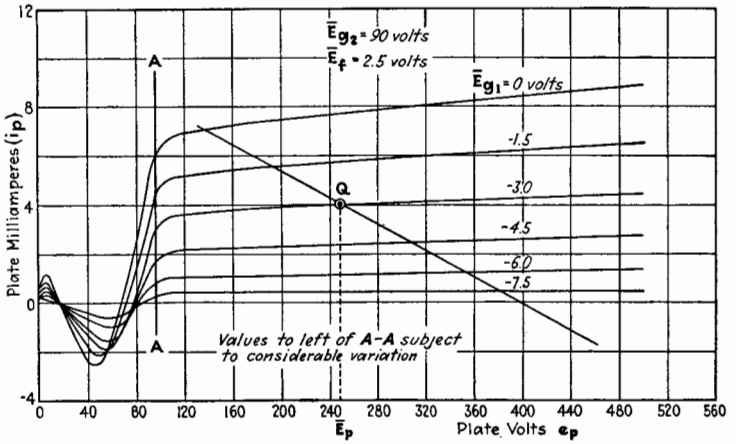


FIG. 332.—Curves of i_p *vs.* e_p for tetrode UY224, showing load-resistance line.

per stage of a Class A amplifier using this tube. The expected high amplification, corresponding to u_p for the tetrode, is not realized in practice, because of the impossibility of obtaining and using a plate load having an impedance much greater than r_p of the tetrode. Because the tetrode must be used with a plate load of high impedance, it is most useful as a voltage amplifier. For example, the tetrode may be successfully used in resistance-coupled amplifiers and in tuned radio-frequency amplifiers. If the screen-grid tetrode is used in a transformer-coupled amplifier at audio frequencies, frequency distortion results because of the relatively low impedance of the transformers.

The plate-current plate-voltage diagram, with a plate-load resistance line drawn in, is shown in Fig. 332. The resistance

line corresponds to a resistance of 37,500 ohms. The voltage amplification can be calculated by the use of Eq. (463), Chap. XV, page 401. In this equation, A_n^2 and $g_{g_{(n+1)}}$ can be neglected but, because of the small value of k_{pn}, the first fraction in the second bracket of the denominator cannot be neglected as was done in obtaining Eq. (464) for the simple triode. If the quiescent point is at Q of Fig. 332, the values for the constants of Eq. (463) are as follows:

$k_{pn} = 1.67 \cdot 10^{-6}$ mho; $u_{pn} = 615$; $K_{bn} = 26.7 \cdot 10^{-6}$ mho; $K_{c_{(n+1)}} = 0.2 \cdot 10^{-6}$ mho; $C_{pn} + C_{g_{(n+1)}} = 20\mu\mu$f; and $C = 0.05\mu$f.

With these values, the voltage amplification at several frequencies is given in the following table.

TABLE XVIII

Frequency, cycles per second	$(V. A.)_n$
1	30.4
10	35.9
100	36.0
1,000	36.0
10,000	35.9
100,000	33.0
1,000,000	8.0

This table shows that the performance of the amplifier just described is unusually satisfactory as compared to a similar amplifier using simple triodes. A plate-load resistance having a value greater than 37,500 ohms, which was assumed in the example, may be used with the result that a higher $(V.A.)_n$ is obtained.

On account of the high plate voltage demanded by the screen-grid tetrode, the plate-battery voltage must be high. For the amplifier just described, the plate-battery voltage is 540 volts. A higher plate-load resistance would demand a still higher battery voltage.

The screen-grid tetrode is especially well adapted to voltage amplifiers at radio frequencies. It uses a tuned plate load, either in the form of a simple tuned circuit, as described in Chap. XVI, or in the form of a tuned transformer, as described in Chap. XVII. The theory presented in these chapters applies

also when the screen-grid tetrode is used. Generally, the formulas can be simplified by neglecting A_n and that part of the input admittance of the next stage which is due to C_{pg} of this next stage. In other words, $g_{g_{(n+1)}}$ is practically zero and $C_{g_{(n+1)}}$ is equal to $C_{g_1g_2} + C_{g_1k}$. Because of the high amplification per stage, complete shielding of the separate stages is very important.

244. Screen-grid Tetrode as Detector.—The screen-grid tube can be used either as a small-signal detector or as a detector for large signal voltages. The theory of Chap. XX applies where small signal voltages are used; the theory of Chap. XXII when large signal voltages are used.

The screen-grid tube is often used for detection with large signal voltages. In this case better results are generally obtained

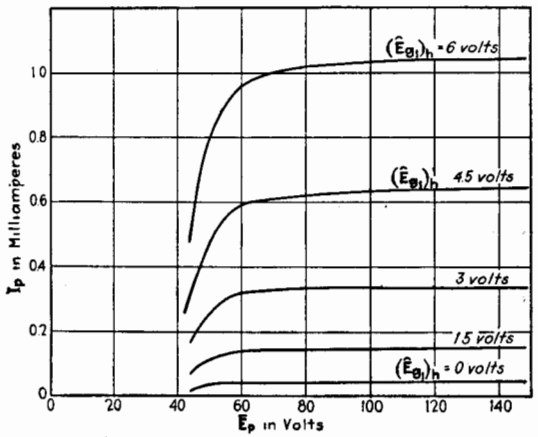


FIG. 333.—Plate-circuit detection characteristics of tetrode ER224. $\bar{E}_A = 2.5$ volts; $\bar{E}_{g_2} = 45$ volts; $\bar{E}_{g_1} = -4.5$ volts.

if the voltage of the screen grid is made less than the usual value for the tube used as an amplifier. The screen-grid tube used as a detector for large signal voltages is of the order of from five to ten times more sensitive than a simple triode.

Either plate- or grid-circuit rectification may be used. The rectification diagram for plate detection, taken for the same type tube for which the characteristic curves just described were taken, is shown in Fig. 333. The corresponding rectification diagram for grid-circuit detection is shown in Fig. 334. The theory of Chap. XXII can be applied for further analysis of detection.

245. Tetrode with Second Grid Connected to Plate.—There is in use to some extent at the present time a tetrode having two

ordinary grids (not screen grid) as shown in tube *c* of Plate V, facing page 588. The tube is designed so that the second grid may be connected as in either Fig. 320*c* or Fig. 320*e*.

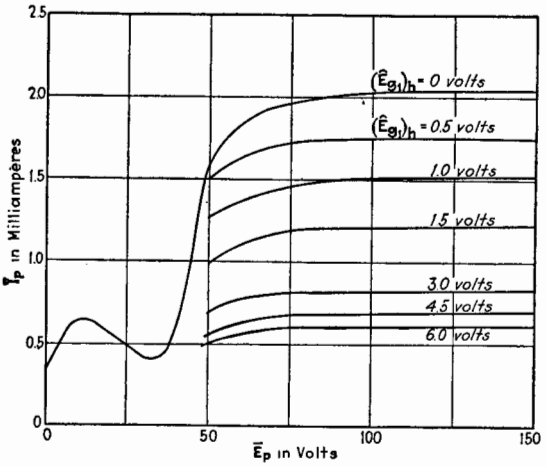


FIG. 334.—Grid-circuit detection characteristics of tetrode ER224. $\bar{E}_A = 2.5$ volts; $\bar{E}_{v_2} = 45$ volts.

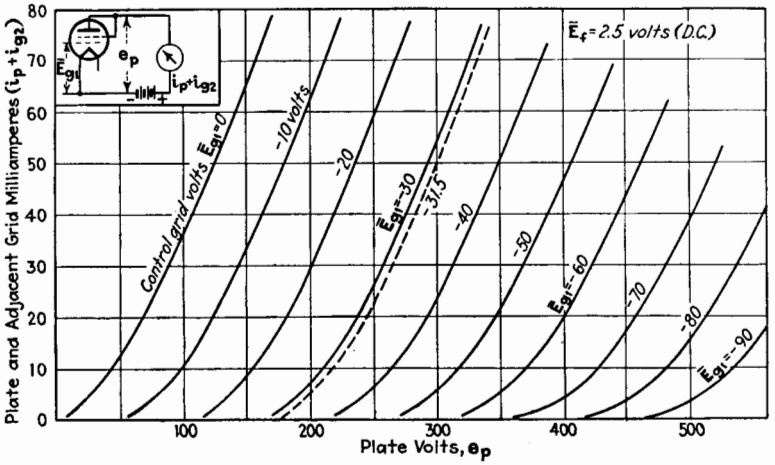


FIG. 335.—Curves of $i_p + i_{v_2}$ *vs.* e_p for tetrode type 46. (*R.C.A. Radiotron Handbook.*)

When the tetrode is used as in Fig. 320*c*, the potential of the second grid being the same as that of the plate, the characteristics of the tube are much the same as for a simple triode having

a plate situated at some position between the second grid and the plate of the tetrode.

The plate-current plate-voltage characteristic curves of the tube *c* of Plate V when connected as just described are shown in Fig. 335. The voltage ratio u_p is about 5.6, r_p for a plate current of 30 milliamp. is about 2,000 ohms, and s_p for the same current is about 2,700 micromhos.

