RADIO COMMUNICATION SERIES

BEVERLY DUDLEY, CONSULTING EDITOR

Graphical Constructions for Vacuum Tube Circuits

This book is produced in full compliance with the government's regulations for conserving paper and other essential materials.

Graphical Constructions for Vacuum Tube Circuits

BY

ALBERT PREISMAN, A.B.E.E.

Director of Engineering Texts and Consulting Engineer
Capitol Radio Engineering Institute

FIRST EDITION
SECOND IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK AND LONDON
1943

GRAPHICAL CONSTRUCTIONS FOR
VACUUM TUBE CIRCUITS

THE MAPLE PRESS COMPANY, YORK, PA.

TO

Edith, Rita and Cynthia

PREFACE

This book is designed to fill a gap in the literature on vacuum tubes, *viz.*, graphical constructions. By graphical constructions are meant those geometric manipulations by which are obtained solutions to problems on nonlinear circuits, particularly those involving vacuum tubes. It is therefore evident that ordinary graphs and charts, used for the easy solution of analytical formulas, are not the subject matter of this book.

While the author realizes that the engineer and scientist usually favor the analytical method of approach, he is also aware that many practical problems are amenable solely to graphical or experimental methods of attack, and he feels that this book may serve a useful purpose in presenting the former of these two methods. However, he has not hesitated to employ analytical methods in conjunction with the graphical where such procedure was of value, and thus the reader will often find an analytical derivation in the body of the text, as in the chapters on balanced amplifiers and on detection.

Much of the material incorporated here is original, and a good deal of this appeared in the *RCA Review* and in *Communications*. The author is indebted to these periodicals for permission to include this material in the present text. However, in many instances the discussion has been expanded and also revised, as in the chapter on balanced amplifiers.

In order to forestall any criticism regarding the bibliography at the end of each chapter, the author hastens to explain that his choice was governed by the following considerations:

1. Only those articles which he had read and digested were included.
2. Is the reference basic and still correct?
3. Is the article the most recent, and does it correct errors in previous articles?
4. Is it readily available to the American public? (Only in rare instances are foreign references cited.)

While such a choice may result in a list far less imposing than those found in other texts, it is hoped that the reader will find the references more readily available, less repetitious, and also less contradictory and confusing.

There has been no attempt to make this book a complete exposition of graphical methods. If it gives the reader a fundamental grasp of the subject and proves of value to him in his work, its purpose will have been achieved.

No work, no matter how humble, is due solely to one man's effort. I am only too happy to acknowledge the aid and encouragement given me by my wife, who helped greatly in the typing and preparation of the manuscript. I also wish to acknowledge the assistance given me by Dr. Alfred N. Goldsmith, who was also instrumental in obtaining the comments and criticisms of others, particularly E. W. Herold, who reviewed the third chapter and furnished me with many helpful suggestions and criticisms. And finally I wish to express my thanks to those other members of the Radio Corporation of America who passed on the merits and value of this book.

ALBERT PREISMAN.

SILVER SPRING, MD.,
August, 1943.

CONTENTS

GRAPHICAL CONSTRUCTIONS FOR VACUUM TUBE CIRCUITS

CHAPTER I

THE NONLINEAR-CIRCUIT PROBLEM

1. Introduction.—Vacuum tubes belong to a large class of circuit elements known as nonlinear parameters, and accordingly the study of *nonlinear circuits* has become an important branch of engineering. It must not be supposed, however, that it constitutes a subclass of circuit theory. On the contrary, the most general type of circuit is the nonlinear type, and linear passive networks constitute only a special class of the above. Nevertheless, most books on circuit theory deal with linear passive networks, and the literature that has accumulated on this special type of circuit is both enormous and ever increasing. In recent years, particularly since the advent of vacuum tubes, the need for a treatment of the more general subject of nonlinear circuits has become more and more acute, but the subject does not appear to have had the attention that it warrants. At any rate, the published results of such studies have not been satisfying or sufficiently general to be regarded as a real start in this most difficult of problems. From a practical viewpoint, the methods proposed have been, in the main, special and not sufficiently simple to attract the average engineer.

However, since the need for methods of solution becomes more pressing each day, it is the intention to present here applications of the methods to practical vacuum-tube problems. While those employed are mainly graphical, any combination of analytical and graphical solutions that gives an answer in the least time or with the least effort will be utilized.

2. General Considerations—Complete Solution.—To indicate the difficulty of the problem, it may be worth while to review very briefly linear-circuit theory. If we have an n-mesh network, we

can write n equations for the circuit—one equation for each mesh. Thus

$$\begin{aligned} i_1Z_{11} + i_2Z_{12} + \cdots + i_nZ_{1n} &= e_1 \\ i_1Z_{21} + i_2Z_{22} + \cdots + i_nZ_{2n} &= e_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ i_1Z_n + i_2Z_{n2} + \cdots + i_nZ_{nn} &= e_n \end{aligned}$$

where Z_{11} is the total impedance of mesh 1; Z_{22}, of mesh 2, etc.; Z_{12} the impedance common to meshes 1 and 2, Z_{n2} the impedance common to meshes n and 2, etc., and i_1 and e_1 the current and voltage, respectively, in mesh 1, etc.

Suppose that Z_{11} consists of an inductance, capacity, and resistance in series. It can then be written operationally as $[Lpi_1 + i_1R_1 + (i_1/pC_1)]$ where $p = d/dt$ and $1/p = \int dt$. In the same way, impedances in the other meshes can be written as above. It will be observed that the equations are integro-differential equations involving the time or, if the parameters are distributed along the length, involving distance as well as time. For lumped circuits, only the time is involved.

The latter can be reduced to ordinary second-order differential equations by taking the derivative of the mesh equations with respect to time and then employing any standard method of solution and inserting initial conditions in order to evaluate the integration constants. Such solutions are available because the equations are linear in the derivatives and the function itself; *i.e.*, mathematicians have discovered ways of solving these. The solution is known as the *classical method.*

These equations can alternatively be solved by Heaviside's operational methods or by Laplacian transforms, and the solution will be complete with initial conditions automatically inserted. Because these methods are simpler and more straightforward, they are generally preferred to the older classical method.

In either case, another feature that is of great importance is the fact that each voltage imposes its current flow independently of the other voltages present in the network; *i.e.*, the principle of superposition holds, and the effects are directly additive.

Now suppose that L or R or C is not a constant, but a function of the current through or voltage across it. Specifically, let $R = \phi(i)$, so that R is now a nonlinear parameter. We now have in our mesh equation a term $i\phi(i)$, and our differential equations

are no longer linear. No general mathematical methods are known for solving such equations. The *principle of superposition* no longer holds: each voltage reacts (cross modulates) with all the others, indeed, may even be considered to react with itself, so that a sinusoidal voltage may be considered to cross modulate with itself to produce harmonic currents and voltages.

The solution thus becomes far more complicated. Some simple special nonlinear differential equations, such as the Bessel equation, can be solved by means of series expansions. However, as stated above, these solutions are special, and the results so involved that it is usually impossible to find the optimum value of the parameters which will give maximum results, such as, for instance, maximum power output in a particular impedance. The solutions are thus far from satisfactory.

3. The Steady-state Solution.—The above discussion is concerned with the complete solution of the network, the transient as well as steady state. Suppose only the latter is desired: is the problem simplified to a satisfactory degree? The answer is, unfortunately, no. In linear networks, the operational expressions reduce to complex expressions, for p may be replaced by $j\omega$. The problem then reduces from a set of differential equations to a set of algebraic equations involving the complex variable $j\omega$, and the method of solution is well known to the average electrical engineer.

Since R, L, and C are assumed constant, the algebraic equations are linear in the various unknown currents; and these may be solved for by the method of determinants. Indeed, a whole host of general network theorems may be evolved from the theorems on determinants; the latter may be manipulated by matrix algebra, for instance, to enable one to find circuits equivalent to the given circuit. These theorems and manipulations have been the subject matter of many texts and constitute a remarkable body of knowledge in themselves.

Once again, let us inquire what happens if, for example, some R is a function of the current i, *i.e.*, if $R = \phi(i)$. Our steady-state equations are no longer of the first degree, the method of determinants is no longer valid, and all the theorems dependent upon it no longer hold. We now have to solve n equations in the various i's of degree higher than the first, and no general method of solution is known for such (nonlinear) equations.

4. The Terminal Characteristic.—What can be done in the situation described above? We must go back and study the circuit anew. For simplicity (if such a term can be applied) let us confine our attention from now on to nonlinear resistances. One of the first questions to arise is what do we mean by resistance? The answer may be the ratio of the voltage across the resistor to the current through it. If the resistance is linear, then this ratio is a constant and independent of the voltage and current. But if the resistance is nonlinear, then the ratio of voltage to current is a function of either, and we cannot specify the value of resistance until we know the current. But since the latter is usually the unknown and in ordinary (linear) circuit theory is expressed as a function of the voltage and circuit parameters, we are immediately enmeshed in the vicious circle of simultaneous solution, with no general method of extricating ourselves.

Since the ratio of voltage to current is no longer a constant, so that the simple relationship known as *Ohm's law* no longer holds, we are forced back to the more fundamental concept of the functional relationship between current and voltage, rather than the ratio of the two. This relationship is known as the load line or, better still, as the terminal characteristic. For a linear resistance, the terminal characteristic is a first-degree, or linear, equation, *viz.*,

$$i = ke$$

where k is a constant, known as the *conductance*. Its graph is a straight line, and this explains why it is called a linear resistance. For a nonlinear resistance the terminal characteristic is other than a first-degree equation. It may be of second degree or higher; it may be transcendental; it may even be discontinuous.

Attempts have been made to assign resistance values to such a parameter. Thus, the ratio of the voltage to the corresponding current has been called the d.c. resistance, and the ratio of a change in voltage to the corresponding change in current has been called the variational, incremental, or a.c. resistance of the device. Graphically, if the terminal characteristic be plotted, the former refers to the cotangent of the angle that the secant (line joining the origin to the point at which the resistance is to be found) to the terminal characteristic curve makes with the voltage axis, and the latter refers to the cotangent of the angle

that the tangent line to the curve makes with the voltage axis, *i.e.*, the reciprocal of the derivative of the curve.

The question now arises as to just how the terminal characteristic is to be expressed. For a vacuum-tube diode, for example, the plate current is practically proportional to the $\frac{3}{2}$ power of the plate voltage, when the latter is positive. It is zero for negative plate voltages. The terminal characteristic is therefore a function whose derivatives are discontinuous over the range of plate voltage normally encountered in rectifier practice, *i.e.*, for positive and negative values. In a triode, for example, the range of electrode voltages is often chosen so that the plate current is never driven down to zero (cutoff), in order to avoid the above type of discontinuity and consequent generation of distortion products. In balanced amplifiers, on the other hand, such discontinuity of action of either tube may be permitted (see Chap. V).

5. Power-series Representation.—One method for handling the forms of the terminal characteristic discussed above is that of the *power series*. Here the relationship between e and i is written as a series in ascending powers of e.

$$i_1 = k_1 e_1 + k_2 e_1^2 + k_3 e_1^3 + \cdots + k_n e_1^n \tag{1}$$

Now suppose that the above parameter is in series with a linear resistance (to simplify matters), and a voltage e, which is some function of time. We wish to find the current flowing through the two resistors in series. This is the typical problem encountered in *nonlinear circuit theory*. We know the relationship between the current and voltage across each parameter, but we cannot tell how much of e is across either until we know the current through each. But we cannot know this until we know the voltage across either, for the current is a nonlinear function of the voltage in the case of the first-mentioned parameter. This is the vicious circle mentioned previously. There are, however, two pieces of information available in the form of *Kirchhoff's laws*. The first is that the current is the same for both resistors, since they are in series, and the second is that the sum of the two voltage drops is equal and opposite to the impressed voltage e.

With this in mind, we can proceed as follows: Let the terminal characteristic for the linear resistor be

$$i_2 = \left(\frac{1}{R}\right) e_2 \qquad \text{or} \qquad e_2 = i_2 R \tag{2}$$

Then

$$e_1 = e - e_2 = e - i_2 R \tag{3}$$

and

$$i_1 = i_2 \tag{4}$$

so that the subscripts for the current can be dropped. If we substitute Eq. (3) in Eq. (1), we obtain

$$i = k_1(e - iR) + k_2(e - iR)^2 + k_3(e - iR)^3 + \cdots + k_n(e - iR)^n \tag{5}$$

This is now an equation in one unknown of the nth degree and can be solved, at least approximately. We note that, if the second resistor were also nonlinear, then Eq. (2) would not be so simple, we should face the task of solving simultaneously two equations of degree higher than the first, and the method of substituting from Eq. (3) in Eq. (1) would not have been so simple to apply, if at all possible. Indeed, even in the simpler case of one resistance linear, the solution is not at all easy, and a method of successive approximations has been worked out by Carson[1]* to facilitate matters.

If both resistors had been linear and of value R_1 and R_2, respectively, then the method outlined above would have led to the result

$$i = \frac{e}{R_1 + R_2}$$

which is the well-known expression for Ohm's law as applied to two resistors in series. Indeed, a linear resistance is one whose terminal characteristic is expressible by a power series having only one term, and that of the first degree. Simultaneous solution is then easy and, in the case of an n-mesh network, can be readily effected by the use of determinants.

6. Objections to the Power-series Method.—The power-series method is open to some serious objections, of both a practical and a theoretical nature. From a practical viewpoint, it is cumbersome, laborious, and too involved for everyday application. However, where the curvature of the characteristic is not

* The superior numerals refer to the items listed in the Bibliography at the end of each chapter.

too "violent," so that a few terms of the series are adequate, it is of great value, as it reveals the fundamental mechanism of harmonic distortion, modulation, and detection.

From a theoretical viewpoint, its use for characteristics that have sharp bends is questionable or at least nonorthodox. A sharp bend, in mathematical parlance, means discontinuities in the successive derivatives of the function; *i.e.*, they cease to exist at the point where the "break" occurs. Now the usual justification for a power-series representation of a function is that it is really a *Taylor* or a *Maclaurin expansion*, and hence the coefficients represent the values of the successive derivatives at the point about which the expansion is made. If these derivatives do not exist, then the Taylor or the Maclaurin expansion does not exist. Nor are these expansions valid when extended to a point of the curve for which the derivatives do not exist.

However, it is not necessary that the coefficients be evaluated by the Taylor expansion. They may be chosen by the process of *least squares*, or zonal harmonics. As many points of the curve are substituted as there are terms required in the series and the resulting equations (in which e and i are known) are then solved for the coefficients, which are now the unknowns. This is not difficult, since the equations are linear in the coefficients.

Suppose a characteristic with a sharp bend is approximated by a five-term power series, and it is then decided to use a six- or seven-term power series for greater accuracy. Since the characteristic has a sharp bend or break, its derivatives do not exist at this point, and a single Taylor series is not valid. The power-series coefficients must be chosen in some other manner as indicated above. Now it will be found that, if a seven- instead of a five-term series is desired, all the first five coefficients will be different from their original values; *i.e.*, one cannot merely add to the original five terms two more terms. The change in the values of the first five coefficients may or may not be small; this is immaterial, for the series is not the usual type of convergent series, where more terms can be added to those originally determined to obtain greater accuracy. The latter feature is true of a Taylor series and of the Fourier series. Thus, objections of a fundamental, mathematical nature may be raised concerning the use of a power series to represent a nonanalytic function (one whose derivatives do not exist at every point). However, it is

found that a power series with a sufficient number of terms represents a function with a sharp bend in it to a sufficient degree of accuracy for all practical purposes, and the main objection to a power series for such a function is that a prohibitive number of terms are required.

In passing, it is of interest to note that, if the current is a function of two voltages, such as that of the grid and plate of a triode, then a double power series is required, unless the amplification factor of the tube is a constant. Such a series is employed in Chap. III. For tetrode, pentode, and other multielectrode tubes, power series in many variables can be employed. Needless to say, the computations become very involved.

7. Fourier-series Method of Representation.—In order to avoid the use of series with so many terms, a *Fourier-series method of expansion* has been employed. The Fourier series is usually employed to represent the periodic wave shape of some quantity that is a function of time, but it may be used to represent the functional relationship between any two quantities, such as current and voltage. Although the series then exhibits the current as a periodic function of the voltage, this is of no consequence if the range of operations is limited to one cycle or less.

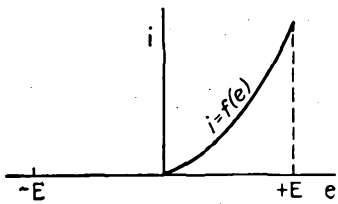


FIG. 1.—Diode-rectifier terminal characteristic.

We shall now illustrate this method. In Fig. 1 is shown a terminal characteristic such as that of a diode rectifier. We choose an upper and lower limit for the voltage range, *viz.*, $+E$ and $-E$, respectively. We now introduce an angle variable

$$\theta = \frac{\pi e}{E}$$

such that to every value of e there is a corresponding value of θ. Note that for $e = \pm E$, $\theta = \pm\pi$, respectively; *i.e.*, as e ranges from $-E$ to $+E$, θ ranges from $-\pi$ to $+\pi$, which is one cycle.

The current i can now be expressed in terms of a Fourier series as a function of θ and hence of e. Thus

$$i = \frac{b_0}{2} + \sum_{n=1}^{\infty} a_n \sin n\theta + b_n \cos n\theta \qquad (6)$$

As mentioned previously, the range of θ is limited to one cycle; that is, e varies from $-E$ to $+E$. The value of E must therefore be chosen large enough to include all variations in e that are to be encountered.

The values of the coefficients are to be calculated in the usual manner, analytically,

$$b_n = \frac{1}{E}\int_{-E}^{E} f(e)\cos n\theta\, d\theta \quad \text{and} \quad a_n = \frac{1}{E}\int_{-E}^{E} f(e)\sin n\theta\, d\theta \quad (7)$$

or from the graphical plot of $f(e)$ by a schedule analysis. The advantage in representing the terminal characteristic $f(e)$ by this series rather than by a power series is that the former requires fewer terms for a given accuracy of representation if the characteristic has a sharp bend and the questions as to the validity of the series are absent. The restrictions are very few, such as the requirement of a finite number of maxima and minima per cycle and no infinite values. Most functions encountered in practice fulfill these restrictions.

A further refinement suggested by Barrow[2] is artificially to prolong the characteristic in any arbitrary manner, which also means to extend the limits for representation of one cycle. A schedule analysis may then be made, and the desired portion of the characteristic will be represented with the same accuracy as before with fewer terms and hence less ensuing labor.

Suppose a nonlinear and a linear resistance are connected in series with a given voltage. Let the former be represented by a Fourier series such as Eq. (6), and the latter by R. If the same procedure as that described for the power-series method be followed here in order to find the current through the resistors, a transcendental equation instead of a nonlinear algebraic equation will have to be solved and no general method of solution exists. If both resistances are nonlinear, the situation is even worse. Hence this method is hardly suited to solving nonlinear circuits of the type described above.

8. Application of the Fourier-series Method.—The Fourier-series method may be applied in the evaluation of the distortion products produced in a circuit if the relation between current through and voltage across the circuit is known. Such an overall relation, or terminal characteristic, must first be found, say, by

the power series or a graphical method, and then this relation expressed as a Fourier series, of the form given by Eq. (6).

Now, let $e = f(t)$; *i.e.*, let e be a given function of time. Specifically, let

$$e = E_m \sin \omega t \tag{8}$$

If Eq. (8) be substituted in Eq. (6), terms of the form

$$b_n \cos\left(\frac{n\pi E_m \sin \omega t}{E}\right) \quad \text{and} \quad a_n \sin\left(\frac{n\pi E_m \sin \omega t}{E}\right)$$

are encountered in the expansion for i, the current. These can be expanded in terms of Bessel functions by methods due to Jacobi.

$$\left.\begin{aligned} \cos \frac{n\pi E_m \sin \omega t}{E} &= J_0 \frac{n\pi E_m}{E} + 2 \sum_{k=1}^{\infty} J_{2k} \frac{n\pi E_m}{E} \cos 2K\omega t \\ \sin \frac{n\pi E_m \sin \omega t}{E} &= 2 \sum_{k=1}^{\infty} J_{2k-1} \frac{n\pi E_m}{E} \sin (2k - 1)\omega t \end{aligned}\right\} \tag{9}$$

in which $J_K(n\pi E_m/E)$ is a Bessel function of the first kind of order k and modulus $(n\pi E_m/E)$. Since Bessel functions are becoming better known to engineers, this expansion is not an impractical operation for the technical man. The terms in the resulting series may be regrouped in individual series involving $\cos 2k\omega t$ and $\sin (2k - 1)\omega t$ so as better to exhibit and evaluate the various frequency components. However, since this all implies that the over-all terminal characteristic has already been determined and then facilitates the evaluation of the distortion products, it is not the method that we seek, *viz.*, to determine the over-all terminal characteristic when the individual terminal characteristics, the method of connection, and the impressed voltage are known. Moreover, the Fourier-series method applies to single-valued functions only. We shall see later that if reactances are present in the circuit the over-all terminal characteristic is a closed loop, which means that the current is a multivalued function of the voltage. It is therefore evident that the Fourier series cannot apply to this type of circuit, which is an unfortunate limitation.

9. The Graphical Method.—The discussion of the graphical method, which is the main theme of this book, will be deferred to subsequent chapters. We shall proceed first to some elementary considerations concerning vacuum tubes.

BIBLIOGRAPHY

1. CARSON, J. R..: *Proc. I.R.E.*, 1919.
2. BARROW, W. L.: Contribution to the Theory of Non-linear Circuits with Large Applied Voltages, *Proc. I.R.E.*, August, 1934.

CHAPTER II

THERMIONIC VACUUM TUBES

1. Introduction.—In this chapter some of the fundamental characteristics of thermionic vacuum tubes will be discussed. It will thus serve as an introduction to the matters developed in succeeding chapters.

The thermionic vacuum tube is one of the most important nonlinear resistances in use today. It owes its properties to the fact that in it are to be found electrons "in the open," as it were, divorced from the positive ions in whose company they are normally to be found. We therefore begin with a rough picture of conditions in the ordinary metallic conductor.

In a solid metallic conductor, we find an orderly array of atoms in definite geometric configurations known as a *crystal lattice structure*. Consider a single crystal of the metal. The atoms are bound together by certain atomic forces that constrain them to maintain the crystal configuration in spite of thermal agitation present at, say, room temperatures.

Each atom is made up of a positive nucleus, around which rotate electrons in orbits at various distances from the nucleus. In the crystal structure, the atoms are so close together that the outermost orbital electrons are associated as much with one nucleus or atom as with another and hence are free to move about in the interior of the crystal, and indeed from one crystal to another of the conductor. Depending upon the metal, there may be one, two, or even three free electrons per atom. These free electrons form a kind of gas within the conductor, and it is their circulation around a closed metallic path that constitutes the ordinary current flow in an electrical circuit.

2. The Potential Barrier.—It is evident that except for a relatively slight hindrance to their motion, known as *ohmic resistance*, the free electrons find little difficulty in darting about in the conductor under the impress of thermal agitation. The reason is to be found in the balanced attractions of the various positive ions

(atoms that have lost one or more electrons) if the electron is not too close to any one ion. The region within a crystal may therefore be visualized as a space consisting of strong attractive centers where the atoms are located and relatively force-free regions in the remaining parts of the crystal. The corresponding electrical potential field may therefore be compared to the gravitational field of a series of plateaus pitted with deep holes (representing the location of atoms). If the holes are filled with "bound" orbital electrons, then the "free" electrons can hurdle these holes and roam the plateaus. The analogy is admittedly naïve but does indicate in a sense the situation concerning bound and free electrons.

Suppose an electron tries to pass out of the interior of the metal through the boundary surface. Immediately the forces of attraction of the ions unite to pull it back, for the electron is now leaving them all behind, whereas before it was immersed in their force fields. The electron thus experiences a potential barrier to its escape; this barrier is known as the *work function* of the material, *i.e.*, the work required to move an electron through the boundary surface and sufficiently far from it to render the attractive forces negligible. The electron is then completely free to pursue its course outside the metal.

3. Methods of Producing Emission.—There are several ways of imparting sufficient kinetic energy to the free electrons:

1. Bombarding them with other electrons, or particles, such as positive ions or metastable atoms. If these possess sufficient kinetic energy, they can impart this energy or a suitable portion of it to the free electrons to enable them to overcome the potential barrier at the surface and thus escape. This is known as *secondary emission.*

2. Illuminating the metal with light photons whose frequency f, hence energy content fh (where h is Planck's constant, $= 6.55 \times 10^{-27}$ erg sec.) is sufficient to enable the electrons to pass the potential barrier. This is known as *photoelectric emission.*

3. Producing a strong electric field at the surface of the metal and thus pulling them out. If this is done by impressing a potential between the metal and another electrode a reasonable distance away (and making the latter the anode), it will be found that a high voltage is necessary. However, high fields may be produced by positive ions close to the surface without the need of

an external potential, as in the case of positive mercury-gas ions, or possibly metastable atoms near the surface of a pool of mercury (cathode) or, in the case of barium ions, on the surface of the core metal of an oxide-coated filament. However, the method as normally considered is that of the application of a strong field due to an external potential and is known as *high-field* or *auto-electronic emission.*

4. Heating the crystal lattice structure and thus imparting to the free electrons sufficient energy to escape. This is known as *thermionic emission* and is the method employed in thermionic vacuum tubes.

4. Thermionic Emission.—Formulas for the thermionic emission of electrons have been developed. The formula employed today is the Richardson-Dushman equation

$$i = A_0 T^2 \exp \frac{-\varphi e}{kT} \tag{1}$$

where i is the emission current per square centimeter, A_0 is a constant involving certain constants of nature, T is the absolute temperature, φ is the work function of the emitter, e is the electron charge ($= 1.591 \times 10^{-19}$ coulomb), and k is Boltzmann's constant ($= 1.371 \times 10^{-23}$ joule per degree). This equation is based on the assumption that the electrons obey the Fermi-Dirac statistics within the metal. The constant A_0 should equal 120.4 amp. per centimeter per degree from theoretical considerations, but experimental results indicate a wide variation in its measured value, due probably to the difficulty in obtaining a clean surface on the metal measured.

However, the most important factor is the exponential. Small variations in φ and T will cause large variations in i, and so we conclude that the material best suited for a thermionic emitter is one that has the lowest work function and can withstand the highest temperature without deteriorating. Hence, possession of one of these characteristics alone is not sufficient to warrant the use of a material for emission purposes.

5. Tungsten Cathodes.—Tungsten is a metal well suited for thermionic emission. It is exceedingly refractory and has the highest melting point of all metals (3655°K.). It can safely be operated at 2400°K. for wires of small diameter and at higher temperatures for wires of larger diameter. At these tempera-

tures, the exponential $-e/kT$ becomes sufficiently small for practical use of tungsten as a cathode even though the work function is 4.54 volts, a rather high value. Tungsten is particularly well suited for high-power transmitting tubes, where the cathode has to be especially rugged to withstand all the mechanical and chemical requirements, particularly bombardment by high-velocity positive ions in high-voltage operation.

6. Thoriated-tungsten Cathodes.—The heating power required for tungsten is rather high, and so other thermionic emitters have been sought. The most practical are the thoriated-tungsten and the oxide-coated cathodes, which we shall discuss in the order named. The potential barrier of a metal depends markedly upon the nature of its surface. Indeed, it is difficult in practice to obtain a pure metal surface, owing to absorbed gases and other impurities. Tungsten, for example, when contaminated with oxygen experiences an increase in its work function and hence a decrease in its emission at a given temperature. A monatomic film of thorium, on the other hand, reduces the potential barrier and hence the work function to a much lower value, so that copious emission is obtained at much lower temperatures, and hence filament heating power.[1] The work function is about 2.63 volts, as compared with 4.65 volts for pure tungsten.

It is to be noted that a pure thorium filament would not be as satisfactory as the above composite structure; for it could not be operated at the requisite temperature for satisfactory emission without undue evaporation, and also the work function is higher than that for the composite structure. The reason the monatomic film can be heated to a higher temperature than a solid thorium wire without evaporating from the tungsten base is that the force of adhesion between tungsten and thorium is greater than the force of cohesion between two thorium atoms. Indeed, this accounts for the fact that the thorium layer is monatomic: any double layer would quickly be destroyed by the outer layer of thorium evaporating from the thorium layer underneath it.

In manufacture, the tungsten is mixed with 1 to 2 per cent of thoria (thorium oxide) and processed into a filament. The tube is degassed, and then activation is started. The filament is flashed for one or two minutes at about 2800°K. Some of the thoria is reduced to pure thorium by the tungsten, the resulting tungsten oxide being vaporized and deposited upon the cooler

parts of the tube. (This takes place in spite of the greater heat of formation of thorium oxide as compared with tungsten oxide because of the volatility of thorium and diffusion to the surface, *i.e.*, the mass-action law.)

The filament then is operated at about 2100°K. The thorium diffuses rapidly along the grain boundaries of the tungsten crystals to the surface and then migrates over the surface to cover a large fraction of it. At this temperature, however, the evaporation of thorium from the surface is low.

The temperature is then reduced to the operating value of between 1800 and 2000°K., and at this temperature the diffusion is reduced markedly, although it is still greater than the rate of evaporation of a monatomic film. However, if this tends to produce a double layer of thorium, the outer layer will quickly evaporate, so that essentially emission is from a monatomic film

Such a film is sensitive to bombardment by high-velocity positive ions, which will tend to strip it off. It has been found, however, that if the thoriated-tungsten filament is heated to about 1600°K. in a hydrocarbon vapor a shell of tungsten carbide is formed. This is brittle, but the inner core of tungsten maintains the mechanical strength of the filament. Such carbonized filaments show a reduction to one-sixth in evaporation of the thorium layer at 2200°K. so that they can be operated at a higher temperature than the thoriated-tungsten filament. This, in turn, increases the rate of diffusion of thorium to the surface, thus enabling the damage due to positive-ion bombardment and oxidation to be more quickly repaired. Even so, however, this type of emitter is limited to tubes operating at plate potentials of 4,000 volts and lower.

After some time, thoriated-tungsten filaments lose their emissive powers owing to the consumption of the thorium produced by the initial flashing. Since there is still a reserve of thoria, the activation process can be repeated until all the reserve is consumed.

7. Oxide-coated Cathodes.—The second type of composite cathode is the oxide-coated structure. This consists of a metal base such as nickel with a few per cent of cobalt, silicon, or Konel metal (alloy of nickel, cobalt, iron, and titanium), coated with the oxides of barium and strontium. The latter are applied to a filament type of cathode by dragging the wire through suspen-

sions of the carbonates in water and baking each successive thin layer or, in the case of indirectly heated cathode sleeves, by spraying them with a mixture of the carbonates suspended in a solution of nitrocellulose, which subsequently acts as a binder. The coating is then reduced to the oxides of the metal by heating them to a temperature of 1400°K. in a vacuum. The carbon monoxide that is evolved is continuously pumped off.

The cathode must now be activated. This is accomplished by heating it to a temperature between 1000 and 1500°K. for several minutes and then at a lower temperature, while plate potential is applied, for a longer period of time. Emission now begins to take place at an increasing rate until it reaches its normal value, which, at 1000°K., is equivalent to that of a tungsten filament at 2300°K.

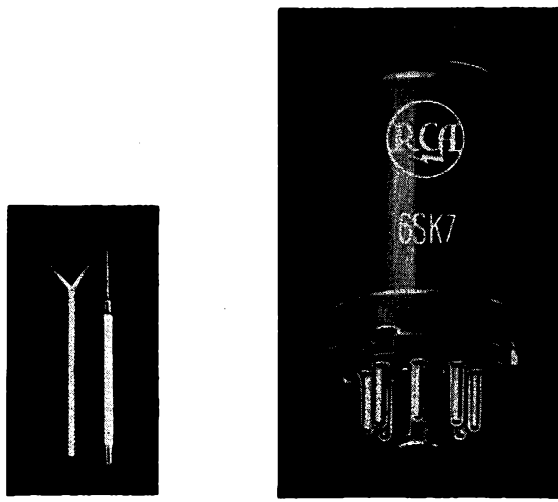
The oxide-coated cathode requires the lowest temperature of the three types discussed for copious emission and hence is employed wherever possible. Since the coating is sensitive to positive-ion bombardment and since a tube employing this type of cathode cannot be evacuated so completely as those employing the other two types, its use is limited to the lower voltage receiving type of tube. It has, however, a life of several thousand hours and, owing to its low operating temperature, is ideally suited for indirectly heated cathode purposes.

The mechanism of emission appears to be as follows: During the activation process, free barium metal is produced in the coating. This diffuses in part to the surface to form a monatomic film partly covering the coating, the remainder being diffused throughout the coating. This free barium is produced by reduction and electrolysis of the oxides as a result of the elevated temperature and anode potential employed during the activation process. The barium is thus absorbed on the core metal and in the coating and is in the form both of atoms and ions, since it loses its outer electrons with relative ease.

The presence of absorbed ions (adions) near the core metal reduces the work function of the latter so that electrons can escape from within to the coating. They then pass through and relatively freely out of the latter into the space surrounding the cathode. The conductivity of the coating is predominantly electronic in the presence of free barium but is also partly ionic. The latter form of conductivity tends to maintain the free barium

content of the coating in spite of evaporation of the metal from the surface.