246. Tetrode with Two Grids Connected.—When the two grids of a tetrode are connected and used as the control grid while polarized at a negative potential, the voltage ratio u_p

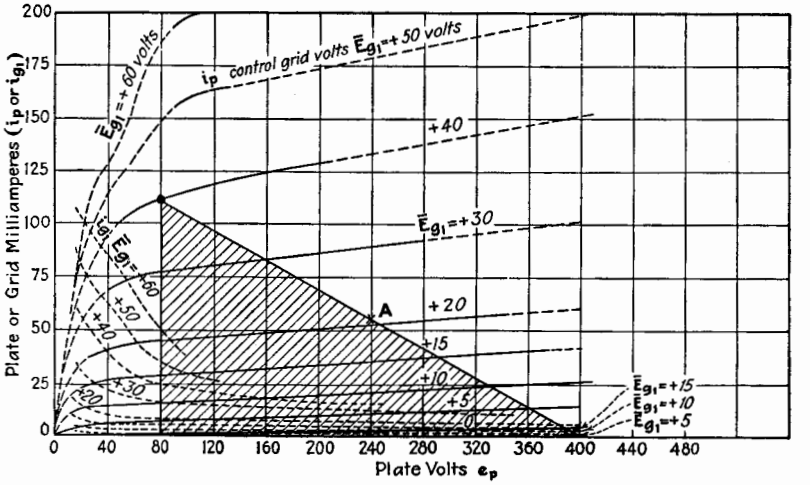


FIG. 336.—Curves of i_p *vs.* e_p for tetrode type 46 used as Class B amplifier. $\bar{E}_{g_1} = \bar{E}_{g_2}$. (*R.C.A. Radiotron Handbook.*)

of the equivalent triode is high, as shown by Eq. (848). Secondary emission, so pronounced in the screen-grid tetrode, is practically absent. The plate-current plate-voltage characteristic curves of the tube just described are shown in Fig. 336. This figure should be compared with Fig. 335.

The voltage ratio for this tube is about 65 and r_p is of the order of 24,000 ohms. The value of s_p is about the same as when the second grid is connected to the plate.

Practically no plate current flows unless the control grid is positive, as is shown by Fig. 336. The control grid, when positive, draws a current shown by the dotted lines of the figure. Because of these characteristics, this tube is not used as a simple

Class A amplifier tube, but as a power-output tube connected in a special way, shown in Fig. 337. As shown by the characteristic curves and the diagram of connections, two tubes are used with their grids connected to the opposite terminals of the secondary winding of the input transformer. In this arrangement, one tube passes current during the positive loop of input

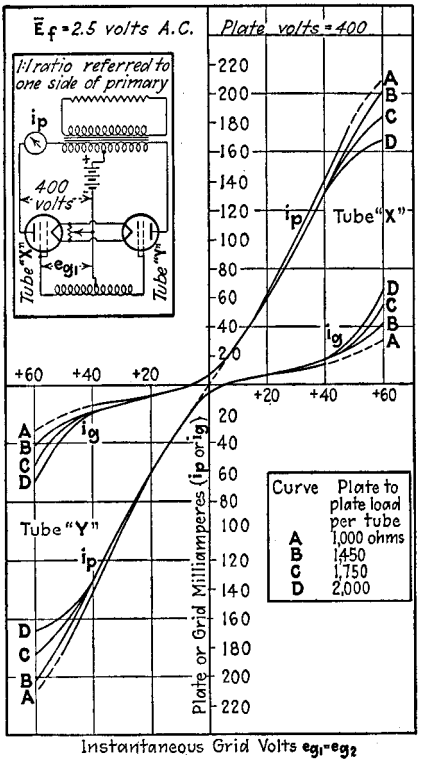


FIG. 337.—Characteristics of two tetrodes type 46 used as Class B amplifiers.

voltage and the other tube during the negative loop of input voltage. The amplifier tube in the previous stage must be able to supply the grid current without undue distortion.

This arrangement is known in the art as a *push-push amplifier*. Although each tube operates as a Class C amplifier, the two tubes taken together are known as a Class B amplifier. It has the advantage of very large power output at high efficiency. If the grid-voltage amplitude is 40 volts, the a-c. power output of both

tubes at fundamental frequency is given by the shaded area below the assumed resistance line, shown in Fig. 336. The quiescent point is almost on the horizontal axis. The average point for the current taken by both tubes is approximately at the mid-point of the resistance line. Under these conditions, nearly the theoretical maximum efficiency of 50 per cent for sinusoidal operation is attained.

One disadvantage of this arrangement is the large harmonic content of the output compared to the best performance of a simple triode power amplifier. High efficiency is obtained with a sacrifice in quality of reproduction.

III. PENTODES

The pentode was used in Europe before it came into general use in this country. In Europe, the lighting service in different places is not standardized as regards voltage and frequency so that, to a considerable extent, set designers were forced to use dry batteries for plate supply. The use of low plate voltages was therefore of great importance. The advantage of the space-charge grid as used in the tetrode and also as used in the pentode permits the tubes to operate satisfactorily on low plate voltages. One of the chief advantages of the pentode is large power output obtainable with small tubes operating at low voltages.

Although many different connections are possible with a pentode, as shown in Fig. 324, only two are used extensively in practice at the present time. These two connections are shown in Fig. 324*c* and *h*.

247. Pentode with Suppressor Grid.—When a pentode is connected as in Fig. 324*c*, the first or inner grid is the control grid and is polarized negatively. The second grid, frequently called the screen grid, is connected to the cathode through a battery which gives the grid a positive potential of 200 volts or more. The high potential attracts the electrons through the first grid and projects some electrons through the next or third grid to the plate. The third grid is generally connected directly to the cathode and is often called the *suppressor* grid. This name comes from the action of this grid in suppressing the secondary emission from the plate which gave the undesirable portion of the characteristic curves to the left of line AA in Fig. 327 for the tetrode.

Static characteristic curves of the pentode, tube *b* of Plate VI, are shown in Figs. 338 and 339. The action of the suppressor

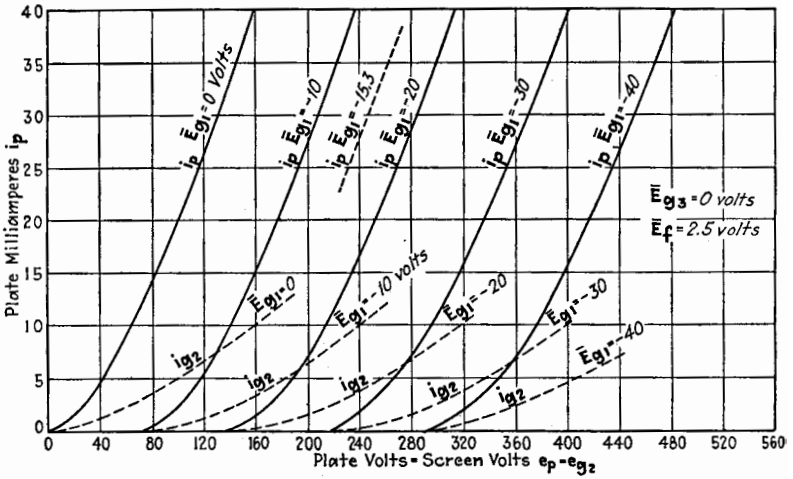


FIG. 338.—Curves of i_p *vs.* e_p of pentode type 247. $e_{g_2} = e_p$. (*R.C.A. Radiotron Handbook.*)

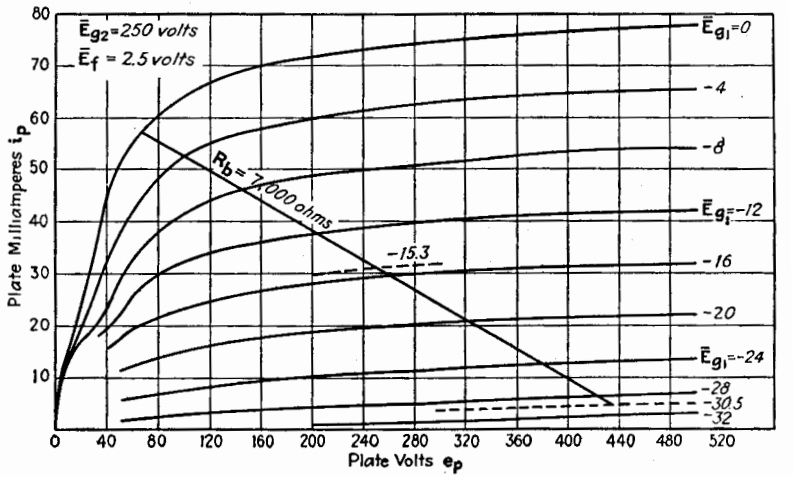


FIG. 339.—Curves of i_p *vs.* e_p for pentode type 247 with suppressor grid. $\bar{E}_{g_2}$ = 250 volts; $\bar{E}_{g_3} = 0$. (*R.C.A. Radiotron Handbook.*)

grid is made evident by the absence of the negative slopes of the plate-current curves in Fig. 339. The absence of this region of negative slope permits the load resistance line to be moved to

the left, as in Fig. 339, and therefore permits lower plate voltages to be used.

The voltage ratio u_p of the pentode shown is about 150, and the variational plate resistance r_p is about 60,000 ohms.

The second grid of the pentode just described is similar to the other grids of the tube and has a relatively coarse mesh. This grid reduces the capacitance C_{pg_1} to a considerable degree, but, since this tube is to be used as a power-output tube, the design is not intended primarily to reduce this capacitance to a low value.

The pentode shown in *c* of Plate VI has an added shield connected to the cathode, the purpose of which is to reduce the capacitance C_{pg_1}. The bulb is so shaped that an external shield can be brought near to the internal shield. The voltage ratio of this pentode is greater than 1,500, and its plate resistance r_p is of the order of 1.5 megohms.

248. Pentode with Space-charge Grid.—A space-charge grid, similar to that described in a tetrode, may be incorporated in a pentode, as in Fig. 324*h*. The inner grid having a positive potential is the space-charge grid and serves to aid the electrons in their passage through the other grids. The second grid is the control grid. The third grid is generally used as a screen grid. The arrangement is a screen-grid tetrode with a space-charge grid added which permits lower plate voltages to be used than would be necessary without the space-charge grid.