8. Indirectly Heated Cathodes.—The oxide-coated emitter is the only one suitable for indirectly heated cathodes, owing to its low operating temperature. This type of cathode consists of a heater wire coated with a layer about 0.5 mm. thick of the oxides of aluminum and beryllium, the combination then being placed in a nickel or Konel-metal sleeve on which has been placed an oxide coating. The heater operates at about 1000°K. and maintains the cathode at about 850°C., which is sufficient for copious emis-



a *b*

FIG. 2.—Oxide-coated cathode and complete tube. (*Courtesy of RCA.*)

sion from the oxide coating. Such indirectly heated cathodes are equipotential surfaces. Between such a cathode and the plate can be impressed a direct voltage, with resultant d.c. flow unless the latter is purposely modified by a desired a.c. potential applied between a third electrode (called a *grid*) and the cathode. This action will be explained later. At the same time, 60-cycle alternating current can be applied to the heater, thus eliminating storage batteries or other d.c. sources for this purpose, and yet not having an unwanted component of this frequency appear in the output circuit as hum. This is of particular importance in audio amplifiers and radio receivers. Successful a.c. operation of such devices dates from the time indirectly heated cathode tubes were

introduced to the industry. Since, however, oxide-coated cathodes are not suitable for large transmitter tubes, this type of cathode cannot, unfortunately, be employed in such tubes; instead, filament type cathodes are still used.

In Fig. 2*a* may be seen an oxide-coated cathode, and in Fig. 2*b* a complete tube of this type.

9. The Tube as a Nonlinear Parameter.—If another electrode, called a *plate*, is introduced into the tube (now called a *diode*) and the plate is made positive with respect to the cathode, a current will flow if the cathode is heated to produce thermionic emission. If the plate is made negative with respect to the cathode, no current will flow, for the applied potential is assumed less than that required to produce field emission from the plate. By the same token, no current will flow if the plate is made positive with respect to the cathode but the latter is not heated.

The unilateral conductivity of the tube when the cathode is emitting makes it useful as a rectifier of alternating voltages and also indicates that such a device is a nonlinear parameter. This is so because in a linear parameter the current is directly proportional to the voltage regardless of whether the latter is positive or negative, whereas, in the case of the tube, the current is zero when the (plate) voltage is negative. Moreover, it will be found that—except at ultra-high frequencies—there is no lag between the current flow and the applied voltage, so that we can state more specifically that the vacuum tube is a nonlinear resistance parameter. It is for this reason that Chap. I deals with nonlinear circuits, particularly nonlinear resistances: we have to deal with these when we deal with vacuum tubes.

10. Space Charge.—At first thought it would appear that a small voltage would be sufficient to draw all the emitted electrons through the obstacle-free space or vacuum over to the plate, since the thermal energy has done that which the voltage was unable to do, *viz.*, carry the electrons through the surface of the cathode (emission). A moment's reflection, however, will indicate that this is not so, that there are obstructions in the space in the form of mutual repulsions between the electrons themselves and that these hinder the plate potential in its effort to transport the electrons. This repelling effect is known as *space charge*. In the conductor it is neutralized by the positive space-charge effect of the metallic ions in the crystal lattice structure, and the only

opposition to the movement of the electrons is that of ohmic resistance—possibly the result of collisions of the electrons with the ions. But when the electrons are "out in the open," remote from the ions, their negative space-charge effect becomes evident, and it will be found that a low plate voltage causes but a small current to flow, a higher plate voltage causes a greater current to flow, etc., until a voltage is reached which causes all the electrons emitted to flow. Higher voltages than this can cause no further increase in current, and the latter maximum value is known as the *temperature-saturation value* of current, whereas the former values are known as *voltage-saturation values.*

11. Child's Law.—In 1911, Child[2] gave the first analysis of the relationship between current and voltage in a vacuum tube. An account of this analysis will be given here because it is one of the few examples of nonlinearity in which the internal mechanism causing this is capable of quantitative exposition.

Child idealized the configuration by assuming that the cathode and plate were infinite parallel planes separated by a distance d. He further assumed that the emission was infinite and therefore far in excess of any required for the current flow to be produced; *i.e.*, the plate voltage would be less than that producing saturation current. He further assumed that the electrons were emitted with no initial velocities. The effect of initial velocities is to project the electrons out into the interelectrode space in spite of the space charge; but since the emission velocities are low (equivalent to about 0.6 volt, average, for an oxide-coated cathode), the error in neglecting them is small. Indeed, in spite of the idealization of the problem, the results are in good agreement with those experimentally obtained for actual tubes.

Consider a cloud of electrons in the interelectrode space. Those nearer the plate are repelled toward it by those nearer the cathode; the latter are repelled by the former toward the cathode. Those on the cathode have the greatest number of electrons in front of them and thus experience the greatest force opposing their moving toward the plate. At the same time, these must move toward the plate if others nearer the anode are to deposit on it, for Kirchhoff's law as to the continuity of current flow must be satisfied here as in the case of any other type of electric current flow. Hence, the current flow will be of such magnitude that the resultant negative space charge of those electrons already in

transit will be balanced by the forward pull of the anode upon the electrons just about to leave the cathode, *i.e.*, upon those hardest to move. We shall now formulate the relationship between this equilibrium current and the plate voltage.

In electromagnetic theory for steady current flow we have Poisson's equation

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

where V is the potential at any point in the interelectrode space, x is the distance of the point from the cathode, and ρ is the charge density at the point in statcoulombs per cubic centimeter. Owing to the assumption of infinite parallel-plane electrodes, the electrostatic lines of force will all be perpendicular to the electrodes, and hence the divergence of the electrostatic lines of forces in the y and z directions (indicated by d^2V/dy^2 and d^2V/dz^2, respectively) will be zero.

If there is no charge density at the point in question, ($\rho = 0$), then Eq. (2) becomes Laplace's equation, *viz.*

$$\frac{d^2V}{dx^2} = 0 \qquad (3)$$

Both Poisson's and Laplace's equations are based upon Gauss's theorem or, more fundamentally, upon Coulomb's law, which is an example of a central force. The proof of these theorems can be found in any standard text on electromagnetic theory.[3]

In our example, we assume that there is a charge ρ throughout the interelectrode space, so that Poisson's equation is to be used. We note that ρ is not a constant, but a function of x. This relationship is given by the equation of continuity, *i.e.*, the current is everywhere the same and furthermore is equal to the charge times its velocity. Hence, we can write

$$I = \rho v_x \qquad (4)$$

where I is the current per unit area of the electrode surfaces and v_x is the velocity of the electrons at the point distant x from the cathode.

It is also true, from the definition of potential as the work done on a unit charge in transporting it from a point of zero potential to the point whose potential is V, that the kinetic energy imparted

to an electron coming from the cathode (assumed at zero potential) to the point in question is

$$\tfrac{1}{2}mv_x^2 = Ve \tag{5}$$

where m is the mass of the electron and e is its charge.

We can eliminate ρ and v_x from Eqs. (2), (4), and (5) and obtain

$$\frac{d^2V}{dx^2} = 2\pi I\sqrt{\frac{2m}{eV}} \tag{6}$$

The solution of this equation can be effected by multiplying through by $2(dV/dx)$ and integrating. Thus

$$2\frac{dV}{dx}\frac{d^2V}{dx^2} = \frac{d}{dx}\left(\frac{dV}{dx}\right)^2 = 4\pi I\sqrt{\frac{2m}{e}}\,V^{-\frac{1}{2}}\frac{dV}{dx} \tag{7}$$

and

$$\left(\frac{dV}{dx}\right)^2 = 8\pi I\sqrt{\frac{2m}{e}}\,V^{\frac{1}{2}} + C \tag{8}$$

We can now insert the boundary conditions:

1. For $x = 0$, $-dV/dx = 0$, since the latter represents the potential gradient, or force on a unit charge, and this is zero at the cathode for the equilibrium current, as explained previously.

2. For $x = 0$, $V = 0$, since by hypothesis we assumed the cathode to be at zero potential.

From these we find C, the integrating constant by Eq. (8), to be zero, so that

$$\frac{dV}{dx} = \sqrt{8\pi I}\left(\frac{2m}{e}\right)^{\frac{1}{4}}V^{\frac{1}{4}} \tag{9}$$

This first-order differential equation can readily be solved by separation of the variables and yields

$$V^{\frac{3}{2}} = \frac{9}{16}\,8\pi I\sqrt{\frac{2m}{e}}\,x^2$$

or

$$I = \frac{V^{\frac{3}{2}}}{x^2}\,\frac{1}{9\pi}\sqrt{\frac{2e}{m}} \tag{10}$$

which, as indicated in Eq. (2), is in electrostatic units. This can be converted into practical units, and the constants calculated. If we further substitute d for x, we obtain

$$I = 2.336 \times 10^{-6} \frac{V_p^{3/2}}{d^2} \text{ amp. per cm.} \tag{11}$$

where V_p is the potential between the plate and cathode. Equation (11) is known as *Child's law.*

12. Discussion of Child's Law.—An examination of Eq. (11) reveals the following:

1. For a fixed plate voltage V_p, the voltage saturation current, as it is called, varies inversely as d^2. This indicates the effect of varying the spacing of the electrodes of the diode.

2. Since I, the voltage saturation current, is the same for all values of x, the potential at any point at a distance x from the cathode is

$$V = 3\sqrt[3]{\frac{3\pi^2 m I^2}{2}} \cdot x^{4/3} = (\text{const.}) \cdot x^{4/3} \tag{12}$$

This indicates the variation of potential with distance due to the presence of the negative electrons in the interelectrode region. We note that, if there were no electrons there, V would be a linear function of x, as in the case of the parallel-plate condenser.

3. Other facts that can easily be derived are that the potential gradient

$$E = -\frac{dv}{dx} = -(\text{const.}) \cdot x^{1/3} \tag{13}$$

that the velocity

$$v_x = (\text{const.}) \cdot x^{2/3} \tag{14}$$

and that

$$\rho = (\text{const.}) \cdot x^{-2/3} \tag{15}$$

For $x = 0$, $\rho = \infty$, which means that the emission must be assumed to be infinite, in order to represent a finite current at the cathode when the electron velocities there are assumed zero.

However, the most interesting fact is that presented by Eq. (11) itself, *viz.*, that the current density, hence the total current, is proportional to the $\frac{3}{2}$ rather than the first power of the plate voltage. This indicates that even for positive voltages the diode is nonlinear. Of course, as indicated in Chap. I, the entire terminal characteristic for positive and negative plate voltages is nonlinear, but the fact that even the positive range is nonlinear is of interest in connection with the following: Suppose that the

plate voltage was composed of a d.c. component V_1 and an a.c. component V_2, whose peak value was less than V_1, so that $V_p = V_1 + V_2$ was always positive.* Then, if the terminal characteristic for positive V_p were linear, V_2 would produce an a.c. component of current of the same wave shape as V_2, whereas, if the $\frac{3}{2}$ power holds, the current will be a distorted copy of V_2.

Physically, the reason for the $\frac{3}{2}$ power is that, if the voltage is increased, the current is increased because of the greater number of electrons moved and because they are moving with a greater velocity than before. Since, for equilibrium conditions, the number of electrons en route to the plate must increase to a point sufficient to reduce the pull of the plate on the electrons at the cathode to zero and they are moving with a higher velocity, the current must increase more than in proportion to V_p, that is, as $V_p^{3/2}$.

Langmuir derived Eq. (11) independently, as well as for the case where the plate is an infinite cylinder and the cathode is an infinite equipotential emitter situated on the axis of the plate. In this case, I varies as $V_p^{3/2}$ too, but only inversely as the radius of the cylinder if it is large compared with the radius of the cathode. He has further shown that the $\frac{3}{2}$-power law holds theoretically for all geometric configurations of the electrodes and that the difference is only in the constants of the equations.

13. Departures from the Theory.—In actual tubes the departure from the $\frac{3}{2}$-power law may be appreciable. One of the reasons for this is that the electrodes are not infinite in extent, so that the fringing of the electrostatic field at the finite boundaries modifies the effects. Another is the effect upon the electrostatic field of supporting wires, and a third is the fact that the electrons are emitted from the cathode with initial velocities. The effect of the latter is to cause an excess of electrons in front of the cathode, so that the field at the cathode is negative and electrons at the cathode are repelled into it in spite of the pull of the plate. However, the initial velocities of most of the electrons are sufficient to overcome this negative gradient and indeed cause it to be established in the first place.

* Note that voltages are represented in this chapter by V, in order that E may be reserved for the potential gradient. In other chapters, E is used for voltage to conform with the nomenclature suggested by the Institute of Radio Engineers.

A fourth effect is that of the potential drop along the length of a filament type cathode (direct emitter), which affects the potential distribution between any part of the cathode and the plate. For an indirectly heated cathode this effect is absent, however.

These departures warrant, in many cases, the use of a power series in integral powers of V for the representation of the functional relationship between I and V. Indeed, the chief use of the $\frac{3}{2}$-power law is in correcting the value of the d.c. component of the plate current if the direct plate voltage is other than that normally specified. Another example is that of the correction of the d.c. component of the plate current in a pentode tube if the screen-grid voltage is other than that normally specified by the manufacturer.

A final departure from the theory is obviously that where V_p is so high that temperature-saturation current is caused to flow. The current then levels off, as shown in Fig. 3. Such leveling off is not pronounced in a composite type emitter, such as an oxide-coated cathode, owing, possibly, to the drawing out of further electrons from the crevices in the uneven cathode surface by higher values of V_p.

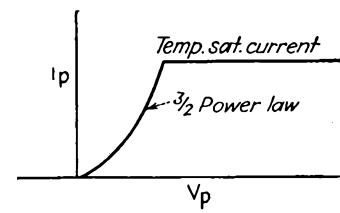


FIG. 3.—Temperature-saturation characteristic.

We note from this, however, that, if the cathode emits more electrons than will normally be required in operation, the plate current will be determined by voltage-saturation conditions, Child's law will hold, at least approximately, and all types of cathodes having adequate emission will exhibit the same sort of terminal characteristic. Hence, in practice, the cathode is designed to emit such a large number of electrons that, even if the emission falls off with time, the terminal characteristic over the normal range of plate voltages will remain unchanged.

14. The Triode Tube.—We have seen that, owing to the presence of a space charge, the flow of the electrons is limited to a value given by Child's law for a given plate voltage. This value is much less than that which would flow, for example, in a copper conductor of the same cross section as the cathode or plate; *i.e.*, the d.c. resistance of the diode is much greater than that of the copper conductor. The resistance of the former is due to the

opposition to the flow of electrons by their neighbors ahead of them; the resistance of the copper conductor is due probably to the collisions of the electron gas with the ions of the crystal lattice structure and this resistance is relatively small in value. The space-charge effect of the electrons in the copper is neutralized by that of the positive ions.

On the other hand, it is practically impossible to modify the space-charge conditions in a conductor by the insertion of another control electrode, whereas, in the diode, such insertion has a profound effect upon the space-charge conditions and hence current flow. In the latter case, the electrons, as stated before, are "in the open" and hence at the command of the control electrode, whereas, in the conductor, the control effect of the many positive ions is so jumbled as to be negligible from a statistical viewpoint for the majority of the electrons of the outer orbits, and the effect of a control electrode is rendered negligible by the contradictory and hence balancing effects of the many positive ions.

The use of a control electrode, in the form of a mesh, or grid, occurred to De Forest[5] in 1907 and shortly after to von Baeyer[6] in Germany. The resulting three-electrode tube is now known as a *triode* and exhibits remarkable control characteristics that form the basis of most of present-day electronic technique.

In 1913 van der Bijl[7] gave what was possibly the first analytical treatment of the action of the grid. It will not be possible in this chapter to discuss the derivation. A more recent work by Vodges and Elder[8] treats this matter very elegantly by the method of conformal mapping. For the purposes of this text, a physical interpretation will suffice.

Equation (12) showed that the space current I will be increased if the spacing between the electrodes, d, is decreased. Hence, if a third electrode is introduced between the plate and cathode, it may be expected to exert a more profound effect upon the space current than the plate. In addition, however, the grid shields the cathode from the electrostatic field of the plate and thus further reduces the effect of the plate upon the space current. As a result, the grid plays a more dominant role than the plate in determining the space current, even that portion of it which passes through the interstices of the grid en route to the plate and called the *plate current.* The relative effectiveness of the two

electrodes is measured by a quantity called the amplification factor, which we shall now discuss.

15. The Amplification Factor.—The grid may be made either positive or negative with respect to the cathode. If positive, it will aid the plate voltage in producing current flow but will divert electrons to itself; if negative, it will oppose the plate voltage but draw no current and hence cause no expenditure of energy in the electrical control source connected between it and the cathode. Hence, it is usually operated at a negative potential, called a *negative bias*. If this bias is too great, it will prevent the plate from drawing electrons away from the cathode, *i.e.*, no current will flow. The tube is now said to be *at cutoff*, and the bias producing this condition is called the *cutoff bias*. In normal amplifier operation, the tube is operated above the cutoff point.

Suppose, in a particular tube, it is found that if the plate voltage is raised 10 volts but the grid is made negative by one more volt, as measured from an initial set of values, the plate current remains unchanged. Evidently the increased negative grid bias has just balanced the increased plate potential. The ratio of the two opposing changes is 10, and this is called the *amplification factor* of the tube (denoted by the symbol μ).

It may then be found that, if the plate potential is increased by 20 volts, the grid must be biased back by 2.2 volts. The amplification factor is now only 9.09. This indicates that the amplification factor is somewhat variable. Hence, recourse is had to the calculus, and the factor is defined as

$$\mu = \left.\frac{dV_p}{dV_g}\right]_{i=\text{const.}} \tag{16}$$

where dV_p and dV_g are differential changes in the plate and grid voltages, respectively. Since they are of opposite sign, μ is inherently a negative number, although, if this is understood in subsequent work, the positive sign may be employed and proper interpretation given to the results obtained.

16. Factors Affecting the Amplification Factor of a Triode.—The magnitude of μ depends upon several factors, such as the geometric shapes of the electrodes, the spacing between the grid and plate, and the fineness of mesh of the grid structure. A typical triode construction is shown in Fig. 4; it will be noted that the grid is in the form of a helical coil of wire. The amplification factor is

increased if the pitch of the helix is reduced and the diameter of the grid wire is increased, for the grid can then shield the cathode from the electrostatic field of the plate more completely and thus render the plate less effective in producing plate current. In practice, triode tubes have amplification factors ranging from about 3 to 100, while tetrode and pentode tubes have values of over 1,000 in many cases, as will be explained in the following chapter.

As indicated in the previous section, the μ of a triode may be variable with the tube voltages—indeed, always is variable to

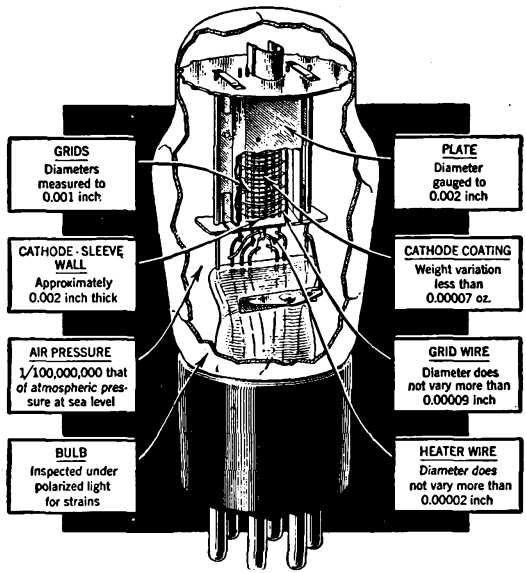


FIG. 4.—Typical triode construction. (*Courtesy of RCA.*)

some extent in an actual tube. This variability may be explained as due to asymmetries in the tube structure. For example, suppose the pitch of the grid helix varies somewhat along the length. Suppose, for definiteness, that the helix is tight for half its length and coarse for the other half. The tube may be regarded as essentially two tubes in parallel, one with a high μ (tight helix) and the other with a low μ (coarse helix). As the negative bias of the actual grid is increased, the plate current of the higher μ portion of the tube will reach cutoff first. From thence to the cutoff of the other portion of the tube, the latter will appear to have a lower μ. Hence the amplification factor

of an ordinary tube having some unavoidable asymmetry in structure due to tolerances in manufacture will always show a decrease in μ as cutoff is approached. The effect, however, is usually small.

In some tubes, called *variable-μ* or *supercontrol tubes*, the effect is purposely made large, in order to afford a gradual cutoff characteristic. In Fig. 5 is shown the relationship between a constant-μ and a variable-μ tube for comparison. In this figure, plate current i_p is plotted against grid voltage V_c for a given plate voltage. It will be noted how protracted the variable-μ cutoff is compared with the constant-μ cutoff.

Variable-μ tubes are employed in radio-frequency (r.f.) and intermediate-frequency (i.f.) amplifiers where automatic variation in amplification is desired to compensate for weak and strong broadcast signals. This compensation is known as *automatic volume control* (a.v.c.). They are also employed in audio amplifiers to vary the amplification according to whether the signal is weak or strong so as to accommodate a large volume range within the signal-handling capacities of the apparatus. This application is known as *compression* or *expansion*, depending upon whether the μ is decreased or increased, respectively, as the signal increases. It is to be noted, however, that variable-μ tubes are practically always of the pentode or multielectrode form, rather than the triode.

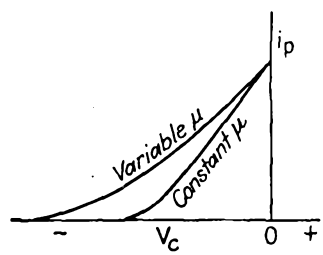


FIG. 5.—Transfer characteristics for constant- and variable-μ tubes.

17. Practical Application.—Since the grid is more effective than the plate in determining the plate current, it is evident that more effect will be obtained by injecting the given signal voltage in the grid circuit than directly in the plate circuit, for a greater variation in plate current will be obtained by the former mode of operation. Thus, suppose the signal voltage V_s is a 100-cycle tone. If the source of this is connected in series with the grid and the bias battery, V_c, as shown in Fig. 6, the plate current will be a pulsating current instead of being a steady d.c. flow. This, in the absence of appreciable distortion, can be resolved into the original d.c. component plus an a.c. component of 100 cycles. The latter is due to V_s.

In flowing through Z_L (Fig. 6), the so-called "load impedance," this a.c. component will develop 100-cycle power in it and also a 100-cycle voltage across it.

If V_s were connected in series in the plate circuit, it would also produce a 100-cycle component in i_p, but the component would be smaller than before, the power expended in Z_L would be less, and the voltage across Z_L would be less too. Moreover, whatever 100-cycle power was expended in Z_L would come directly from the source of V_s, whereas, if V_s is injected into the grid circuit and the grid maintained sufficiently negative, no grid current can flow and no power is drawn from the source of V_s. In this case, the power developed in Z_L comes from the d.c. plate battery, of voltage V_{bb}. The ratio of a.c. output power (in Z_L) to a.c. input power is now theoretically infinite and, in actual practice (owing to some slight losses), exceedingly great.* This ratio is known as the *power amplification factor* and is of great importance when very weak power sources of V_s, such as high-fidelity microphones and photocells, are employed.

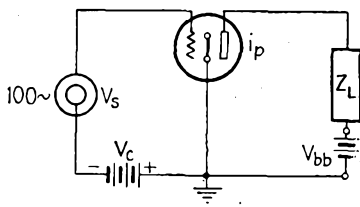


FIG. 6.—Circuit for voltage and power amplification.

Usually, the power output in Z_L is insufficient for the purpose at hand. In that case the output voltage across Z_L can be used to actuate the grid of a second tube, etc., until finally a stage is reached whose output power is sufficient. This final stage is known as a *power output stage;* usually, special tubes adapted for large power outputs are employed here, known as *power tubes.* The previous stages are known as *voltage amplifier stages.* The ratio of the output voltage to the input voltage of such a stage is known as the *stage amplification* or *stage gain.* Since these will be analyzed more fully later, no further discussion will be presented at this point.

18. Other Tube Parameters.—There are two other fundamental tube parameters that will be discussed at this point, the plate resistance R_p and the transconductance G_m. It will have been noted that the triode is operated so that direct plate current

* At ultra-high frequencies, however, the input power may approach the output power in value, owing to transit-time effects.

flows. The action of the grid is to produce a variation in this current, and this variation is called the *a.c. component.*

If we divide the d.c. plate potential V_{bb} by the direct plate current I_b, we have a resistance R_b, called the *d.c. resistance* of the tube. This parameter, however, is seldom employed in analytical methods of solution. On the other hand, the ratio of alternating plate voltage (if alternating voltage be injected into the plate circuit) to the resultant a.c. component of plate current is a more useful ratio analytically and is called the *a.c., variational,* or *incremental plate resistance* (see Chap. I). It is evident that the alternating voltage superimposed on V_{bb} may be regarded as a variation or incremental change in V_{bb}—hence the variety of names for this resistance. The latter is a function of the magnitude of the a.c. component of the plate as well as the grid bias voltage and hence, analogously to the expression for μ, is written as

$$R_p = \frac{dV_p}{di_p}\Bigg]_{V_g = \text{const.}} \tag{17}$$

where dV_p is the differential change in plate voltage and di_p is the resulting differential change in plate current, under the condition that the grid voltage V_g remains constant. This parameter, as will be shown, is a very important one in vacuum-tube theory.

The transconductance G_m is a measure of the effectiveness of the grid voltage in producing a change in plate current. In this case, as in that describing the meaning of the R_p of a tube, it is assumed that there is no load in the plate circuit of the tube: even the plate (B) supply is assumed to have no internal impedance. Therefore, the only opposition to the flow of the a.c. component of plate current is the tube itself, *i.e.*, its internal space-charge effects.

Referring to the G_m once again, we find that, for example, a 1-volt increase in grid voltage allows the plate current of a tube to rise 2 ma. above its previous value, under the condition that the plate voltage is maintained constant. The G_m is then 0.002 amp. divided by 1 volt, or 0.002 amp. per volt, which is 0.002 mho, or 2,000 micromhos. This is analogous to a.c. plate conductance (reciprocal of R_p), except that the voltage is applied in the grid circuit and causes a transfer effect in the plate circuit—

hence the name "transconductance." It is also often called "mutual conductance."

The G_m is also a function of the grid voltage and plate voltage and if defined by the differential expression

$$G_m = \frac{di_p}{dV_g}\Bigg]_{V_p = \text{const.}}^{*} \tag{18}$$

where di_p is the differential change in plate current due to the differential change in grid voltage, V_g, while the plate voltage V_p is maintained constant. This will be found to be a useful parameter too.

Now the μ of the tube was defined as the relative effectiveness of the grid compared with the plate in determining the plate current; that is, dV_g in Eq. (18) is equivalent to $1/\mu$ times a certain increment of plate voltage, dV_p. Hence, Eq. (18) can be written as

$$G_m = \frac{di_p}{dV_p/\mu} = \frac{\mu}{R_p} \tag{19}$$

since $di_p/dV_p = 1/R_p$. The derived form for G_m given by Eq. (19) is particularly useful in circuit analysis.

19. Vacuum-tube Operation from the Physical Viewpoint.—If the grid is viewed as varying the internal space charge of the tube and, in this way, the current flow, then the tube may be regarded as an adjustable resistance of magnitude determined by the grid voltage. The resistance referred to is its d.c. resistance, and the tube is thus a kind of rheostat controlled by a (grid) voltage rather than by a rheostat arm manually operated. If the grid voltage varies with time, then the d.c. resistance of the tube will vary with time, and the tube is then called a *time-variable parameter.* This type of parameter was not discussed in Chap. I. The differential equations describing a circuit containing linear and time-variable parameters involve di/dt multiplied by constant resistances and also by resistances that are functions of the same variable as the impressed voltages, *viz.*, time. Such equations are, in general, easier to solve than nonlinear differen-

* Strictly speaking, these are partial derivatives, since i_p is a function of V_g and V_p. The partial-derivative representation will be used in Chap. V, where both voltages are allowed to vary simultaneously as in normal operation.

tial equations, which involve di/dt multiplied by resistances that are a function of the current, i.

If a tube has a characteristic such that, if the grid voltage be fixed at some value, the a.c. resistance is independent of the current drawn by the plate voltage, then the tube is said to be a *linear tube.* (Obviously, this holds only for positive plate voltages great enough to counterbalance the negative grid voltage and prevent cutoff.) Since the current is a function of both the grid and plate voltages, the plot of such a characteristic (see Chap. III) is a surface in three-dimensional space. For a linear tube this surface is a cylinder of some sort.

If, in addition, the μ of the tube is constant, then the cylindrical surface becomes a plane in space. In such a tube, if the range of operation is such that plate current flows at all times, the output voltage is an undistorted copy of the input (grid) voltage. Such a tube is highly desirable for amplification purposes, but it can be only approximated by actual tubes. This is evident from the following: For a linear tube, if the grid voltage V_g be kept constant, i_p must be in proportion to $V_p - \mu V_g$, since the a.c. resistance R_p is constant and $V_p - \mu V_g$ is the portion of the plate voltage that can cause plate current to flow. (Note that μV_g is assumed negative, so that it cancels part of V_p.) The relationship is represented in Fig. 7 for two values of V_g (solid lines), *viz.*, V_{g1} and V_{g2}. Note that, when V_p equals μV_{g1}, cutoff is reached for the curve to the left labeled V_{g1} and, when $V_p = \mu V_{g2}$, cutoff is reached for the curve labeled V_{g2}. Also, note that, for point P, the d.c. resistance is the cotangent of θ_1, while the a.c. resistance, or R_p, is equal to cot θ_2. These matters will be discussed more fully in the following chapter.

We have shown for the diode that, according to Child's law, i_p varies as V_p. While no rigorous derivation has been worked out for the triode tube, a somewhat similar relation holds approximately for the triode, *viz.*,

$$i_p = A(V_p + \mu V_g)^{3/2} \tag{20}$$

For actual tubes the relation between i_p and V_p may depart somewhat from Eq. (20), but the exponent is in practically all cases other than unity. Hence, the actual triode curves are other than straight lines and may be of the shapes suggested by the dotted lines Fig. 7. Such curvature is further increased, in general, by

variability of the amplification factor, particularly in the neighborhood of cutoff.

The reader may be puzzled as to why the above discussion is concerned with a variation in the plate voltage, since it was explained previously that the given variation in voltage known as the signal voltage was preferably introduced into the grid circuit. The answer to this is that, for useful operation, a load impedance Z_L (see Fig. 6) is inserted in the plate circuit, so that the variations in plate current, produced by the signal voltage acting on the grid, produce variations in the voltage across Z_L which are a copy of the grid signal voltage. Since the impressed

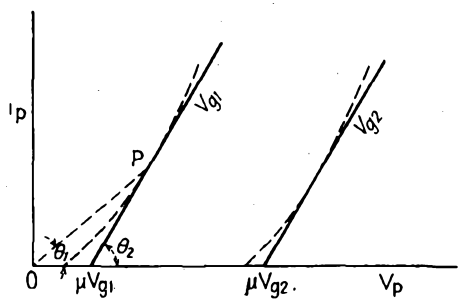


FIG. 7.—Linear and actual tube-plate characteristics.

B voltage V_{bb} is direct, then if there is an alternating voltage V_L, across Z_L, there must be a voltage $-V_L$ between the plate and cathode, as well as a direct voltage, in order that the sum of the voltage drops equals V_{bb}. This means that, when Z_L is present, V_p is not constant but fluctuates, and hence we are concerned with the effect upon i_p of these fluctuations, as well as of those in the grid circuit.

If the current is proportional to $(V_p + \mu V_g)^{3/2}$ instead of the first power, the current will be a distorted copy of V_g and will produce a distorted voltage across Z_L, and also one of opposite phase, as part of V_p, which will react to make the current somewhat differently distorted, etc., in short, the exact form of i_p (and V_L) for equilibrium conditions may be quite different from what it is when $Z_L = 0$ and $V_p = V_{bb}$, a constant direct voltage. Indeed, this is the basic problem of nonlinear circuits (see Chap. I): given a known voltage—here V_{bb}—in series with an impedance Z_L—which may be linear—and a nonlinear impedance—that of

the tube—to find the current flow through this circuit and the voltages set up, particularly the output voltage V_L.

If the signal voltage is sufficiently small, then the variations in i_p and V_p will be correspondingly small, and the relation between i_p and the grid and plate voltages may then be represented fairly accurately by

$$i_p = A(V_p + \mu V_g) \tag{21}$$

For a linear tube, this is an exact representation. Where Eq. (21) is adequate, a simple equivalent circuit, known as the *equivalent plate circuit*, may be employed, and the analysis of the tube behavior greatly simplified.

20. The Equivalent-plate-circuit Theorem.—Let the plate current i_p be composed of the normal d.c. component I_b, due to the B-supply voltage V_{bb} and the grid bias voltage V_c, and an a.c. component I_p, due to a signal voltage V_s injected into the grid circuit. Then, according to Eq. (21),

$$i_p = I_b + I_p = A[(V_{bb} - I_bZ_L - I_pZ_L) + \mu(V_s + V_c)] \tag{22}$$

Note that Z_L may be a different impedance to the d.c. component I_b from that to the a.c. component I_p. In the absence of the signal voltage V_s, we have

$$i_p = I_b = A(V_{bb} + \mu V_c) \tag{23}$$

Subtracting Eq. (23) from Eq. (22), we obtain

$$I_p = -AI_pZ_L + A\mu V_s$$

$$I_p = \mu V_s \frac{A}{1 + AZ_L} \tag{24}$$

It will now be shown that $A = 1/R_p$. If the grid voltage be kept constant, V_s must equal zero. Under these conditions, from Eq. (21)

$$\left.\frac{di_p}{dV_p}\right]_{(V_g = V_c = \text{const.})} = A = \frac{1}{R_p} \tag{25}$$

from the definition of R_p given by Eq. (17).

Substituting this value of A in Eq. (24), we finally obtain

$$I_p = \frac{\mu V_s}{R_p + Z_L} \tag{26}$$

This is the equivalent-plate-circuit theorem. It states that, as far as the a.c. component of the plate current is concerned, it is

as if the tube with its variable d.c. resistance were replaced by an apparent source whose internal impedance is R_p and whose generated voltage is μV_s. The circuit equivalent to that of the plate is shown in Fig. 8, and it is evident by an application of Ohm's law to this circuit that Eq. (26) will be obtained.