As the art progresses, other combinations of electrodes and other connections will undoubtedly develop. There are at present several special-purpose tubes which are shown in the X-ray pictures of Plates V and VI. All of these will not be described, but they are shown to illustrate the trend of present-day development.

IV. SPECIAL TUBES

249. Variable-u Tube.—Distortion in an amplifier due to nonlinearity of the characteristic curve was discussed in Chap. XII. Consider specifically the particular kinds of distortion and interference which can exist in a radio-frequency amplifier. Assuming that the plate circuits of the radio-frequency amplifier are tuned, only those frequencies produced by distortion which are in the neighborhood of the resonance frequency pass through the amplifier.

One common form of interference in a receiving set is produced by two carrier waves having frequencies such that their difference is sufficiently near the resonance frequency, denoted by n_1, to pass through the tuned circuits. The difference frequency is produced by detection in the first amplifier tube, provided its characteristic curve is nonlinear. As shown in Sec. 202 of Chap. XIX, if Eq. (593) represents the nonlinear characteristic curve, the difference frequencies are produced by all terms of the equation of powers of e higher than the first, but especially by the even powers of e. The coefficients A^{I}, A^{II}, etc., of Eq. (593) are the first, second, etc., derivatives of $\bar{\imath}$ with respect to $\bar{e}$, evaluated at the quiescent point. Hence, this kind of interference depends largely upon all of the derivatives of order greater than the first.

There is another kind of interference known as *cross talk*. This interference may occur when a strong modulated carrier acts upon the amplifier if at the same time a second carrier, to which the amplifier is tuned, is being received. Cross talk occurs whether or not the carrier being purposely received is modulated. The frequency of the interfering carrier may be only sufficiently different from n_1 for the selective circuits of the amplifier to prevent it from passing directly through the amplifier.

Cross talk is due to the modulation of the desired carrier by the message conveyed by the interfering carrier. The way in which this occurs can best be understood by a mathematical formulation of the conditions. Let the plate current of the amplifier, measured from the quiescent point, be

$$\bar{\imath}_p = A^{\text{I}}\bar{e}_g + \tfrac{1}{2}A^{\text{II}}\bar{e}_g^2 + \tfrac{1}{6}A^{\text{III}}\bar{e}_g^3 + \cdots \tag{859}$$

where $\bar{e}_g$ is the grid voltage also measured from the quiescent point. Let $e_1 = \sqrt{2}E_1 \sin \omega_1 t$ be the voltage acting on the grid due to the carrier of the desired signal; let $e_2 = \sqrt{2}E_2(1 + m_2 \sin \omega_1 t) \sin \omega_2 t$ represent the voltage due to the modulated interfering carrier. The total a-c. grid voltage is $e_1 + e_2$. This voltage is to be substituted for $\bar{e}_g$ in Eq. (859). The terms having frequencies in the neighborhood of $n_1 = \omega_1/2\pi$ are

$$\begin{aligned}(\bar{\imath}_p)_{n_1} = {} & A^{\text{I}}e_1 + (\tfrac{1}{6}e_1^3 + \tfrac{1}{2}e_2^2e_1)A^{\text{III}} \\ & + (\tfrac{1}{120}e_1^5 + \tfrac{1}{12}e_2^2e_1^3 + \tfrac{1}{24}e_2^4e_1)A^{\text{V}} + \cdots\end{aligned} \tag{860}$$

Substituting the expression for e_1 and e_2 and expanding,

$$(\bar{\imath}_p)_{n_1} = \left[\frac{di_p}{de_g} + \left(\frac{1}{2}\cdot\frac{d^3i_p}{de_g^3}E_2^2 + \frac{1}{24}\cdot\frac{d^5i_p}{de_g^5}E\,E_1^2\right)(1 + m_2\sin\omega_l t)^2E_2^2 + \frac{1}{16}\cdot\frac{d^5i_p}{de_g^5}E_2^4(1 + m_2\sin\omega_l t)^4\right]\sqrt{2}E_1\sin\omega_1 t \quad (861)$$

The first term of Eq. (861) is due only to the desired signal as amplified by an amplifier having a linear characteristic curve. The rest of the terms in Eq. (861) are due to cross modulation of the desired carrier by the modulation of the interfering signal. This cross modulation evidently depends upon the third, fifth, etc., derivatives of the curve.

Cross talk is most pronounced in receiving sets having *automatic volume control.* Automatic volume control is an automatic control of the audio output so that it remains more or less constant for a considerable range of input voltages. It is usually accomplished by causing the steady component of current, as produced by rectification of the incoming signal by the detector, to vary the negative polarizing potential of the first radio-frequency amplifier. The varying polarizing potential is produced by passing the rectified current through a suitable resistance located in the grid circuit of the amplifier. With this arrangement an intense signal causes the polarizing potential to be highly negative and often moves the quiescent point to the region of curvature of the plate-current grid-voltage curve of the amplifier. This curvature produces cross talk because of the large values of the high-order terms in the power-series development of the characteristic curve.

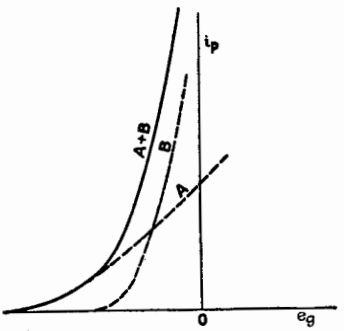


FIG. 340.—Equivalent plate-current curve for two tubes in parallel.

In 1930 Ballantine and Snow[5] described a special tube whose characteristic curve is modified so as to decrease the high-order derivatives for low values of plate current. This modified characteristic curve might be obtained by operating two triodes in parallel, one having a small value of u_p and the other a large

[5] BALLANTINE and SNOW, *Proc. I.R.E.*, **18**, 2102 (1930).

value of u_p. The two plate-current curves for these two tubes are illustrated by the dotted lines A and B of Fig. 340. The total plate current from the two tubes is shown by the full-line curve of Fig. 340. The cut-off of plate current is obviously less sharp than for curve B alone. Such a characteristic curve is well

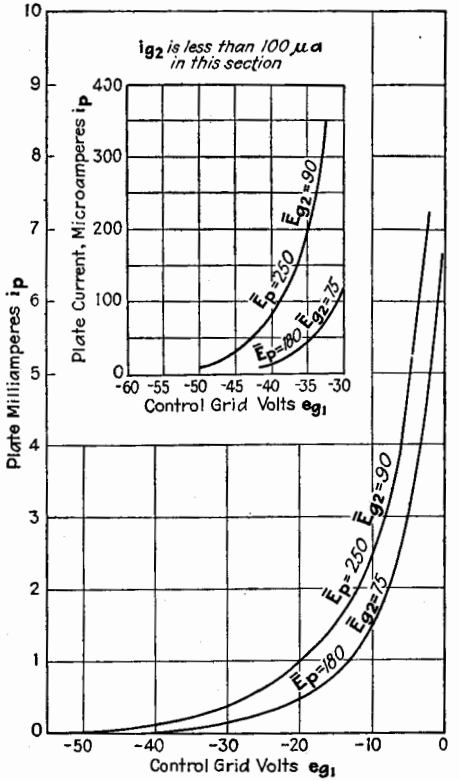


FIG. 341.—Curves of i_p *vs.* e_{g_1} for variable-u pentode, type 235 with suppressor grid. (*R.C.A. Radiotron Handbook.*)

adapted to the use of automatic volume control. Weak signals are amplified on the steep portion of the curve; strong signals cause the polarizing potential to increase negatively so that the operating point moves to the portion of the curve having a small slope.

Instead of using two tubes, Snow incorporated the properties of two tubes in a single tube. This may be done by varying, along the length of the tube, one or more of the structural dimensions of the tube. The common construction is to wind

the grid so that a portion of it has a coarser mesh than the rest of the grid. The plate-current curves of a screen-grid tetrode having a grid of variable mesh but which is similar in other respects to tube *b*, Plate V, are shown in Fig. 341.

As explained in Chap. VII, the voltage ratio of two tubes having different values of u_p, operating in parallel and on the

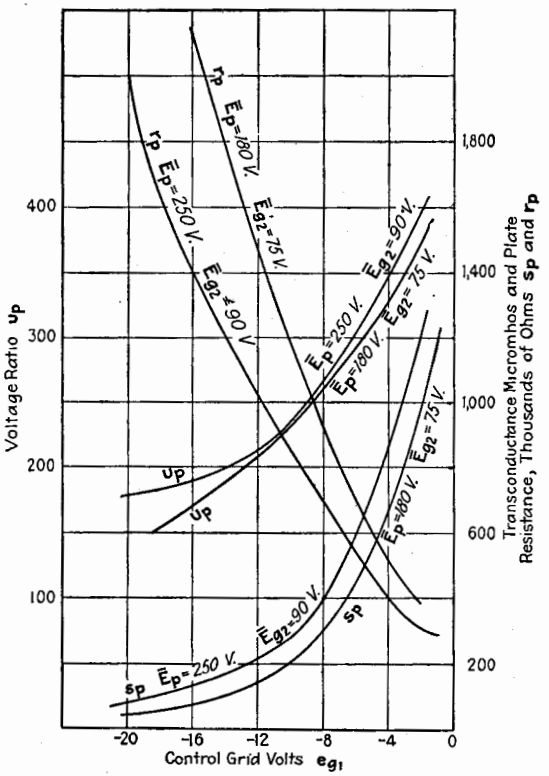


FIG. 342.—Equivalent triode coefficients of variable-u pentode, type 235. (*R.C.A. Radiotron Handbook.*)

same plate and grid voltages, always increases as the grid voltage increases. For the same reasons, the voltage ratio of the single special tube just described varies as shown in Fig. 342 which gives the dynamic characteristics of this tube.