This is one of the most frequently used and hence most important equations for vacuum tubes. The reason for this is that the vacuum tube, as mentioned previously, is a time-variable parameter and gives rise to circuit equations difficult to solve. The equivalent circuit is an ordinary linear type and easy to solve and hence is preferable to the actual in the solution of vacuum-tube problems. Accordingly, it is employed wherever possible. It cannot be employed, however, to evaluate the losses in the tube itself, which produce heating of the plate and are hence denoted as "plate dissipation." But, for a linear tube at least, it gives the correct values of a.c. component of current and voltages in the external load impedance, and these are the quantities in which we are usually most interested.

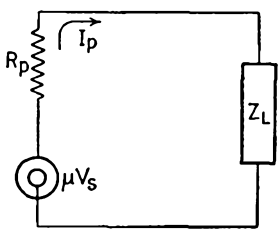


FIG. 8.—Equivalent plate circuit for a.c. component.

In passing we note that the above theorem depends upon a more fundamental one, the *compensation theorem*, which states that a voltage drop due to some impedance may be replaced (compensated for) by a generator generating a voltage equal to this voltage drop. Here we are compensating for the variable resistive voltage drop in the tube by an equivalent generated voltage μV_s and a fixed resistance R_p.

21. Equivalent Constant-current Source.—A variation of the equivalent-plate-circuit theorem follows at once from ordinary linear-network theory. Thus, let us first evaluate the voltage V_L across Z_L. This is evidently

$$V_L = I_p Z_L = \mu e_s \frac{Z_L}{R_p + Z_L} \tag{27}$$

Now multiply numerator and denominator by R_p, and obtain

$$V_L = \frac{\mu}{R_p} e_s \frac{R_p Z_L}{R_p + Z_L} = G_m e_s \frac{R_p Z_L}{R_p + Z_L} \tag{28}$$

The quantity $G_m e_s$ has the dimensions of current. This current may be considered as being produced by a constant-current generator. A constant-current generator is a source that feeds an unvarying magnitude of current into any finite value of load impedance connected to it. One physical interpretation of such a source is that it has infinite internal impedance and infinite generated voltage, of such order that their ratio is the finite value of current cited above. Evidently, such a generator will send the same value of current through any finite impedance connected in series with it.

The constant current $G_m e_s$ flows through the internal resistance R_p of the actual source and the load impedance in parallel and, in doing so, sets up the same voltage drop V_L as is set up across Z_L, when regarded in series with the internal resistance R_p and the finite generated voltage μV_s.

As stated before, these circuits are equivalent for any linear circuit, and not because the one under discussion happens to involve a vacuum tube. Which equivalent circuit to use, the constant voltage or the constant current, is a question of convenience. It has been regarded by some that the constant-voltage type is preferable if the source resistance is low, as in the case of the average triode tube (2,000 to 10,000 ohms), and the constant-current type is preferable for high-impedance sources, such as pentode tubes (one megohm or thereabouts). However, either type of representation can be used for any type of circuit, and the constant-current source has been found to be of great convenience in the analysis of many vacuum-tube circuits. We shall illustrate this point with an example.

A well-known type of video amplifier circuit employed in television as a voltage amplifier stage is shown in Fig. 9 and is known as a *series-peaking circuit.* Here C_p represents the unavoidable capacitance existing between the plate and ground of the first (left-hand) tube, and C_i is the likewise unavoidable capacitance associated with the input-circuit of the second (right-hand) tube. While these capacitances are mainly within the tube (although they must also include the stray wiring capacitances of the external load parameters to ground), it is convenient to consider them external to the tube and as part of the external load impedance. The latter consists principally of R_L, a resistance, and L, an inductance. The parameter C_g and R_g have a sufficiently high

susceptance and resistance, respectively, to be considered negligible in effect at the frequencies under consideration—one million cycles and higher. Finally, the input signal voltage is V_1, and the output amplified signal voltage, applied to the next tube, is V_2.

Let the driving-point impedance looking into terminals 1-1 be denoted by Z_D, where this includes C_p, and the transfer constant, which is the ratio of voltage across terminals 2-2 to that across terminals 1-1, be denoted by T. Then, if the first tube be a pentode tube, its R_p may be regarded as so high compared with Z_D as to be a negligible shunt to it. In that case, if we apply Eq. (28), we obtain

$$V_2 = V_1 G_m Z_D T \tag{29}$$

since the voltage across terminals 1-1 is $V_1 G_m Z_D$ and hence that across 2-2 is $V_1 G_m Z_D T$, as above. The gain, α, of this stage is defined as

$$\alpha = \frac{V_2}{V_1} = G_m Z_D T \tag{30}$$

The problem of finding the gain is thus that of finding the values of Z_D and T as functions of frequency—a straightforward problem in linear theory. The only manner in which the tube enters is through its G_m.

Suppose optimum values of R_L and L are thus found in terms of C_p, C, and the frequency f, which will make this stage have a satisfactory flat response over a certain portion of the frequency spectrum, say, up to 5 megacycles. Now suppose it is desired to find the optimum values for a low R_p triode tube. Is it necessary to start anew, since now the R_p is an appreciable shunt across Z_D? Not at all. We can represent the triode tube by its equivalent pentode (constant-current) form. Here we merely place R_p across terminals 1-1 in Fig. 9. Now, if R_p is greater than the previously determined value of R_L in the case of the pentode tube, we can use instead of R_L a higher resistance R_L', of such value that R_L' and R_p in parallel equal the previously determined value R_L. The triode circuit will then function in identical manner with the pentode tube. Since the latter permitted the gain to be determined by simple analysis of Z_D and T, it was natural to derive the gain for this type of tube first.

Of course, this procedure is possible only if the external load circuit starts off with a resistance R_L. However, this is the case for many circuits. An interesting example is that cited by White[9] for the low-frequency (l.f.) compensation of a video amplifier stage. There are other uses for the constant-current equivalent circuit, as in r.f. amplifier stages, but lack of space precludes presentation of these applications.

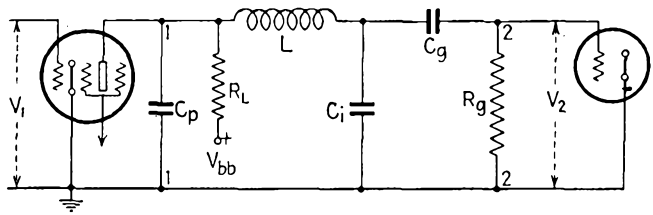


FIG. 9.—Series-peaking video circuit.

22. Further Discussion of the Tube Resistances.—The equivalent circuits presented have been those for a linear tube, *i.e.*, one whose a.c. resistance R_p is constant. This, of course, does not imply that the d.c. resistance is constant. In Fig. 10 are shown three plate-current curves for a linear tube, corresponding to three grid voltages V_{g1}, V_{g2}, and V_{g3}. The a.c. plate resistance (if all curves, for simplicity, are assumed to have the same slope) is the cotangent of angle *BDM*, *CEM*, or *GFM* and is independent of plate voltages that exceed the cutoff values. The d.c. resistance, however, is not independent of the plate voltage. Thus, suppose we set the grid voltage at V_{g1}. Then, for two particular values of plate voltage, *OH* and *OJ*, the two d.c. plate resistances are the cotangents of angles *AOM* and *BOM*, respectively, and hence different. Indeed, for a grid voltage V_{g2} and a plate voltage *OK*, the d.c. plate resistance has still another value, the cotangent of *COK*, and so on, for other curves. Thus, the d.c. plate resistance is a function of the plate and grid voltages for all values of these, whereas the a.c. plate resistance may be independent of these for a limited range of values of these two voltages. It

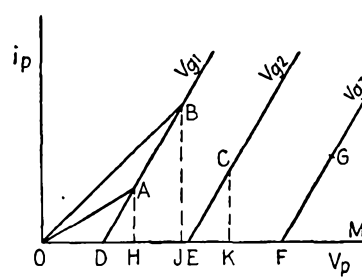


FIG. 10.—D.c. and a.c. plate resistances for a linear tube.

is for this reason that the latter resistance is preferred in the power-series method of solution, which is the method essentially employed in the previous two sections, although the power series consisted of only one term, the first-degree term. In the graphical method of solution, the d.c. resistance will generally be found to be the preferred parameter (note this fact in reading Chap. IV, particularly).

The vacuum tube is peculiar in that the d.c. and a.c. resistances are so widely different in value. Indeed, a little thought will indicate that the equivalent circuit is evident without the need of the derivation given, since an examination of Fig. 10 shows that the grid voltage seems to act more as a fixed potential μ times as great in the plate circuit than as a means of varying the internal resistance of the tube. This is to be expected physically from the fact that the space-charge effect of the grid is a fixed effect in the plate circuit, rather than an effect proportional to the plate current, which is the manner in which a resistive voltage drop would operate. The tube thus acts like the circuit shown in Fig. 11, in which C represents a kind of electrical check valve or diode that offers no resistance to the flow of current in the direction denoted by the arrow but does not permit current to flow in the opposite direction. In this circuit, if V_p exceeds μV_g, current flows owing to the effective voltage $V_p + \mu V_g$ where μV_g is opposite in polarity to V_p, as shown, and is limited (in the absence of an external load) solely by R_p. If the current in this circuit were plotted against V_p, curves similar to those shown in Fig. 10 would be obtained for different fixed values of μV_g.

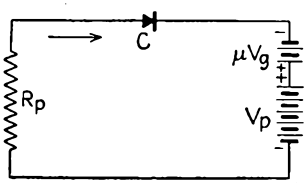


FIG. 11.—Equivalent circuit for a triode.

Nevertheless, it is more correct from an energy viewpoint to view the linear tube as a variable d.c. resistance, controlled by grid voltage, rather than as a fixed a.c. resistance R_p in series with two voltages V_p and μV_g. This is because the source of μV_g does not furnish or absorb any energy from the circuit if μV_g is negative, whereas in Fig. 11 it is evident that the source of μV_g receives energy from the source of V_p when V_p exceeds μV_g. However, as far as considerations pertaining to an external load impedance are concerned, the representation equivalent to Fig. 11 is valid and much simpler than the representation of the tube

as a variable d.c. resistance. This is why the equivalent circuit similar to Fig. 11 is employed when external load relationships only are required.

It is of interest to note that there are devices which are variable resistors for which the a.c. and d.c. resistances are equal. An example of this is the carbon button microphone. Here the resistance of the carbon granules of the button depends upon the pressure of the diaphragm on them. Let us assume such a button is under a normal pressure (say, atmospheric) and is in series with a direct voltage V. Let the normal resistance due to the normal pressure be R, and the variations in R due to variations in the pressure be $\pm\Delta R$. For normal, steady pressure, the circuit equation is given by

$$IR = V \tag{31}$$

For variations in pressure, we obtain variations $\pm\Delta I$ in the current, such that

$$(I \mp \Delta I)(R \pm \Delta R) = V \tag{32}$$

Subtracting Eq. (31) from Eq. (32) we obtain

$$\mp\Delta IR - \Delta I\,\Delta R = \pm I\,\Delta R$$

If ΔR is sufficiently small compared with R, then ΔI will be small too, and $\Delta I\,\Delta R$ can be ignored, so that

$$\mp\Delta IR = \pm I\,\Delta R \tag{33}$$

Equation (33) states that the variations in resistance $\pm\Delta R$ may be replaced by an equivalent voltage $\pm I\,\Delta R$, which, acting in a circuit of resistance R, causes a current $\mp\Delta I$ to flow. Equation (33) is thus similar to Eq. (31). We thus have here, too, a circuit of fixed resistances and variable voltages equivalent to the actual circuit of fixed voltage V and variable resistance $R \pm \Delta R$.

There is an important difference, however, between this equivalent circuit and the equivalent plate circuit of the triode. In the former, the apparent voltage $\pm I\,\Delta R$ is a function of the direct voltage V through the quantity I, as well as a function of the variations in pressure through the quantity $\pm\Delta R$. In the equivalent plate circuit, on the other hand, the apparent voltage μV_g is independent of direct plate voltage and current and a function solely of μ and V_g.

This difference is also exhibited by the actual terminal characteristics of the microphone as compared with those of the triode. In Fig. 12 are shown three terminal characteristics for the carbon button for normal pressure, higher pressure, and lower pressure (curves marked R, $R - \Delta R$, and $R + \Delta R$, respectively). We also note here that the d.c. resistance equals the a.c. resistance for any particular value of pressure, as mentioned previously; but now we note, in addition, that both resistances vary, in contrast to the constancy of the R_p of a linear tube.

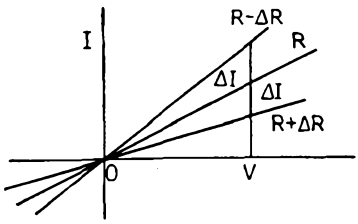


FIG. 12.—Terminal characteristics for a carbon-button microphone.

23. Equivalent Circuit for a Nonlinear Tube.—The reader may wonder if, similar to the equivalent plate circuit for a linear tube, there exists an equivalent circuit for the nonlinear tube, *i.e.*, for a tube with such curved characteristics that even for small grid-voltage excursions, or swings, the approximation of a linear tube is insufficient or for an actual tube in which the grid swings are too great for the linear-tube approximation to hold. The answer is yes; the equivalent circuit for the nonlinear tube is the more general, and that for the linear tube is but a special case of the former. Unfortunately, the derivation for the nonlinear tube is too involved to warrant presentation here but may be inferred from the series of terms occurring in Carson's* method of successive approximations for power-series solutions. Suffice it to say that the equivalent circuit involves a series of generators in circuit with the R_p of the tube and the external load impedance. The R_p may be that given by the reciprocal of the slope at the operating point—the point on the characteristics from which the a.c. excursions take place. The generated voltages are quantities involving V_p, V_p^2, V_p^3, . . . , μV_g, $(\mu V_g)^2$, $(\mu V_g)^3$, . . . If sufficient of these voltages are included, the equivalent circuit will give correctly, not only the fundamental component of the current through the external load impedance, but the distortion terms as well. Generally, however, so many generators must be evaluated if the characteristics are greatly curved as to discourage

* *Proc. I.R.E.*, 1919.

the average engineer from employing this method in the solution of actual problems.

We note, in passing, that this more general equivalent circuit replaces, by the use of the compensation theorem, not only the variable d.c. resistance, but also the variable a.c. resistance of the nonlinear tube, with a constant a.c. resistance and sufficient apparent generated voltages to compensate for the variability of the a.c. as well as the d.c. resistance. It is an interesting artifice from a theoretical point of view; and if a few terms are sufficient satisfactorily to represent the curvature of the characteristics, it can be used in actual problems without a prohibitive expenditure of labor. In Chap. III, as an alternative, will be given a graphical proof of the equivalent-plate-circuit theorem.

24. Conclusion.—This concludes our rather brief discussion of the thermionic vacuum tube, particularly the triode. Further reference will be made in succeeding chapters to the tetrode, pentode, and diode tubes. However, from now on the presentation will be mainly graphical in nature, although power-series methods will be employed where they throw further light upon the operation of the tube and its circuit.

BIBLIOGRAPHY

1. LANGMUIR and ROGERS: *Phys. Rev.*, **4**, 544, 1914.
2. CHILD, C. D.: *Phys. Rev.*, **32**, 492, 1911.
3. PAGE, LEIGH: "Introduction to Theoretical Physics."
4. LANGMUIR, I.: *Phys. Rev.*, **2**, 450, 1913.
5. DEFOREST, L.: U.S. Patent 841387 (1907); U.S. Patent 879532 (1908).
6. VON BAEYER: *Verhandl. deut. physik. Ges.*, **7**, 109, 1908.
7. VAN DER BIJL, H. J.: *Verhandl. deut. physik. Ges.*, **15**, 338, 1913; *Phys. Rev.*, **12**, 180, 1918.
8. VODGES, F. B., and F. R. ELDER: *Phys. Rev.*, **24**, 683, 1924.
9. WHITE, E. C.: English Patent 456450.

CHAPTER III

ELEMENTARY GRAPHICAL CONSTRUCTIONS

1. Introduction.—Before beginning the discussion of graphical methods of solution, it may be well to mention that these methods are not necessarily restricted to the solving of vacuum tube problems but can be applied to mechanical and magnetic circuits as well. The chief difficulty in the case of the latter is that the primary data on the magnetic circuit are not accurately known, particularly the relationship between magnetomotive force H and flux density B when the former is a complex wave function of time. In this case the terminal characteristic (B-H curve) is a hysteresis loop containing minor hysteresis loops, and the exact behavior of B vs. H for these minor loops is at least open to discussion. Often, derived curves obtained from secondary data, such as that of incremental permeability vs. H, are used, a procedure somewhat analogous to that employed in solving detector circuits (Chap. VI).

Another problem that might profitably be attacked by graphical methods is that involving nonlinear compliances, such as those encountered in loud-speaker structures. Here it may be that the production of subharmonics can be demonstrated in a more illuminating manner than by analytical methods.

As a final suggested example, the generated-voltage vs. field-current and armature-speed characteristics of a shunt generator can be studied by graphical means, and the behavior of the machine under various operating conditions nicely predicted. For further information concerning this the reader should consult "Direct-current Machinery" by McFarland.

2. General Considerations.—In the previous chapter the mechanism of thermionic emission from various kinds of emitters, the action of the space charge in a tube and Child's law, the effect of the introduction of a grid in the tube, the a.c. and d.c. plate resistances, the amplification factor μ, the transconductance G_m, and the equivalent-plate-circuit theorems were all discussed.

The derivation of the latter theorems depended upon the use of a simple power series to represent the terminal characteristics of the tube, *viz.*, a single term involving the plate and grid voltages to the first degree. It was pointed out that, if the tube were nonlinear, a prohibitive number of terms of higher degree might be necessary adequately to represent the characteristics. The Fourier method of representation was not employed because it required that the over-all terminal characteristic of tube and external load impedance be known, rather than the terminal characteristics of the individual components and their method of connection.

The graphical method takes the curve or curves as determined experimentally and operates geometrically upon them. That is, the graphical method accepts the characteristic "as is" and does not concern itself with the inner meaning of its shape any more than is absolutely necessary. It therefore operates with directness and dispatch and avoids the complications inherent in the analytical method.

In view of the above, the question may arise as to why the graphical method is not used to the exclusion of the analytical method. The answer is that the former has several serious disadvantages as well as the above advantages. The disadvantages are mainly as follows:

1. The accuracy of graphical results depends upon the mechanical factors involved in drawing the curves and geometrical constructions, such as the thickness of the lines.

2. The graphical method is special: it usually gives a particular answer to a particular set of initial conditions. It does not indicate, as a general rule, the optimum initial conditions to give the best results. Thus, if we are given the power pack voltage applied to a tube, the grid bias, the grid signal voltage ("grid swing"), and load impedance, we can find the plate-current variation and, by an easy calculation, the power output of the tube. We cannot, however, find directly the optimum bias, grid swing, and load impedance for maximum power output for a given power pack voltage except by a series of trials. The analytical method will give us this information, though often in complicated form. Possibly the main reason for the foregoing defect in the graphical method is the fact that it has not as yet been developed to the fullest extent; it may overcome this

limitation as it is improved. We are as yet bound to a great extent by the old Greek tradition of ruler and compass constructions, and this tends to limit the scope of the graphical method.

3. The method fails under certain complicated conditions such as when the operating point changes appreciably with the signal voltage. However, successive approximations can be made to the true state of affairs, and it is further questionable whether the analytical method is really superior for these more complicated conditions.

4. Graphical constructions become far too involved, if not actually impossible, when more than three variables are involved, as in the case of a tetrode or pentode tube, unless all but three of the variables can be kept fixed, whereupon these become parameters rather than variables.

It may therefore be concluded that there is no general method of attack for vacuum-tube problems and that the best that can be done today is to employ both methods in an effort to obtain even working or approximate solutions. It is to be hoped that some day a general treatment of nonlinear networks (of which the vacuum tube is but one example) will be developed that compares with the treatment of linear networks.

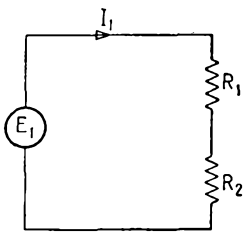


FIG. 13.—Two linear resistances in series.

3. Simple-series Linear Circuit.—To proceed with the graphical method, suppose two constant resistances R_1 and R_2 are in series with an e.m.f. E_1 (Fig. 13). The current flow I_1 is simply given analytically by Ohm's law as

$$I_1 = \frac{E_1}{R_1 + R_2}$$

The voltage drop across R_1 is then I_1R_1; that across R_2, I_1R_2. Denote the voltage drop across R_1 by E_A; that across R_2 by E_B.

Then

$$I_1R_1 = E_A \tag{1}$$

$$I_1R_2 = E_B = E_1 - E_A \tag{2}$$

Suppose it is desired to solve this circuit graphically. The independent variable will be voltage; the dependent variable, current I_1. In order that there be but one independent variable,

say, E_A, the relationships given in Eq. (2) may be used in the form where I_1R_2 is equal to $(E_1 - E_A)$ rather than to E_B. If these two equations are plotted, the straight lines OC and BC (Fig. 14), respectively, are obtained.

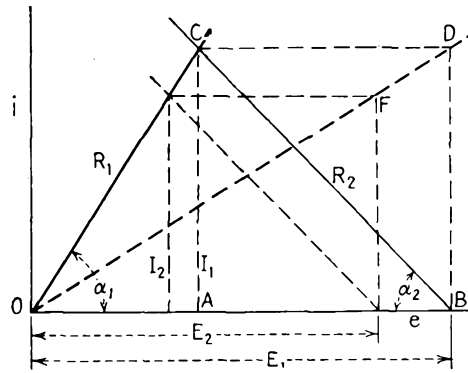


FIG. 14.—Graphical constructions for linear series circuit.

The slope of OC is evidently

$$\cot \alpha_1 = \frac{E_A}{I_1} = R_1 \tag{3}$$

The plot of Eq. (2), *viz.*, BC, is evidently shifted from the origin by a distance E_1 and is of a negative slope corresponding to

$$\cot (\pi - \alpha_2) = \frac{E_1 - E_A}{I_1} = -\cot \alpha_2 = \frac{-E_B}{I_1} = -R_2 \tag{4}$$

OC and BC are called the *terminal characteristics* or *load lines* of R_1 and R_2, respectively; and if one of them is plotted from the origin, the other must be plotted from a shifted origin and with a negative slope, when it represents a resistor in series with one corresponding to the other load line. Where these two load lines intersect, at C, is the simultaneous solution of Eqs. (1) and (2).

This intersection gives a value of current I_1 that is the same for both resistors. But this is the value of current that is sought, since in a series circuit the current is the same throughout the circuit. Hence I_1, the current in the circuit of Fig. 13, has been determined graphically by the intersection of the two load lines. The distance OA represents the voltage E_A required across R_1 to force the current I_1 through it, and AB represents the voltage E_B required across R_2 to force I_1 through it. It is

evident from Fig. 14 that

$$OA + AB = E_1 \tag{5}$$

and

$$OA = I_1 \cot \alpha_1 = I_1 R_1 \tag{6}$$

and

$$AB = I_1 \cot \alpha_2 = I_1 R_2 \tag{7}$$

so that the graphical method checks with the analytical method.

Suppose another voltage E_2 is employed, shown in Fig. 14 as less than E_1. The same procedure is followed as for E_1, and current I_2 is obtained. This may be repeated for any other voltage or voltages desired, and the corresponding currents obtained. If the circuit is viewed as a whole, its action may be represented by means of a load line too. Thus, for a voltage E_1, the current is I_1. The ordinate DB equal to I_1 is drawn at the end of E_1. Similarly, point F is obtained for voltage E_2. Points D and F are on the load line for the two resistors R_1 and R_2; *i.e.*, they are on the load line for a resistor R equal to $R_1 + R_2$ or the total circuit resistance. This load line passes through the origin O, so that it is represented by the line OFD. This is a straight line, since those for the two resistors are straight.

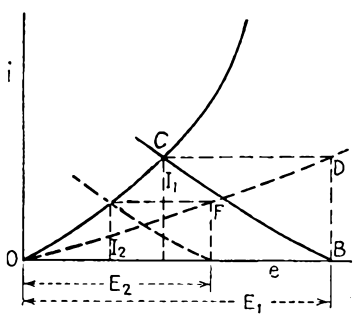


FIG. 15.—Graphical constructions for nonlinear series circuit.

In this particular example the graphical solution exhibits no advantage over the analytical, so that as yet its utility requires demonstration. Let us now take the case of two nonlinear resistors in series with an e.m.f. These will give rise to a figure similar to Fig. 12, except that R_1 and R_2 now vary with the current through them. The load line for each is hence a curve and they can be plotted from the ends of E_1, as shown in Fig. 15. Their intersection at C gives the desired value of current I_1. For some other voltage E_2 the current I_2 is obtained. If these current values are projected over to D and E, respectively, points on the load line for R_1 and R_2 in series are obtained, and this load line has been drawn in as OED. The latter is evidently a curved line, also.

Apparently, the graphical solution is as simple as that for the fixed resistors. Let us now review the analytical method. Instead of Eq. (1), we now have for R_1

$$i = k_1 e + k_2 e^2 + k_3 e^3 + \cdots + k_n e^n \tag{8}$$

and for R_2

$$i = k_1'(E - e) + k_2'(E - e)^2 + \cdots + k_m'(E - e)^m \tag{9}$$

where E is the total applied voltage. These two power series have been purposely written with a different number of terms to suggest the fact that the two load lines have different curvatures, *i.e.*, that the two resistors R_1 and R_2 vary in different ways with the current. To find the current through the two in series, Eqs. (8) and (9) must be solved simultaneously. Such equations, in general, are not solvable if of degree higher than the fourth, and the roots can be found only approximately, by Horner's method, for example. The above thus demonstrates in a striking manner the directness of the graphical method, and this is true in the case of many nonlinear circuits.

In Chaps. I and II was mentioned an analytical method that gives a solution by a series of approximations.[1] This results in involved computations, but it is general in scope. The circuit represented by Fig. 15 can be, in physical form, that of a driver tube energizing the grids of power tubes in a class AB_2 arrangement. Thus the driver tube may exhibit some nonlinearity as represented by load line OC, and the grid circuit of the power tube may exhibit nonlinearity as represented by load line BC. This matter will be analyzed more thoroughly in the case of a class AB_2 circuit.

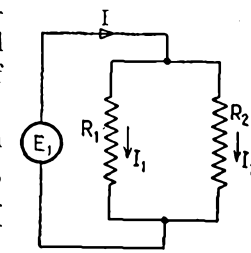


FIG. 16.—Parallel circuit.

4. Parallel Circuit.—In Fig. 16 are shown two variable resistors R_1 and R_2 in parallel, and an e.m.f. E_1 impressed across their terminals. In this case the total current I should be given in order that this problem be equivalent to that of the series circuit; *i.e.*, I should be known, and then the component currents I_1 and I_2 through R_1 and R_2, respectively, as well as E_1 can be determined. In practice, however, E_1 is usually given, and I, I_1, and I_2 are to be determined. Hence, the method will differ

somewhat from that employed for the series circuit; in fact, it will be similar to the problem of finding the component voltage drops, as well as the total impressed voltage in a series circuit when the current is known.

Referring to Fig. 17, we note that the axes are reversed: the voltages are plotted as ordinates, and the currents as abscissa; *i.e.*, the conductances of the two resistors are utilized. OA represents the conductance load line for R_1, and AB that for R_2. Length OB represents any desired value of total current I; OD represents I_1, and DB represents I_2. AD represents the value of E_1 that could cause these currents to flow. If some other value of total current I is then chosen and the process repeated, the corresponding values of I_1 and I_2 will be found, and also E_1 for

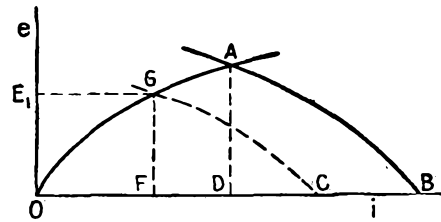


FIG. 17.—Graphical constructions for parallel circuit.

this value I. The load line for the two resistors in parallel can then be obtained by plotting the various values of I against the corresponding values of E_1. Then, for the given value of E, the value of I can be found, and the construction shown in Fig. 17 repeated for this value of I to obtain the corresponding components I_1 and I_2.

The constructions can be shortened, however, as shown in the figure. Thus, the given value of E_1 is projected to curve OA. The intersection is at G, whose projection on the i axis is F. Then OF is the value of I_1 for the given value of E_1. The same thing could be done for AB, and I_2 found. Then I would be the sum of I_1 and I_2. Or load line AB can be shifted a distance FD, so that its plot is the broken line GC. Then OF represents I_1, FC represents I_2, and OC represents I.

Parallel circuits of nonlinear resistances are not so often encountered as series arrangements, so that this construction is seldom employed. However, where required, it is as easy to employ as the construction for the series circuit.

5. Series Parallel Circuits.—Series parallel circuits can be solved by finding the load lines for the series parts of the network and those for the parallel parts and then combining these load lines as the circuit dictates. The above method fails, however, in the case of Wheatstone bridge circuits, but these are beyond the scope of this text.

6. Application to the Vacuum Tube.—If the graphical method is applied to the vacuum tube, such as a triode, then a complication is encountered: the plate current (dependent variable) is a function of two independent variables, the plate voltage and the grid voltage. Under these conditions the plot is a surface;

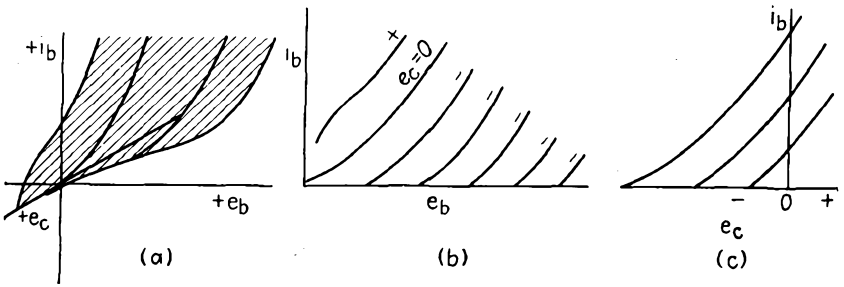


FIG. 18.—Triode surface and characteristic curves.

this is shown in Fig. 18*a* as a shaded area, with representative cross-sections, or "ribs," on it.

It is inconvenient to perform the graphical constructions upon a space model; hence the principles of solid analytic and also of descriptive geometry are employed to obtain projections of this space model upon a plane, *i.e.*, upon a sheet of paper.

The equation of the tube surface may be given as

$$i_b = F(e_c, e_b) \tag{10}$$

Consider the equation

$$e_c = c \tag{11}$$

where c is some constant. Equation (11) represents a plane parallel to the i_b-e_b coordinate plane, and at a distance of c units along the e_c axis from the latter plane.

The simultaneous solution of Eqs. (10) and (11) represents the intersection of the tube surface with the plane, and this intersection may be considered a rib of the tube surface. If this

rib be projected over to the i_b-e_b coordinate plane, it forms a curve as shown by any one of the family in Fig. 18*b*. Different values of c give rise to different curves of the family. It is evident that, since i_b is zero for negative values of e_b, only the first quadrant is necessary to depict the significant features of all curves of the family.

Analytically, the simultaneous solution of Eqs. (10) and (11) means that e_c has been made a parameter, so that i_b thereby becomes a function of e_b alone and thus can be plotted on a plane. As the parameter e_c is changed from one value to another, the corresponding plot for i_b vs. e_b changes from one plane curve to another, and the totality of curves constitutes the plate family of characteristics.

In a similar manner, e_b may be made a parameter and i_b plotted against e_c, and this will result in a family of curves too. This family is known as the plate-current–grid-voltage characteristic (Fig. 18*c*); and since i_b is not necessarily zero for negative values of e_c, the family requires both the first and second quadrants for its sphere of activity. In this family, the curves proceed to the left as the plate voltage is raised. Which family of curves is to be used depends upon the problem being considered. Usually, the grid voltage is known, and the plate current and plate voltage are unknown except at one point, so that the grid voltage is made the parameter and the i_b-e_b family is therefore used. Moreover, the constructions, as will be shown, are usually simpler for this family. In passing, it is well to note that the surface constitutes more than the shaded portion shown: it includes the points in the i_b-e_b plane where i_b is zero. This is mentioned here because later some confusion may arise in the reader's mind in applying the graphical solution. For large grid swings it may appear that a curve ends on the axis before it intersects some line of construction. In that case it is to be remembered that the curve then really proceeds along the axis, so that the intersection of the above line of the construction is with the axis.

We now come to the question of the resistor, which is usually placed in series with the plate of the tube and is moreover practically always linear. Figure 19 shows the electrical circuit. Here R_L is the load resistance in the plate circuit. We have seen earlier in this chapter that the load line for a fixed resistance is

a straight line. This concept must now be modified in view of the three-dimensional surface plotted for the tube characteristic. It is obvious that the relation between current through and voltage across the resistor is independent of the grid voltage applied to a tube which may or may not even be connected to this resistor. This fact can be shown by plotting the characteristic for R_L as a load plane instead of as a load line. This load plane must intersect the e_c-e_b plane in a line parallel to the e_c axis in order to show the load plane's independence of the grid voltage applied to the tube. In Figure 20 is portrayed the load plane for R_L (as a shaded area). Note that the intersection AB is parallel to the e_c axis (hence perpendicular to the e_L axis) and

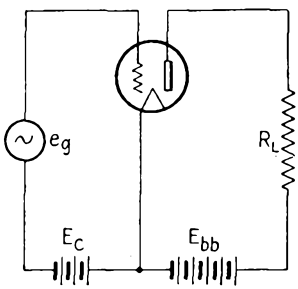


FIG. 19.—Triode circuit.