The structural dimensions of a variable-u tube may be such that over a considerable range the plate current is an *exponential* function of the grid voltage; or

$$i_p = A\epsilon^{be_g} \tag{862}$$

The transconductance is also exponential; or

$$s_p = Ab\epsilon^{be_g} \tag{863}$$

A vacuum tube having a characteristic curve of the form of Eq. (862) can be used for *logarithmic recording* as pointed out by Ballantine.[6] The transconductance of an exponential tube and that of an ordinary tube are plotted on logarithmic paper in Fig. 343, taken from the paper by Ballantine and Snow.[5] Figure 344[6] shows the transconductance of a variable-u tube which is logarithmic over a larger range than the tube of Fig. 343 and is hence better adapted to logarithmic recording.

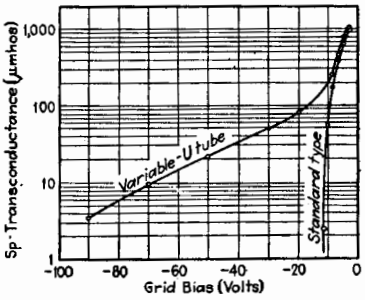


FIG. 343.—Curves of s_p *vs.* e_{g_1} for variable-u tetrode.

250. Low-grid-current Tube The second special-purpose tube to be described is especially adapted to the measurement of very small d-c. currents and is useful wherever an electrometer is used.

Although it is impossible to state who first thought of using a vacuum tube for the measurement of small currents[7] by the methods similar to those used with an electrometer, one of the first extensive discussions of the matter was published by Metcalf and Thompson.[8]

The grid of a triode when negatively polarized controls the flow of electrons to the plate by its effect upon the electrostatic field near the cathode. Although theoretically the flow of current to the grid is only the charging current which flows into the grid capacitance, there is actually a grid current due to other causes. The development of a triode to be used as an electrometer consists in the reduction of the grid current as far as possible.

The grid current in a high-vacuum triode, when the grid is polarized negatively so that no electrons pass directly to the grid,

[6] BALLANTINE, *Electronics*, **2**, 472 (1931).

[7] WOLD, U. S. patent 1,232,879 (1916). WYNN-WILLIAMS, *Proc. Cambridge Phil. Soc.*, **23**, 810 (1927); NELSON, *Rev. Sci. Inst.*, 1, **281** (1930); BENNETT, *Rev. Sci. Inst.*, 1, **466** (1930); NOTTINGHAM, *J. Franklin Inst.*, **209**, 287 (1930).

[8] METCALF and THOMPSON, *Phys. Rev.*, **36**, 1489 (1930).

is due to several causes, listed by Metcalf and Thompson as follows:

1. Leakage over the glass and insulating supports.
2. Positive ions formed by ionization of the residual gas in the tube.
3. Thermionic emission from the grid due to heating of the grid by the cathode power.
4. Positive ions emitted by the cathode.

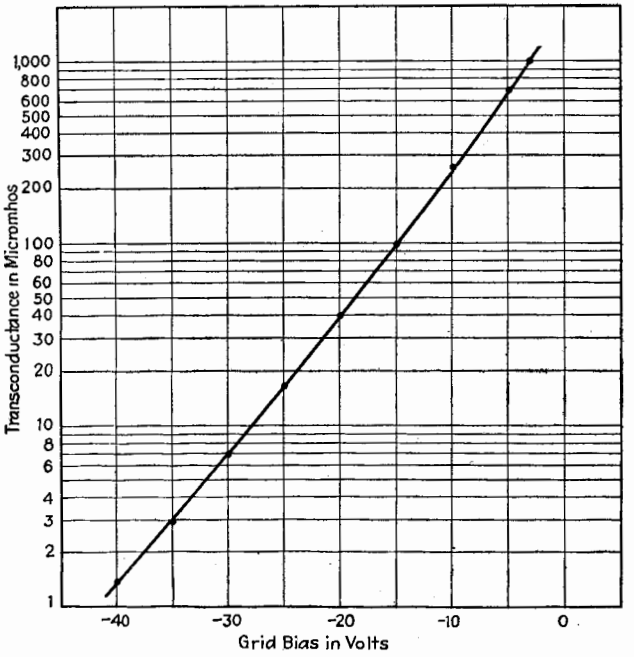


FIG. 344.—Transconductance of an exponential tetrode. (*Ballantine.*)

5. Photoelectrons emitted by the control grid due to light from the cathode.
6. Photoelectrons emitted by the control grid due to soft X-rays produced by plate-current electrons striking the plate.

Several forms of tubes have been designed to reduce grid current due to the above causes. The particular electrometer tube to be described was designed by Metcalf and Thompson and is shown as tube *d* of Plate V facing page 588.

Leakage current is reduced by the use of quartz beads to hold the supporting wires of the grid, these beads being shielded to

prevent surface contamination by sputtering or condensation of products from the filament. The control-grid connection is led out at the top of the tube. Leakage over the outside surface of the glass tube is generally reduced by winding over the tube a few turns of bare wire to serve as a guard ring, and by housing the tube in an enclosure in which a drying agent reduces the humidity of the air surrounding the tube.

Cause 2, *i.e.*, positive-ion current from the gas, is reduced by employing a plate voltage below ionization voltage of the residual gas. The plate voltage specified for the tube described is 6 volts.

Because of the low filament power used in the tube described ($\bar{E}_A = 2.5$ volts, $\bar{I}_a = 0.1$ amp.), the grid does not become hot enough to emit electrons. Hence, Cause 3 listed above is of little importance.

Positive ions from the cathode are prevented from reaching the control grid by interposing, between the control grid and the cathode, a space-charge grid operated at a positive potential of about 4 volts.

Photoelectrons from the control grid due to light from the cathode are reduced by employing a filament of thoriated wire operated at a low temperature.

Finally, photoelectrons due to soft X-rays are reduced by employing a low plate voltage (6 volts). At higher plate voltages the grid current due to this cause is appreciable.

The net grid current due to the combined action of all causes listed is shown in Fig. 345, taken from the instructions for operating the tube described. The grid current is generally less than 10^{-15} amp.; the voltage ratio of the tube is 1.0; and the plate resistance is 40,000 ohms. The grid resistance is of the order of 10^{16} ohms as compared to 10^8 or 10^9 ohms for ordinary tubes.

The electrometer tube is often used in connection with an ionization gauge for the measurement of X-rays, or in connection with a photoelectric cell for the measurement of low intensities of light. In both uses the connections are similar. If the resistance of the input device is comparatively low, the connections shown in Fig. 346*a* are used; if the resistance of the input device is very high, the connections shown in Fig. 346*b* may give a more sensitive arrangement.

Assume for illustration that a photoelectric cell is used, whose constant is

$$K = \frac{di}{dL}$$

where L is the total energy of radiation falling on the cell. If the arrangement of Fig. 346*a* is used,

$$de_g = R_c di = R_c K dL \text{ (approx.)} \tag{864}$$

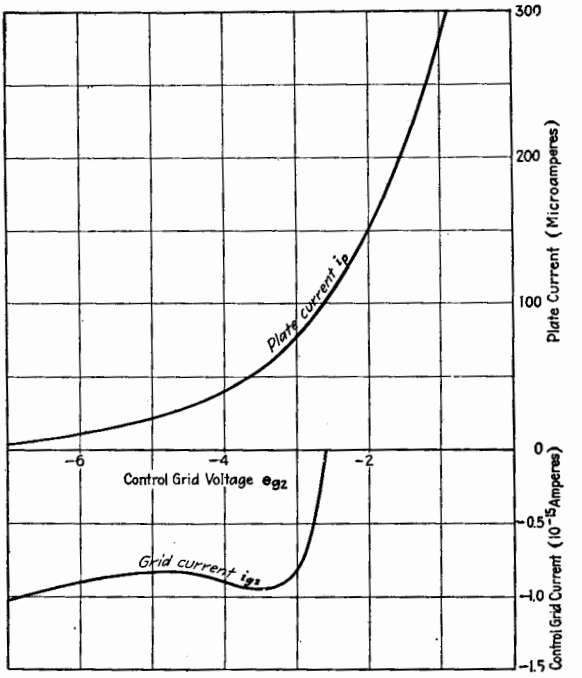


FIG. 345.—Average characteristics of tetrode type FP54. $\bar{E}_f = 2.5$ volts; $\bar{E}_{g_1} = 4$ volts; $\bar{E}_p = 6$ volts.

If the arrangement of Fig. 346*b* is used, the change of grid voltage is the change in voltage drop through the grid variational resistance r_g. Hence

$$de_g = r_g di = r_g K dL \tag{865}$$

Consequently, if r_g is greater than R_c the arrangement of Fig. 346*b* is more sensitive. This discussion assumes that the voltage

across the photoelectric cell is sufficient to produce saturation current through the cell.

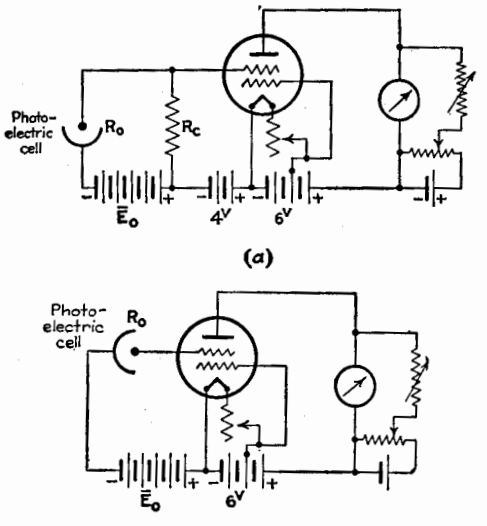


FIG. 346.—Use of low-grid-current tetrode.

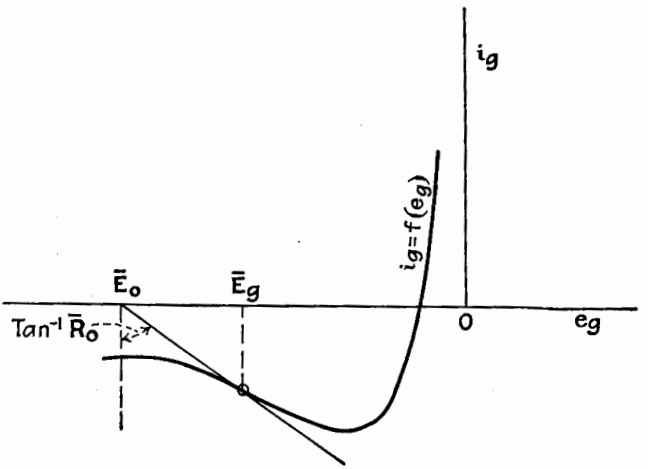


FIG. 347.—Grid-circuit diagram of low-grid-current tetrode.

The grid-current diagram for the electrometer tube, when used as shown in Fig. 346*b*, is given in Fig. 347. The steady grid voltage is determined by the intersection of the resistance line

for the external device and the grid-current curve. The conditions for maximum sensitivity can be derived when the characteristics of the external device connected to the grid are known.

Several more complicated circuits, consisting of bridge arrangements similar to those described in connection with vacuum-tube voltmeters, can be used with a certain gain in stability. These circuits are described by Nottingham[7] and also in a pamphlet[9] issued by the General Electric Company.

General References

JOBST: Über den Zusammenhang zwischen Durchgriff and Entladungsgesetz bei Röhren mit veranderlichem Durchgriff (variable), *Telefunken Ztg.*, **12,** 29 (1931).

BARTON: High Audio Power from Relatively Small Tubes, *Proc. I.R.E.*, **19,** 1131 (1931).

FAY: Operation of Vacuum Tubes as Class B and Class C Amplifiers, *Proc. I.R.E.*, **20,** 548 (1932).

[9] *Gen. Elec. Bull.* 249 A, February, 1932.

APPENDIX A

THEORY OF SUPERIMPOSED CURRENTS

Most problems in electrical engineering are concerned either with direct currents only, or with alternating currents only. In some special problems, those concerned with a microphone in the telephone art, for example, currents flow which are composed of an alternating current superimposed upon a direct current. In problems of this type, the theory and calculation of the circuits may be more involved than in the simpler cases. The elements of the theory of superimposed currents are added here because the currents in vacuum-tube circuits are more often composite than steady or alternating alone.

1. Path of Operation for Superimposed Currents.—Suppose a circuit, equivalent to a pure resistance, offers to a steady current the resistance $\bar{R}$, and to an alternating current of frequency ω the resistance $\tilde{R}$. The hypothetical circuit may consist of a conductor of negligible reactance, for which the skin effect is pronounced, making $\tilde{R}$ different from $\bar{R}$. Often we have to consider the parallel resonance circuit shown in Fig. 1, tuned to the frequency ω. Obviously the resistance of this circuit to a steady current is R which we shall denote as $\bar{R}$. At resonance, the equivalent impedance of the parallel circuit to a current of frequency ω is L/CR. To the harmonics of ω, *i.e.*, to 2ω, 3ω, etc., the capacitive reactance is small and therefore the total impedance is small.

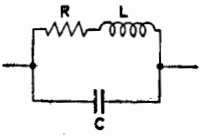


FIG. 1.—Parallel resonance circuit.

Suppose that a steady current $\bar{I}$ flows through the parallel circuit. The e.m.f. across the circuit is $\bar{E} = \bar{I}\bar{R}$. Instead of the steady current, an alternating current, $\tilde{\imath} = \hat{I} \sin \omega t$, may flow through the circuit. The instantaneous e.m.f. across the circuit is then $\tilde{e} = \tilde{\imath}\tilde{R} = \tilde{R}\hat{I} \sin \omega t$. Assume, as is usually true, that $\bar{R}$ is not a function of current and hence not effected by $\tilde{\imath}$, and that $\tilde{R}$ is also not a function of current and hence not affected by $\bar{I}$. This assumption is not true if the conductor is heated

by the current and its resistance is a function of temperature, or if the resistance of the conductor is partly made up of iron losses the value of which depends upon the total flux. Under the given assumption, the two currents $\bar{I}$ and $\tilde{\imath}$ can be superimposed with no cross effects between them, and the total e.m.f. across the circuit is the sum of the two separate e.m.f.'s or

$$e = \bar{E} + \tilde{e} = \bar{I}\bar{R} + \tilde{\imath}\tilde{R} \tag{1}$$

The total current is

$$i = \bar{I} + \tilde{\imath} \tag{2}$$

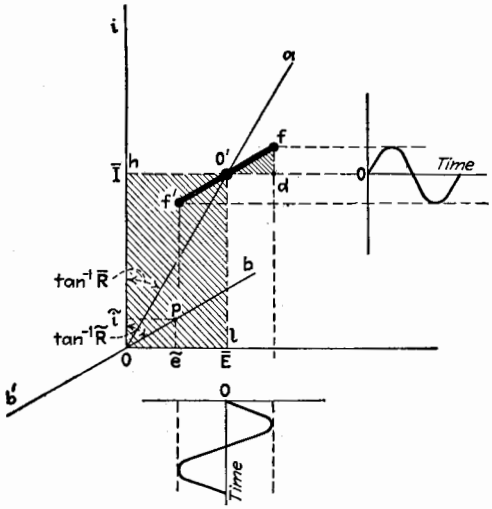


FIG. 2.—Characteristic of a conductor having a different resistance for alternating and direct currents.

Equation (1) is represented graphically in Fig. 2. Line ao is drawn making an angle with the vertical whose tangent is equal to $\bar{R}$. For every point on this line, the ratio of the abscissa to the ordinate is the constant $\bar{R}$. This line represents the steady-current resistance $\bar{R}$. Line bOb' is drawn making an angle with the vertical whose tangent is $\tilde{R}$. This line represents the a-c. resistance $\tilde{R}$. As the potential varies according to the expression $\tilde{e} = \hat{E} \sin \omega t$, the corresponding current $\tilde{\imath}$, at any time t, is determined by projecting horizontally from a point p which moves along the line $b'b$ always vertically above $\tilde{e}$. The extent of travel on each side of the origin is determined by the maximum voltage $\hat{E}$.

If voltages $\bar{E}$ and $\tilde{e}$ act simultaneously, the operating point moves along line ff' drawn through o' parallel to $b'b$. Line ff' is a plot of Eq. (1).

2. Power Relations.—The power delivered to the circuit is

$$P = \frac{1}{T}\int_0^T eidt \tag{3}$$

Substituting the values of e and i from Eqs. (1) and (2)

$$P = \frac{1}{T}\int_0^T (\bar{I}\bar{R} + \tilde{R}\hat{I}\sin\omega t)(\bar{I} + \hat{I}\sin\omega t)dt$$
$$= \bar{I}^2\bar{R} + \frac{\hat{I}^2\tilde{R}}{2} = \bar{I}^2\bar{R} + I^2\tilde{R} \tag{4}$$

Equation (4) shows that the power is the sum of the d-c. and a-c. powers taken separately. The d-c. power is represented

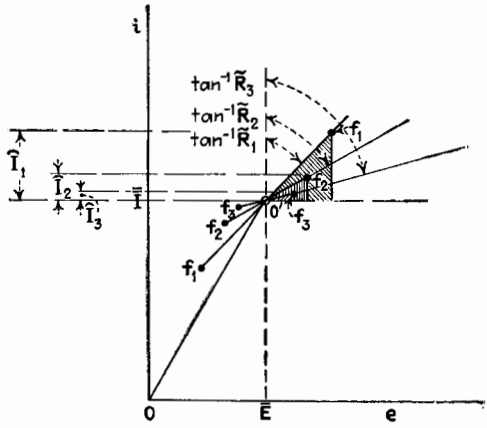


FIG. 3.—Characteristic of a conductor having a different resistance for alternating currents of different frequencies.

by the area of rectangle $hO'lO$, and the a-c. power is represented by the area of triangle $O'fd$.