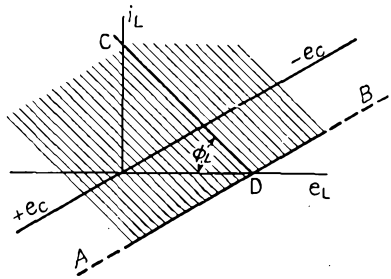


FIG. 20.—Load plane for linear resistor.

that the load plane is perpendicular to the i_L-e_L coordinate plane. The intersection of this load plane with the i_L-e_L plane, or CD, is the more usual load line for R_L, such that

$$\cot \phi_L = R_L \tag{12}$$

If we now wish to find the current flow through the tube and R_L, in Fig. 19, we must find the intersection of the tube surface with the load plane of R_L. The intersection of two surfaces is a curve in space, and this is shown in Fig. 21 as AB. This gives the locus of the plate current (which is also the current through R_L) for different values of grid voltage.

Once again we raise the objection to three-dimensional constructions and once again avoid it by plotting the projection of AB on the i-e plane, that is, on a sheet of paper.

It will be evident that, since the load plane for R_L is drawn perpendicular to the i-e plane, all lines in the former plane, even if curved, will project over to the latter plane as lines coincident

with the intersection of the two planes themselves. Since the intersection of two planes is a straight line, these other lines, such as AB, will appear as straight lines. Hence, if we operate graphically and according to the principles of descriptive geometry on the projections upon the i-e plane of the tube surface and that of the load plane for R_L, we can obtain all the information we require concerning the performance of the circuit. The construction reduces to that of finding the intersection of the load line for R_L with the plate family of characteristics on a plane.

If AB, Fig. 21, be projected to the i-e_c plane, it will appear as a curve, whereas it appears as a straight line on the i-e plane, of

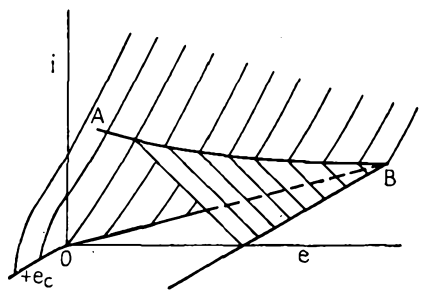


FIG. 21.—Three-dimensional graphical solution for triode circuit.

slope equal to cot ϕ_L [Fig. 20 and Eq. (12)], and hence is easier to draw in the latter case. This is possibly the most important reason why the i-e or, more specifically, the i_b-e_b family of curves is preferred to that of the grid family.

7. Resistance Coupling.—The circuit of Fig. 19 employs what is known as *resistance coupling* and is mainly used in voltage amplifiers to obtain a voltage across R_L, which is an amplified copy of the grid signal voltage e_g. This circuit will now be studied graphically in greater detail by applying the principles outlined in the previous section. As stated there, it is more convenient to use the projections of the surfaces than the surfaces themselves in applying the graphical method. Accordingly, instead of the tube surface, the i_b-e_b family of curves is employed; and, instead of the load surface for R_L, its load line. It is to be noted first that, if the grid voltage were sufficiently negative, the current in the plate circuit would be reduced to zero (plate-current cutoff). Under such conditions there would be no

voltage drop in R_L, and hence the plate voltage would equal E_{bb} (Fig. 19).

This serves to locate one point (O, E_{bb}) of the load line of R_L on the family of characteristics. Since R_L is a linear resistance, its load line is straight, and hence only one other point of it need be found to determine it. This can be done by assuming any convenient voltage value E_L, across R_L, and solving for the resultant current through R_L by Ohm's law. Thus

$$i_L = \frac{E_L}{R_L} \tag{13}$$

The point (i_L, E_L) can then be located on the graph paper, joined to the other point (O, E_{bb}) by a straight line, and thus the load line of R_L obtained.

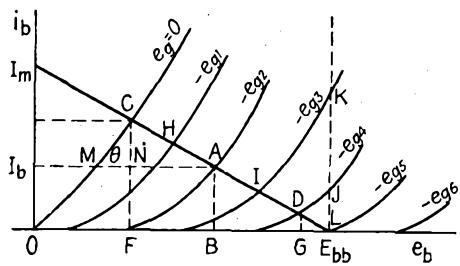


FIG. 22.—Two-dimensional graphical solution for triode circuit.

In Fig. 22 is shown the i_b-e_b family of curves. The voltage between the plate and cathode of the tube (e_b) is to be measured from the origin to the right; the voltage across R_L (Fig. 19), from E_{bb} to the left. For the cutoff grid voltage $-e_{g_6}$ the plate current is zero, the plate voltage is E_{bb}, and the voltage across R_L is zero. If the grid is driven sufficiently positive (value not indicated, since never realized in an actual tube), the plate current is at a maximum, I_m; e_b is zero; and the voltage across R_L is OL, that is, E_{bb}. For intermediate values of the grid voltage, the plate current is between I_m and zero in value, e_b is some value less than E_{bb}, and the voltage across R_L is the difference between E_{bb} and e_b. The graphical construction is therefore exactly similar to that shown in Fig. 15.

Suppose, in the circuit shown in Fig. 19, that e_g is zero. The grid voltage is then merely E_c, so that the plate current is steady, or d.c. Assume that E_c is the value $-e_{g_2}$ shown in Fig. 22. The

intersection of the load line I_mL with the curve for $-e_{g_2}$, or A, gives the value of the d.c. component (shown as I_b). Then OB represents the value of e_b, and LB the value of the voltage drop across R_L (when no signal, e_g is impressed). It is evident from the figure that the sum of the two voltages is E_{bb}, the plate-circuit or B battery voltage. Point A is known as the *quiescent point*, as it is the value of the plate current when the tube voltages are steady, or quiescent.

If a signal voltage e_g is applied, the instantaneous value of the voltage between the grid and cathode changes, since it is now equal to $e_g + E_c$ or e_c. The plate current now varies along the load line I_mL. If e_g is a sine-wave voltage of peak amplitude $E_c(= -e_{g_2})$, then, when it is in the positive direction and peak value, it just cancels E_c, so that the instantaneous grid voltage is zero; and when it is at the negative peak value, it causes the instantaneous grid voltage to be twice $-e_{g_2}$, or $-e_{g_1}$. The plate current accordingly varies from A to C back to A again, then down to D, and then back to A during one cycle of e_g; that is, it varies about the point A to a maximum value C and a minimum value D. Simultaneously, the plate voltage varies from the normal value OB to OF and to OG, while the voltage across R_L varies, respectively, from the normal value of LB to LF and LG. At all times the sum of these two voltages is $OL = E_{bb}$, the B supply voltage. It is to be noted also, from the foregoing, that as the plate current rises the plate voltage drops while the voltage across R_L rises, and vice versa when the plate current drops, so that when e_g is sinusoidal, e_p is approximately a sine wave, 180 deg. out of phase with e_g, while the voltage across R_L (as measured from the B supply end to the plate) is approximately a sine wave, in phase with e_g. This all checks with the analytical treatment of the equivalent-plate-circuit theorem[5] given in Chap. II, although that treatment assumed the tube to be linear, whereas here this restriction need not be made.

8. Graphical Proof of the Equivalent-circuit Theorem.—The above graphical construction affords an interesting proof of the equivalent circuit theorem. Consider triangle ACM, Fig. 22. It represents a circuit containing R_L (whose load line is CA) in series with the R_p of the tube (whose load line is MC). The current CN, which is the a.c. component of i_b for a grid swing from $e_c = -e_{g_2}$ to $e_c = 0$, may alternatively be regarded

as being produced by a voltage equal to MA impressed across the above two resistors in series, and the triangle is the graphical solution of this alternative viewpoint. Since MA is parallel to the e_b axis, it represents a change in plate voltage, together with a compensating change in grid voltage, which leaves the current unchanged at the value $AB = I_b$. By definition, this is the μ of the tube; *i.e.*,

$$\frac{MA}{0 - (-e_{g2})} = \frac{MA}{e_g} = \mu \Big]_{i_b = I_b} \tag{14}$$

so that

$$MA = \mu e_g \tag{15}$$

Hence the vacuum tube, which is a resistor adjustable by grid voltage, in series with a linear resistor R_L and a constant voltage E_{bb}, may be replaced by a circuit equivalent to the actual plate circuit in which a constant resistance of value R_p is in series with R_L and a voltage μe_g. The tube may thus be regarded as equivalent to an active source of generated voltage μ times the input grid signal voltage e_g and of internal resistance R_p.

Any curvature in the tube characteristic, MC, may be regarded as evidence of nonlinearity in R_p. An alternative viewpoint, as embodied in the power-series method, is to regard the tube of constant resistance equal to the initial slope of MC and generating a series of voltages proportional to μe_g, $\mu^2 e_g^2$, $\mu^3 e_g^3$, etc., which, together with the above constant resistance, give rise to current CN.

9. Dynamic Characteristic.—In a simplified analytical treatment it is generally assumed that there is no load in the plate circuit and, in some cases, that the tube characteristic is linear. This is done to simplify the mathematics. As pointed out earlier in this chapter, where these restrictions are not imposed more involved computations based upon Carson's method of approximations can be employed. The reason for these complications is that the vacuum tube has, in general, a nonlinear characteristic. Owing to this, all the voltages of the tube at any instant must be taken into account before the plate current can be determined, since these voltages cross modulate one another, so that the principle of superposition does not hold, and the effect of each voltage cannot be determined independently of the

other voltages, as can be done in ordinary linear circuits (see Chap. I).

If the tube has a load impedance connected to its plate, such as R_L (Fig. 19), and even if this load impedance is linear, then, because of the nonlinearity of the tube, it is necessary to find all the voltages across the tube elements before its plate current can be determined. Where the load impedance is zero, the plate voltage is constant and equal to E_{bb}, the B supply voltage, and the problem is simplified. But where the load impedance is not zero, the plate voltage varies in inverse manner with the plate current, as just demonstrated in the preceding section, and this factor in turn changes the variation of the plate current. It has been found convenient to classify these two cases as follows:

1. The variation of i_b with e_c when the plate load impedance is zero.

2. The variation of i_b with e_c when the plate load impedance is not zero.

The plot of (1) is called the *static characteristic*, because e_b remains static, or constant; the plot of (2) is called the *dynamic characteristic*, because e_b is variable, changing, or dynamic. These two characteristics can be easily determined graphically and, with much greater difficulty, analytically by Carson's method. The above statement holds rigorously, however, only for a load resistance such as R_L (Fig. 19) and but approximately for an impedance such as a choke and resistor in parallel. Where the load impedance is a pure inductance, the graphical method becomes more involved. In general, the dynamic characteristic may be regarded as the functional relationship between the primary cause e_c and the final effect i_b. Intermediate effects of the primary cause, such as e_b, though in turn contributory causes to the variation of i_b, are contained in the coefficients of the functional expression between e_c and i_b. Thus the equation of the dynamic characteristic contains e_c as the independent variable, i_b as the dependent variable, and the effects of R_L (variation of e_b) as parameters of this equation.

10. Graphical Determination of the Dynamic Characteristic.—Figure 22 furnishes all the information necessary for the plotting of the dynamic characteristic, since it gives the different values of i_b for the corresponding values of the grid voltage e_c. The characteristic is plotted in Fig. 23. The points in this figure are

obtained from the intersections of the load line with the plate family of curves, and the letters here are the same as those for Fig. 22. This curve differs from a static characteristic for the tube in that here the plate voltage is different for each point, whereas in a static characteristic it remains constant. It is to be noted that this particular dynamic characteristic is for a load resistance R_L. For some other value of load resistance, the dynamic characteristic will be some other curve. This will be further discussed in Sec. 12 of this chapter.

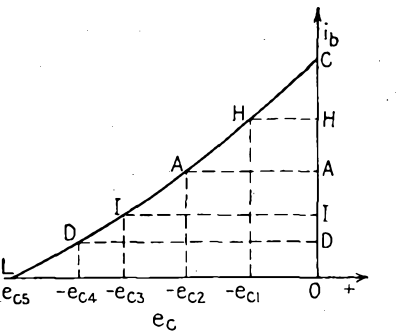


FIG. 23.—Dynamic characteristic.

11. Graphical Determination of the Static Characteristic.—To obtain the static characteristic, the load line for a zero load resistance must be drawn, and its intersection with the plate family of curves found. Evidently, if R_L is zero, its load line is perpendicular to the e_b axis through the point E_{bb} (Fig. 22). It is shown as a broken line in Fig. 22, and two intersections are given, J and K. These are then plotted as the static characteristic in Fig. 24, as well as the dynamic characteristic for comparison.

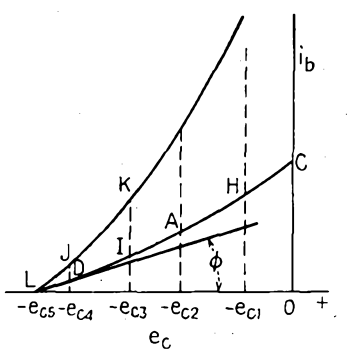


FIG. 24.—Comparison of static and dynamic characteristics.

12. Comparison of the Static and Dynamic Characteristics.—From Figs. 24 and 25, the static and dynamic characteristics may be compared and the following points noted:

1. The static characteristic rises more steeply then the dynamic characteristic.

2. The latter is more nearly straight, which means that the higher order terms in the power series representing it have smaller coefficients. This in turn means that the harmonic distortion generated is less. The higher R_L is, the flatter the curve and the less its slope. The reader may verify for himself that

$$\cot \phi = \frac{\mu}{R_L + R_p} \tag{16}$$

where R_p is the internal plate resistance of the tube at the particular point L in question. From a physical viewpoint, it is to be noted that R_L, which is assumed a constant parameter, masks the variability of R_p—*i.e.*, the nonlinearity of the tube—and that, the higher R_L is, the greater this masking and the more nearly linear is I_b with e_c.

3. In Fig. 25 have been plotted various dynamic characteristics for different values of R_L, and also the static characteristic once more, for comparison. The statements made in (2) are seen to be verified. Curve LJK is the static characteristic, LC the

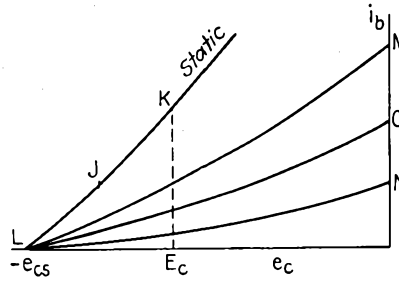


FIG. 25.—Comparison of various dynamic characteristics.

dynamic characteristic for R_L assumed previously, and LM for a smaller value of R_L and LN for a larger value. The point to be emphasized here is that all the curves cut off at the same cutoff bias $-e_{cs}$, a matter that is not often sufficiently stressed. In ordinary class A operation of the tube, a value of grid bias E_c is chosen such that, if a sinusoidal signal voltage e_g is impressed, then, on the positive grid swing, the total grid voltage

$$E_c + e_g = e_c$$

does not exceed zero, so that the grid does not go positive with respect to the cathode and hence draw current. This means that the grid does not require any appreciable power from the source, but only voltage. On the negative grid signal swing, the total grid voltage e_c must not exceed the cutoff point for the tube (point L, Figs. 22 to 25), for otherwise the plate current will not be able to follow the grid signal voltage and hence will not be a true copy of it, in which case the amplified voltage across R_L will not

be a true copy of the signal voltage. Since the positive and negative halves of e_g are equal when it is a sine wave, it is evident that the grid bias E_c must be halfway between the cutoff voltage $e_c = -e_{gs}$ and the zero point $e_c = 0$. Furthermore, the peak value of the grid signal voltage must not exceed E_c. From Fig. 25 it can be seen that the proper bias and maximum signal voltage will be the same for all values of R_L, since the corresponding dynamic characteristics all cut off at the same point L, where $e_c = -e_{gs}$.

13. Effect of Grid Signal Voltage.—From Sec. 12 it is seen that the maximum grid signal swing is the same for all values of R_L. In Fig. 26*A*, *B* and *C* there has been plotted a dynamic characteristic for the tube, curve *LAC*. In all cases the grid

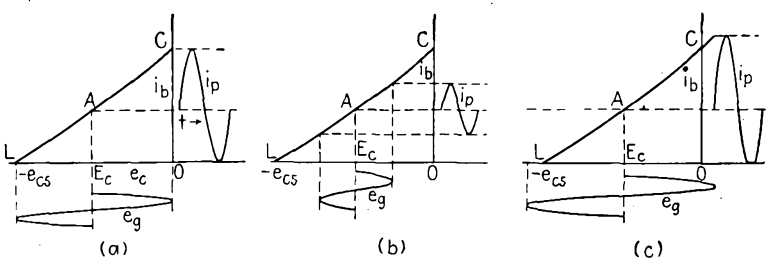


FIG. 26.—Dynamic operation for various values of grid swing.

bias E_c has been chosen halfway between cutoff and $e_c = 0$. In *A* the sinusoidal signal voltage has a peak value equal to E_c. It has been plotted as a function of time, the latter axis extending downward. The corresponding plate current as a function of time has been plotted to the right of the dynamic characteristic *LAC*, since the instantaneous current i_b for each instantaneous value of e_c can be projected over from the dynamic characteristic quite easily. (For the current, the time axis is horizontal, as shown.) From this figure it can be seen that, if the dynamic characteristic is nearly flat, the positive lobe of i_p exceeds the negative lobe (peak values) by very little. It can be shown, both by the power-series[5] method and by the Fourier series, that the excess of one lobe over the other is due mainly to the production of a second harmonic, as well as additional direct current. In Fig. 26*A*, where this excess is small, the distortion and production of additional d.c. component will be small.