3. Effect of Harmonics upon the Path.—The alternating current and e.m.f. have been assumed sinusoidal. The current may contain harmonics of the fundamental frequency $\omega/2\pi$, as $2\omega/2\pi$, $3\omega/2\pi$, etc. If the circuit offers the same resistance to all frequencies, the path is still a straight line as in Fig. 2, but the point O' is determined by the average voltage $\bar{\bar{E}}$ and the average current $\bar{\bar{I}}$. The power is now not given by the area of

triangle $O'fd$. Usually, however, the circuit offers a different resistance for each frequency, as shown in Fig. 3 by the three resistance lines $f_1'f_1$, $f_2'f_2$, $f_3'f_3$. These lines are drawn making angles with the vertical whose tangents are equal to the resistance offered to currents of frequency $\omega/2\pi$, $2\omega/2\pi$, and $3\omega/2\pi$. Three frequencies are used here for illustration, although there may be more in any actual case. If the maximum amplitudes $\hat{I}_1$, $\hat{I}_2$, and $\hat{I}_3$ of the three frequencies are represented by the projections

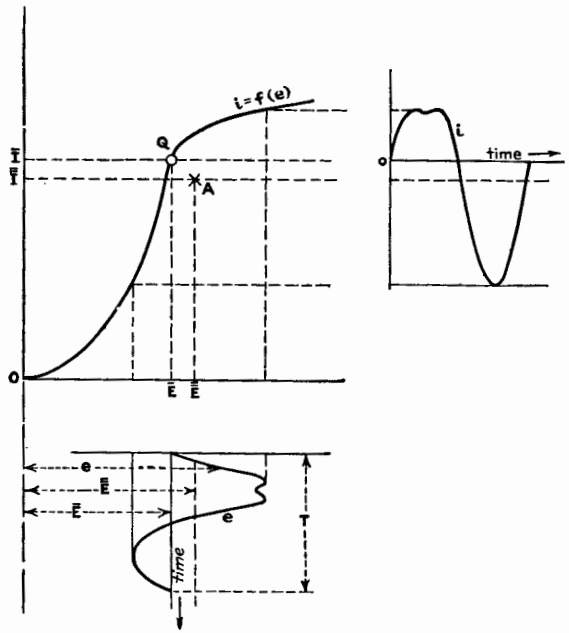


FIG. 4.—Characteristic of a conductor whose resistance is a function of the current.

on the current axis of the distances $O'f_1$, $O'f_2$, and $O'f_3$, the total power is

$$\text{Power} = \bar{I}\bar{E} + \frac{\hat{I}_1^2\tilde{R}_1}{2} + \frac{\hat{I}_2^2\tilde{R}_2}{2} + \frac{\hat{I}_3^2\tilde{R}_3}{2} \tag{5}$$

The total a-c. power is represented by the sum of the areas of the three shaded triangles of Fig. 3. The path of the operating point with harmonics, if the resistance is dependent upon frequency, is some sort of open figure, the form depending upon the phases of the several harmonics with respect to the fundamental current, as well as upon their relative amplitudes and the resistances.

4. Resistance a Function of Current.—Consider a conducting device whose resistance is a function of the current, for example, a diode or an incandescent lamp. With such devices the current and e.m.f. are connected by some functional relation which can best be represented graphically. Let Fig. 4 represent such a graph, and let

$$i = f(e) \tag{6}$$

be the equation of the curve. Assume that a constant e.m.f. $\bar{E}$ is first impressed on the conductor.

A steady current $\bar{I}$ flows such that $\bar{E}$ and $\bar{I}$ satisfy Eq. (6); or

$$\bar{I} = f(\bar{E}) \tag{7}$$

The circle on the curve of Fig. 4 represents the quiescent condition.

Now assume that e varies owing to the superimposition upon $\bar{E}$ of a periodically varying e.m.f. of the period T. This added e.m.f. may be nonsinusoidal and may have an average value not equal to zero. An illustrative wave form is shown plotted in Fig. 4. The total e.m.f. e will have an average value given by

$$\bar{\bar{E}} = \frac{1}{T}\int_0^T e\,dt \tag{8}$$

The current that flows as a result of the varying voltage e is nonsinusoidal and has an average value

$$\bar{\bar{I}} = \frac{1}{T}\int_0^T i\,dt \tag{9}$$

In order that the average point on the i-e diagram determined by $\bar{\bar{E}}$ and $\bar{\bar{I}}$ shall lie on the characteristic curve for the conductor, $\bar{\bar{E}}$ and $\bar{\bar{I}}$ must satisfy Eq. (6); or

$$\bar{\bar{I}} = f(\bar{\bar{E}}) = f\left[\frac{1}{T}\int_0^T e\,dt\right] \tag{10}$$

Equation (10) is in general not true, as is evident by integrating Eq. (6) from 0 to T and dividing by T, which gives

$$\bar{\bar{I}} = \frac{1}{T}\int_0^T i\,dt = \frac{1}{T}\int_0^T f(e)\,dt \tag{11}$$

It is true only if i is a linear function of e, *i.e.*, if $f(e) = A + Be$. This can be proved if the expression just given is substituted in

Eqs. (10) and (11). Therefore, the following general statement may be made.

When the instantaneous current i and the instantaneous e.m.f. e are connected by some relation as $i = f(e)$, the point on the i-e diagram determined by the average current $\bar{I}$ and the average e.m.f. $\bar{E}$ does not lie on the curve $i = f(e)$ unless the current i is a linear function of e.

APPENDIX B

METHOD OF TESTING AN AUDIO TRANSFORMER

The simple audio transformer usually consists of two windings of fine wire on a closed iron core. The inductances of the primary and secondary coils may be hundreds of henries and the coefficient of coupling is very near unity. The determination of the performance of such a transformer as a function of frequency is a straightforward test, although the measurement of the constants of the transformer is difficult. This appendix gives methods of determining these constants.

1. Test of Voltage Ratio of an Audio Transformer.—There are two methods of obtaining the voltage ratio of an audio

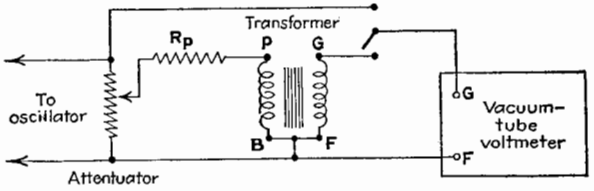


FIG. 1.—Method of determining the voltage ratio of an audio transformer, using a vacuum-tube voltmeter.

transformer. The *first method* makes use of a vacuum-tube voltmeter to obtain the ratio of output to input voltage, while the *second method* is a null or balance method.

First Method.—The connections for the first method are shown in Fig. 1. An attenuator or potentiometer, connected to a variable-frequency oscillator, impresses on the primary winding a voltage sufficient to give always the same voltage output as indicated by the vacuum-tube voltmeter. A switch is used, as shown in the figure, in order that the voltage across the potentiometer may be measured by means of the same voltmeter.

In making the test, the audio transformer should be connected as designed; *i.e.*, one particular terminal of the secondary winding (usually marked *G*) should be connected to the grid of the vacuum-tube voltmeter, and the terminal of the primary winding intended

to connect to the plate should be connected to the high-potential side of the attenuator. This is important, particularly when the impedance across the terminals is high, because the capacitance of the windings has a considerable effect upon the voltage ratio at high frequencies. The capacitance of the vacuum-tube voltmeter and of the wiring is connected across the secondary during the measurement and should be the same in magnitude as that of the circuit to which the transformer is connected when used.

Figure 2 shows curves of the voltage ratio of a particular transformer, taken by the first method. The curves are for different

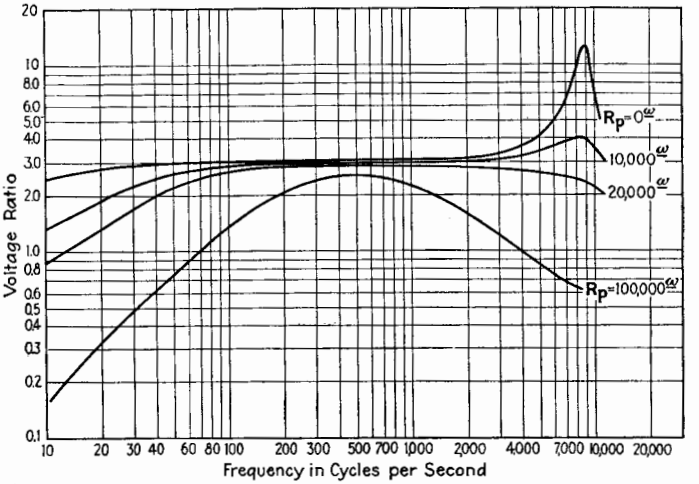


FIG. 2.—Voltage ratio of an audio transformer (TL) obtained by the method of Fig. 1.

values of R_p, the resistance connected in series with the primary coil to simulate the variational plate resistance r_p of a vacuum tube. The dropping off in voltage ratio at low frequencies is mostly due to the decrease in fractional voltage across the primary coil as the inductive reactance of the primary coil decreases. The pronounced peak at about 9,000 cycles per second is a resonance peak, the resonance circuit consisting of the capacitance across the secondary coil and the equivalent inductance of the transformer as viewed from the secondary circuit.

Second Method.—The second method of measuring the voltage ratio is indicated in Fig. 3. The mutual inductance M_a and

either R_a or R_b are varied to give a balance. The voltage ratio is $\frac{R_b}{\sqrt{R_a^2 + M_a^2\omega^2}}$. At very low and at high frequencies, where the telephone receivers and the ear are insensitive, a vacuum-tube voltmeter may be substituted for the telephone receivers to determine the balance. If the coils of the transformer are not wound in opposite directions, it is necessary to reverse the connections to the primary winding to obtain balance. If R_a is of the order of 100 ohms, the capacitance effects in the primary circuit are submerged and this reversal from the direction of connection as prescribed should make little difference in this test.

The second method gives the voltage ratio when *no* capacitance is added across the secondary winding. If the performance

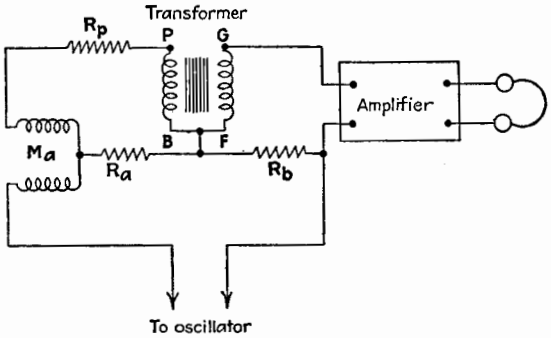


FIG. 3.—Balance method of determining the voltage ratio of an audio transformer.

of the transformer is desired with the usual capacitance across the secondary coil, the input of an amplifier should be connected between the points marked *F* and *G*.

Figure 4 shows curves of the voltage ratio obtained by the second method. The curves are for a different make of transformer from that used in obtaining Fig. 2. The several curves are for different values of R_p as indicated in Fig. 2. The dotted curve in Fig. 4 is for $R_p = 0$ and is taken by the first method. The shift in the position and the change in height of the maximum of the dotted curve and of the full-line curve for $R_p = 0$ are due to the difference in capacitance across the secondary coil in the two cases. The capacitance across the secondary coil for the full-line curve is merely the distributed capacitance of the coil, whereas the capacitance for the dotted curve is the capaci-

tance of the vacuum-tube voltmeter added to the distributed capacitance.

The first method gives more accurate results at high and at low frequencies and is more convenient and much easier in execution. The first method is recommended, except when making the measurements to determine the constants of the transformer as outlined below.

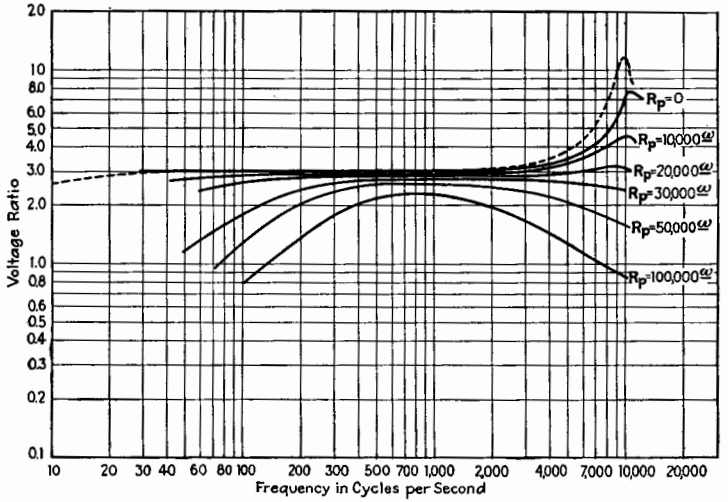


FIG. 4.—Voltage ratio of an audio transformer (AD1). Full lines obtained by the method of Fig. 3. Dotted line obtained by the method of Fig. 1.

2. Measurement of L_2/L_1.—The transformer may be represented by the circuits of Fig. 5, where C_1 and C_2 are the equivalent distributed capacitances of the transformer alone. If the impedance of the primary winding is measured on a bridge when the secondary coil is short-circuited, the measured resistance R_1' and inductance L_1', neglecting C_1, are

$$R_1' = R_1 + \frac{M^2\omega^2 R_2}{R_2^2 + L_2^2\omega^2} = R_1\left[1 + \frac{M^2/L_2^2}{\eta_2^2 + 1} \cdot \frac{R_2}{R_1}\right] \tag{1}$$

$$L_1' = L_1 - \frac{M^2\omega^2 L_2}{R_2^2 + L_2^2\omega^2} = L_1\left[1 - \frac{\tau^2}{\eta_2^2 + 1}\right] \tag{2}$$

Similarly, the resistance R_2' and inductance L_2', measured when the primary coil is short-circuited, are

$$R_2' = R_2\left[1 + \frac{M^2/L_1^2}{\eta_1^2 + 1} \cdot \frac{R_1}{R_2}\right] \tag{3}$$

$$L_2' = L_2\left[1 - \frac{\tau^2}{\eta_1^2 + 1}\right] \tag{4}$$

From Eqs. (2) and (4),

$$\frac{L_2}{L_1} = \frac{L_2'}{L_1'} \tag{5}$$

provided η_1^2 is approximately the same as η_2^2 or both are negligible in comparison with unity. Usually η_1 and η_2 are of the order of magnitude of 0.005, and their squares negligible in comparison with unity.

The measurement of the quantities in Eqs. (1), (2), (3), and (4) should be made at a medium frequency, say 1,000 cycles per second, in order that the capacitance across the winding connected to the bridge may have little effect. As a matter of fact, the effect of this capacitance, though small, should be corrected for by the following formulas:

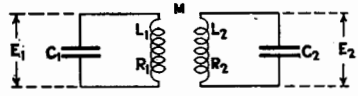


FIG. 5.—Circuits of the audio transformer.

$$\text{Corrected } R_1' = R_1'\frac{\left(\frac{R_1'}{L_1'\omega}\right)^2 + \left(1 - \frac{1}{L_1'C_1\omega^2}\right)^2}{(1/L_1C_1\omega^2)^2} \tag{6}$$

$$\text{Corrected } L_1' = L_1'\frac{\left(\frac{R_1'}{L_1'\omega}\right)^2 + \left(1 - \frac{1}{L_1'C_1\omega^2}\right)^2}{\frac{1}{L_1'C_1\omega^2}\left(\frac{1}{L_1'C_1\omega^2} - \left(\frac{R_1'}{L_1'\omega}\right)^2 - 1\right)} \tag{7}$$

Similar formulas give the corrected values of R_2' and L_2' by changing all subscripts from 1 to 2 in. Eqs. (6) and (7). The values of C_1 and C_2 are obtained by a method to be described.

Another precaution necessary for the highest precision is to make the above measurements when the flux density in the transformer core is the same, because of the fact that both the permeability of the iron and the iron losses depend upon flux density. The measurements should be made so that $I_1L_1' = I_2L_2'$, or so that the voltage across the transformer winding connected to the bridge is the same in both measurements.

3. Measurement of L_1.—The inductance of the primary coil is usually so large that accurate measurement of L_1 on a bridge is almost impossible. The following method for the determination of L_1 is convenient and generally is sufficiently accurate. Refer-

ring to Fig. 6, the primary coil and a high resistance are connected in series and to a source of sinusoidal e.m.f. The frequency of the impressed e.m.f. is chosen low in order that the capacitances across primary and secondary coils may have negligible effects. The 60-cycle mains of the laboratory give a convenient source, the frequency of which is usually accurately known. The resistance R is adjusted until the voltage, as given by a vacuum-tube voltmeter, is the same across the two elements of the circuit. Usually R_1 is negligible in comparison with $L_1\omega$. In such a case $L_1\omega = R$. If R_1 is not negligible, a reading of the voltmeter across the two

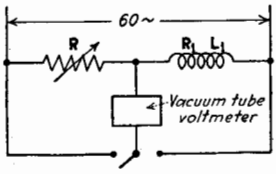


FIG. 6.—Method of measuring the inductance of the primary coil of an audio transformer.

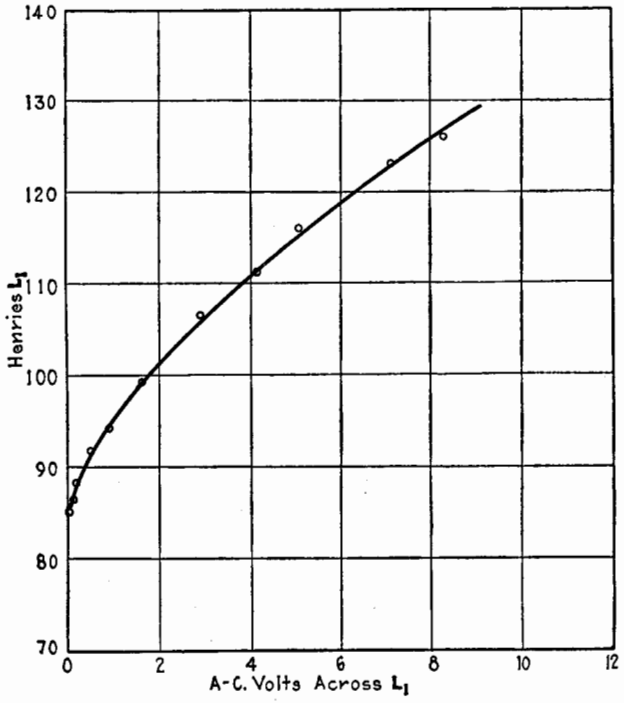


FIG. 7.—Variation of L_1 with a-c. voltage across L_1. Steady current equal to zero. (Transformer AD1.)

elements in series gives the necessary additional data to enable R_1 and L_1 to be calculated.

The magnitude of L_1 varies greatly with the a-c. voltage across the coil. Figure 7 shows the variation of L_1 of a particular transformer (AD1) with a-c. voltage impressed across the coil. In obtaining the readings for Fig. 7 for small a-c. voltages, a resistance amplifier was used in front of the vacuum-tube voltmeter to increase its sensitivity.

The inductance L_1 also depends upon the magnitude of the steady current passing through the coil. The circuit of Fig. 8 was used to obtain the curves of Fig. 9, giving the variation of L_1 with steady current. The a-c. potential difference across the coil was 0.30 volt.

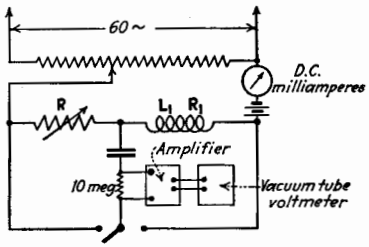


FIG. 8.—Method of measuring L_1 with a steady current flowing through the winding.

4. Determination of the Coefficient of Coupling τ.—The coefficient of coupling τ can easily be calculated from the bridge

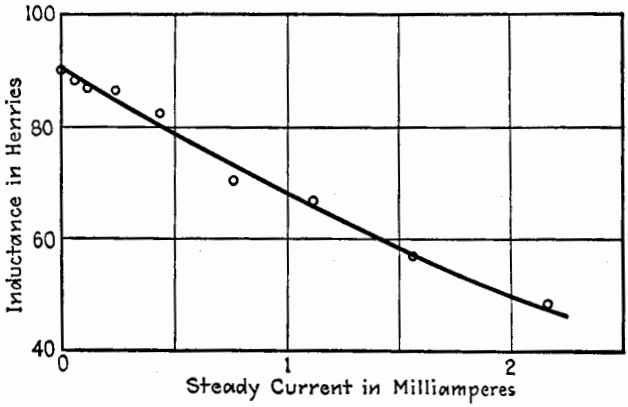


FIG. 9.—Variation of L_1 with steady current.

measurements of the short-circuited transformer. Referring to Eq. (2), if η_2^2 is negligible in comparison with unity,

$$1 - \tau^2 = \frac{L_1'}{L_1} \tag{8}$$

Similarly, Eq. (4) also gives the value of $1 - \tau^2$.

5. Measurement of the Distributed Capacitances C_1 and C_2.—In order to determine the distributed capacitance C_2, the trans-

former is connected as shown in Fig. 3 for obtaining the voltage ratio by the second method.

If E_1 and E_2 are the voltages across the primary and secondary coils, as indicated in Fig. 5, their ratio is given by

$$\frac{\mathbf{E}_2}{\mathbf{E}_1} = \frac{M/C_2}{\left(R_1R_2 - L_1L_2\omega^2 + M^2\omega^2 + \frac{L_1}{C_2}\right) + j\left(R_1L_2\omega + R_2L_1\omega - \frac{R_1}{C_2\omega}\right)} \tag{9}$$

Voltages E_2 and E_1 are in phase provided

$$R_1L_2\omega + R_2L_1\omega - \frac{R_1}{C_2\omega} = 0 \tag{10}$$

If an additional capacitance $(C_2)_{\text{add.}}$ is connected across the secondary coil, this added capacitance can be accurately adjusted for balance when M_a of Fig. 3 is zero. Equation (10) becomes

$$C_2 + (C_2)_{\text{add.}} = \frac{1}{\omega^2} \cdot \frac{1}{\left[L_2 + \frac{R_2}{R_1}L_1\right]} \tag{11}$$

The experimental procedure is to obtain $(C_2)_{\text{add.}}$ for various frequencies. If $L_2 + \frac{R_2}{R_1}L_1$ remains constant, a plot of $(C_2)_{\text{add.}}$ against $1/n^2$ should be a straight line having an intercept equal to C_2. Such a plot for a particular transformer (AD1) is shown in Fig. 10. The value of the distributed capacitance of the secondary coil of the transformer is 90 $\mu\mu$f. The straightness of the line attests the constancy of $L_2 + \frac{R_2}{R_1}L_1$, at least for the lower range of frequencies.

The distributed capacitance of the primary coil can be found by this same process with primary and secondary interchanged. A plot for this measurement is shown in Fig. 11, giving C_1 as 330$\mu\mu$f.

6. Determination of the Ratios $R_1/L_1\omega$ and $R_2/L_2\omega$.—The ratios $R_1/L_1\omega$ and $R_2/L_2\omega$, usually denoted by η_1 and η_2, involve the resistances R_1 and R_2, which are the resistances of the coils as increased by eddy-current losses, iron losses, and skin effect. These resistances are the resistances which would be measured if the distributed capacitances were zero.

To determine η_1, the resistance of the primary winding with the secondary winding short-circuited is measured on a bridge. This resistance is given by Eq. (1) but should be corrected by Eq. (6). Since the coefficient of coupling is practically unity, and η_2^2 is negligible compared with unity, Eq. (1) becomes

$$R_1' = R_1\left[1 + \frac{L_1}{L_2} \cdot \frac{R_2}{R_1}\right] \tag{12}$$

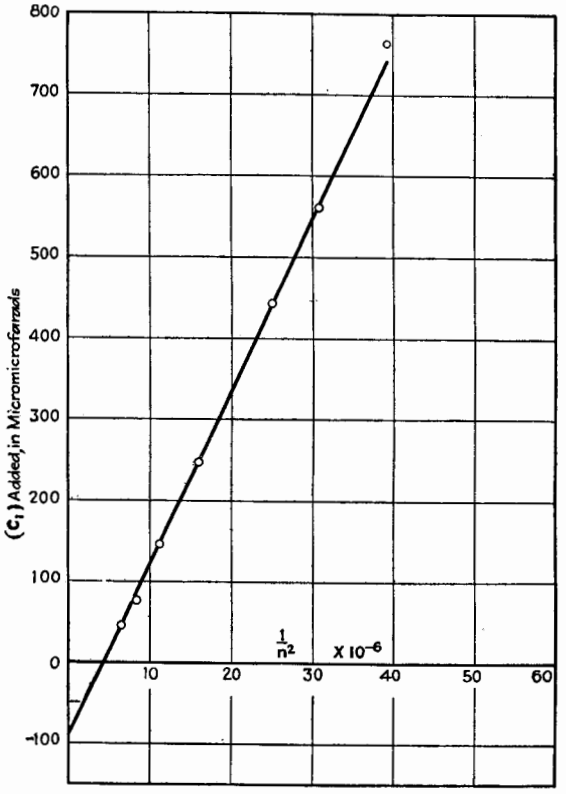


FIG. 10.—Determination of the distributed capacitance of the secondary winding of transformer AD1.

or

$$\frac{R_1}{L_1\omega} = \eta_1 = R_1' \cdot \frac{L_2}{L_1} \cdot \frac{1}{\left(L_2 + \frac{R_2}{R_1} L_1\right)\omega} \tag{13}$$

The value of $L_2 + \frac{R_2}{R_1}L_1$ can be obtained from Eq. (10) in which the ω for zero phase shall be denoted by $\omega_{\phi=0}$. Substituting in Eq. (13),

$$\eta_1 = R_1' \frac{L_2}{L_1} \cdot \frac{\omega_{\phi=0}}{\omega} \cdot C_2\omega_{\phi=0} \tag{14}$$

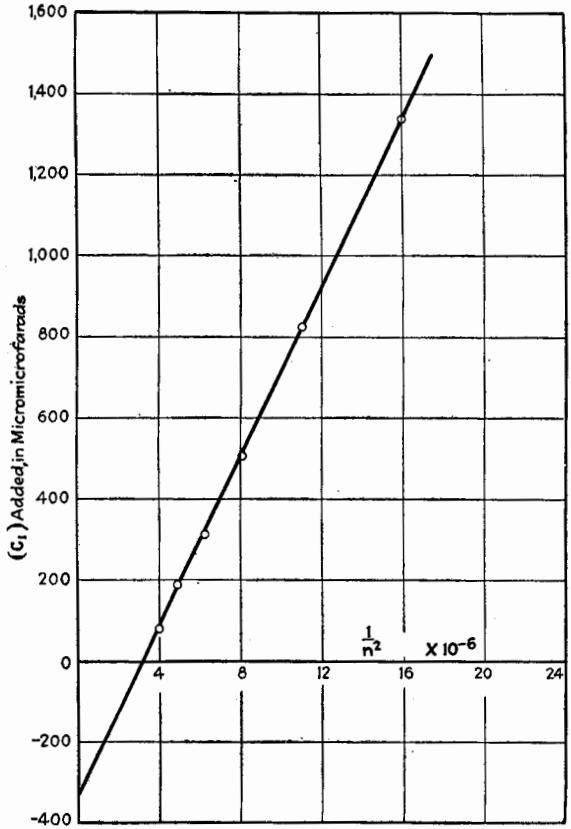


FIG. 11.—Determination of the distributed capacitance of the primary winding of transformer AD1.

The magnitude of $\omega_{\phi=0}$ can be obtained from the point where the line crosses the axis in the plot of added secondary capacitance against $1/n^2$, as shown in Fig. 10.

The corresponding expression for η_2 is given by

$$\eta_2 = R_2' \frac{L_1}{L_2} \cdot \frac{\omega_{\phi=0}}{\omega} \cdot C_1\omega_{\phi=0} \tag{15}$$

where the $\omega_{\phi=0}$ is obtained from the plot, Fig. 11.

Actually, the high-frequency resistances R_1 and R_2 do not differ greatly from the steady-current resistances except possibly at the higher frequencies.

7. Constants of a Particular High-grade Audio Transformer.—The constants are given for the particular transformer (AD1) for which the data for Figs. 4, 7, 9, 10, and 11 were taken. These constants apply to the transformer when no steady current flows through either winding and when the applied a-c. potential used for measurement purposes is very small.

CONSTANTS OF A TYPICAL TRANSFORMER (AD1)

$$\frac{L_2}{L_1} = 9.0$$

Turns ratio = 3.0

$L_1 = 85$ henries
$L_2 = 765$ henries
$C_1 = 330\mu\mu f$
$C_2 = 90\mu\mu f$

Natural resonance frequency of primary coil = 952 cycles per second
Natural resonance frequency of secondary coil = 607 cycles per second

$\tau = 0.998$
$\eta_1 = 0.0026$ at 1,000 cycles per second
$\eta_2 = 0.00139$ at 1,000 cycles per second

General References

WILLIAMS: Low Frequency Intervalve Transformers, *B.I.E.E.*, **1**, 158 (1926).
THOMSON: Audio Frequency Transformers, *Proc. I.R.E.*, **15**, 679 (1927).
THOMSON: Characteristics of Output Transformers, *Proc. I.R.E.*, **16**, 1053 (1928).
KOEHLER: The Design of Transformers for Audio Frequency Amplifiers with Preassigned Characteristics, *Proc. I.R.E.*, **16**, 1742 (1928).

NAME INDEX

SUBJECT INDEX

A

B

C