These deductions are even more apparent when the signal voltage is small, as in Fig. 26*B*. This checks with the power-series

method of development; for where the exciting voltage is small, the higher order terms ($k_2e_c^2$, $k_3e_c^3$, etc.) of the i_b-e_c expansion* are negligible compared with the first-power term k_1e_c, since the square, cube, etc., of a small quantity are small compared with the first power. This means that the power series reduces practically to the first-power term, or is a linear equation. This in turn implies no distortion products: the current is a faithful copy of the grid voltage.

Figure 26*C* depicts the case where e_g exceeds E_c, so that the grid swings positive on the positive half cycle and the plate current reaches cutoff before the grid has swung through the negative peak value. The plate current is evidently distorted, even if the driver tube preceding this tube were capable of supplying the grid losses when the grid goes positive. The distortion, as shown, is mainly due to plate-current cutoff: the negative lobe is flattened at its peak. The asymmetry of the wave implies at least even harmonic distortion, as well as the production of considerable additional d.c. component. The latter phenomenon is known as "self-rectification" in a tube and is permissible only when the distortion products can be filtered out, as in a narrow frequency-range amplifier. It is also permitted in broad-range amplifiers when the action of one tube is supplemented by that of another, as in a balanced-amplifier circuit, operating AB_1 or AB_2 (Chap. V).

14. Voltage Amplification.—The graphical construction affords a means of evaluating the voltage gain of a resistance-coupled amplifier. As can be seen from Fig. 22, the d.c. voltages applied are E_{bb} and $E_c = -e_{g_1}$. For the load resistance R_L chosen and for a grid swing e_g equal to E_c, the peak-to-peak plate-voltage swing is

$$OG - OF = FG$$

so that the peak voltage is $FG/2$. The stage gain is therefore

$$a = \frac{FG}{2e_g} = \frac{FG}{2E_c} \tag{17}$$

Analytically, the stage gain may be evaluated by means of the equivalent circuit theorem as

$$a = \mu e_g \frac{R_L}{R_p + R_L} = \mu E_c \frac{R_L}{R_p + R_L} \tag{18}$$

* $i_b = k_1e_c + k_2e_c^2 + k_3e_c^3 + \cdots$ is the equation of the dynamic characteristic.

In the latter formula, however, R_p and even μ may vary from their value at the quiescent point A, and their nominal values as given by the manufacturer may be for some other quiescent point. Equation (17), on the other hand, is correct for the actual operating conditions, and it will be found that an experimental determination of the gain usually checks Eq. (17) much more closely than it does Eq. (18).

It will be evident that, as R_L is increased, a will increase, until when R_L becomes infinite the gain will become equal to the μ of the tube as measured along the e_b axis. In practical circuits, however, a shunt path consisting of a grid coupling condenser and grid resistor is present through which the a.c. component i_p of i_b can also flow, so that the load impedance to i_p becomes less than that for the d.c. component I_b, which can flow only through R_L. There are then two load lines to consider, a matter best reserved for a following chapter. A further complication is that the load impedance becomes reactive at low frequencies where the reactance of the grid coupling condenser is appreciable compared with the grid resistance and also at the higher frequencies where tube and stray wiring capacities have an appreciable shunting effect.

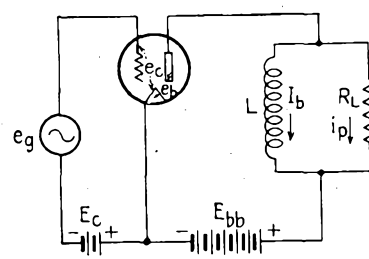


FIG. 27.—Inductive plate-feed circuit.

15. Inductance Plate Feed.—The graphical methods discussed thus far have been concerned with a pure constant resistance in the plate circuit. There is another important type of plate load, however, shown in Fig. 27. This circuit is equivalent in effect to an ideal transformer whose primary is in parallel with the load resistance or across whose secondary there is connected an equivalent load resistance. As may be noted from Fig. 27, L is an ideal inductance whose resistance to direct current is zero, whose reactance to alternating current is infinite, and whose distributed capacity is zero. It may be regarded as an ideal 1:1 autotransformer, which establishes its equivalence to the loaded transformer cited above. R_L is a constant resistance, whose value is therefore independent of frequency. The total load impedance consequently shows zero resistance to direct current and a resistance R_L to alternating current.

When e_g is zero, the plate current is d.c. and flows solely through L, since the latter short-circuits R_L. The plate voltage e_b is then evidently equal to E_{bb}, since there is no voltage drop in the load impedance to direct current. When a signal voltage e_g is impressed, the grid voltage e_c varies by an amount e_g about the direct bias voltage E_c and causes the plate current to vary about its normal steady d.c. value. The variation in the plate current is the a.c. component and, as explained previously in this section, cannot flow through the infinite impedance of the ideal choke L but must flow through R_L. The graphical construction must therefore be modified from that employed for a purely resistive plate load (resistance coupling). The procedure is as follows:

In Fig. 28 is shown the i_b-e_b family of curves. For the grid voltage E_c alone, the load line is vertical and starts at $e_b = E_{bb}$.

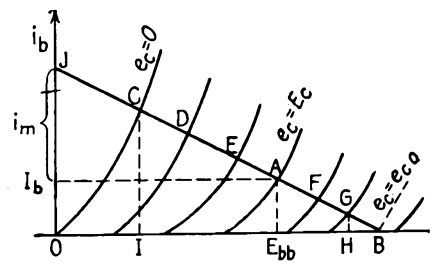


FIG. 28.—Path of operation for inductive plate feed.

The point at which it intersects the curve for which $e_c = E_c$, at A, is the normal d.c. component I_b. This is the quiescent point, about which, as a first approximation, the a.c. component is assumed to alternate. The load line for the latter must pass through A. Another point on the load line is found from the equation $i_m = E_{bb}/R_L$; this value must be added to I_b to give point J, as shown in the figure. The load line JA intersects the voltage axis at point B. This is the point at which the instantaneous value of i_b is zero, or the cutoff point; it occurs when e_c swings sufficiently negative (here shown as $e_c = e_{co}$).

It is immediately apparent that at cutoff the plate voltage exceeds the normal supply voltage E_{bb} by an amount $E_{bb}B$, whereas in resistance coupling the plate voltage at cutoff was just equal to E_{bb}. Also, it is evident that, the higher R_L is, the less is i_m, while point A remains unchanged, so that the load line has less

slope and intersects the voltage axis farther to the right; *i.e.*, the plate voltage at cutoff becomes greater.

This cutoff plate voltage can be evaluated quite simply. Thus, referring to Fig. 27, we know, by Kirchhoff's laws, that the sum of the voltages in the plate circuit must at all times be zero. Also, in an ideal choke, the current (I_b) cannot change—at least suddenly—so that if the internal current of the tube is reduced to zero the choke current I_b must be diverted to R_L. In flowing through R_L it sets up a voltage drop I_bR_L, and this voltage drop will be in the same direction as E_{bb}. Thus,

$$e_b - I_bR_L - E_{bb} = 0$$

or

$$e_b = E_{bb} + I_bR_L \tag{19}$$

From a physical viewpoint, when an attempt is made to decrease the current in a choke the voltage across it rises —here by an amount I_bR_L. This is popularly known as the "inductive kick" and is the cause, for example, of the heavy arc across the switch blades when direct current is interrupted in an inductive circuit. The advantage of this type of load impedance is immediately apparent. It enables the same results to be obtained as for a pure resistance but at a lower B supply voltage.

The above analysis affords another means of determining point B (Fig. 28), so that the load line may be found in two ways, by means of points A and B or points A and J, as desired. It can be seen that the load line is the same as for a resistance-coupled amplifier operated at a plate supply voltage OB, so that the net effect of the choke is to increase in effect the B voltage and therefore, in turn, the maximum permissible grid swing or signal voltage and a.c. component of the plate current for class A operation. The tube can thus deliver more power output; for this reason, power tubes are always connected to their load (usually resistive) through output transformers or chokes depending upon the values of the load resistors.

The phenomena noted above are possibly more easily seen from the i_b-e_c family of curves (Fig. 29). These are obtained from the plate-voltage family of Fig. 28 in exactly the same way as in the case of resistance coupling, *i.e.*, by plotting the intersections $CDEAFGB$. In this way, curve BAC (Fig. 29) is obtained. The static characteristic may also be obtained by drawing a vertical

line through A in Fig. 28 and plotting the intersections here as curve DAE. Note that the tentative operating point A occurs at the steady direct bias voltage E_c and is determined by the static characteristic, while, as has already been explained, all other points of the dynamic curve do not fall on the static curve. As a result, the dynamic cutoff B occurs at a more negative grid voltage than the static cutoff D.

In studying the action of a tube, it is customary, for the sake of simplicity, to assume that a sinusoidal signal grid voltage (e_g) is impressed. In actual operation, the signal voltage impressed

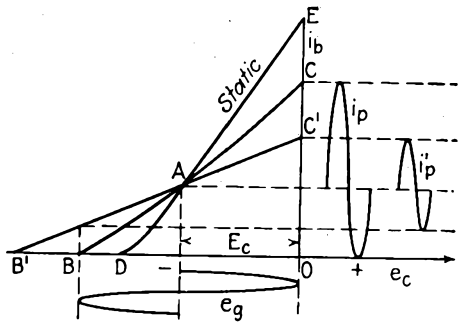


FIG. 29.—Dynamic operation for two values of load resistance.

depends upon the nature of the sound impressed upon the microphone, or, in general, upon the nature of the voltage generated in the pickup device feeding the tube. Consequently, the voltage may be complex, and a thorough analysis of the tube output should take this all into account. However, the effect of a complex wave, such as two sine waves not harmonically related, is to cause harmonic and cross-modulation products to be generated in the plate circuit. If the tube is a component of a broad-range amplifier, these products should be reduced as much as possible. This means a linear relation between i_b and e_c, that is, a straight dynamic characteristic. Then, if a simple sine-wave signal voltage is impressed upon the grid, the plate current can be expected to be practically sinusoidal too. Since, for the linear dynamic characteristic, the principle of superposition holds, we then know that the complex wave can be broken up into its components (such as by the Fourier series if the complex wave is periodic and hence the components are harmonically related) and the action of each studied separately. While the actual grid

voltage may even consist of transient terms, such transient analysis is usually too involved to warrant employing it; moreover, it can be shown that, if the amplifier amplifies equally sine waves of all frequencies in the range, it will react sufficiently faithfully to transient terms.* On the other hand, if the tube is used in a narrow-range amplifier, then the particular kind of complex signal voltage is known, and a special analysis of its behavior can be made. For example, if a carrier and its side bands, such as a modulated r.f. wave, are impressed upon the grid of a detector tube, then the output can be calculated by special methods beyond the powers of the graphical methods outlined here (see Chap. VI).

Hence, it may be concluded that the graphical method shown thus far is best suited for wide-range amplifiers, such as audio amplifiers, in which the distortion products are kept at a minimum; and in this case the action of the tube, when a simple sine-wave signal voltage is impressed, is representative of its action when a complex wave is impressed. We therefore shall study the behavior of the tube represented by Figs. 28 and 29 when a sinusoidal signal voltage e_g is impressed. The a.c. component of the plate current is approximately sinusoidal and is shown as i_p in Fig. 29. The method of obtaining it is obvious from the figure. If a higher value of R_L is used, the dynamic characteristic slopes less and is more nearly straight—line $B'AC'$. Note that this line passes through the same operating point A as line BAC. The a.c. component of the plate current is now i_p' and, while less in magnitude than i_p, is a somewhat more faithful current copy of e_g. Which value of R_L to use depends upon whether the tube is to deliver power into R_L or merely generate a voltage across it that is to be an amplified copy of e_g. However, in the latter case, the preceding type of load (resistance coupling) is usually preferred, and the inductance plate feed is used where maximum power output is desired from the tube.

16. Power Output of a Tube.—We shall now discuss maximum power output into R_L when the latter is shunted by an ideal choke or inductance. The operating point must first be determined, hence grid bias E_c for a given plate supply voltage E_{bb}. It has been seen that the d.c. component of the plate current is

* See, for example, GUILLEMIN, E. A., "Communication Networks," Chap. XI.

determined by the static characteristic, and so it is evident from Fig. 29 that, if a bias voltage greater than E_c is employed, then, for a given value of R_L, the dynamic characteristic is lowered (although approximately the same slope as before) and as a consequence crosses the static curve at a point to the left of A and the axis at a point to the right of B. This means that the cutoff grid voltage is reduced as the d.c. bias is increased, and vice versa.

However, after having chosen the d.c. bias, we can increase R_L, decrease the slope of the dynamic curve, and cause it to cross the e_c axis farther to the left. Hence, a proper choice of R_L can cause the cutoff grid voltage to be twice the normal bias voltage if it is so desired. In Fig. 29 it is seen that for a normal bias of E_c the dynamic curve BAC for a particular value of R_L intersects the e_c axis at B and that

$$OB = 2E_c$$

For class A operation, the grid must not be driven positive, so that it will not draw current and thus impose a load upon the source of its signal voltage, e_g. Evidently, then, the peak positive value of e_g (assumed a sine wave) must not exceed E_c. But, on the negative half cycle, the negative peak value of e_g will be E_c, which, when added to the normal negative d.c. bias E_c, makes the total grid voltage at that instant $2E_c$. Class A operation requires that this peak negative voltage does not exceed the cutoff voltage, so that excessive distortion and rectification are not produced in the plate current. Therefore E_c must be adjusted to such a value that, for a given magnitude of R_L, it is one-half the cutoff grid voltage, and this is its optimum value. Furthermore, the peak signal voltage e_g must not exceed E_c. This is the case for the dynamic curve just cited.

If, after choosing E_c, as above, R_L is now increased, dynamic curve $B'AC'$ is obtained, whose cutoff point B' is too great, so that the plate current is not varied to its furthest extent. For this value of R_L the d.c. bias should be increased until it is just half the new value of cutoff voltage. As it is increased, however, it must be noted that the cutoff voltage decreases, so that the increase in the bias voltage must take this latter factor into account.

If the bias voltage is increased, then the maximum signal voltage, or so-called "excitation," of the tube may be increased.

It may strike the reader that this may impose an excessive demand upon the signal source. This matter, however, is governed by the following practical considerations: In class A operation the grid is not driven positive and hence draws no current. It merely requires an exciting voltage, which is why the vacuum tube is often called a "potential-operated" device. Since power is the product of the voltage by the current, then, if the grid draws no current, it requires no power from the exciting source.* Therefore, from a power viewpoint, the source need not be comparable with the tube in power-handling ability. Furthermore, it is normally possible to obtain in the plate circuit an amplified copy of the signal-voltage input. Hence the input stage need not deliver a voltage that is comparable with that of the output. As a consequence, in practical design the increase in signal voltage required from the preceding stage when the bias voltage of the power stage is increased is of negligible importance both from a power and from a voltage viewpoint as compared with the increased power or voltage output possible from the power tube.

It is therefore desirable to choose R_L and E_c such that the maximum power output is obtained in the plate circuit, regardless of the value of signal input that may be required in such a case. An a.c. component of the plate current will then be obtained that will flow through R_L (while the d.c. component flows through L) and hence will be expended as power in R_L. In practice, R_L may be any load, such as a loud-speaker, that presumably imposes a practically resistive reaction upon the tube.

Before seeking the optimum values of R_L and E_c (hence e_g), it is of interest to see what the a.c. power output of the tube into R_L is for any value the latter may have, and hence the corresponding value of E_c for class A operation. It is to be noted from Fig. 29 that the plate current has a maximum value determined by point C and a minimum value given by point B and that it varies sinusoidally between these limits so that the peak value of the sine wave is half the maximum variation of the total peak plate current. It is further to be noted that in the figure point B

* In such amplifiers where the grids are driven positive, the grid power is small compared with the plate power, so that even in this case the above considerations hold approximately. Ultra-high frequencies are excluded from this discussion, however.

is chosen as the cutoff value where i_b is zero. In practice, the grid never is swung quite to cutoff for a single tube (class A) but to a value where the dynamic characteristic begins to bend markedly. This is done to avoid the excessive distortion products introduced by this portion of the curve, since the remainder of the curve is usually substantially linear. The minimum value of the plate current is therefore never quite zero but a small value, which we shall denote as $I_{\min}$. Let us also denote the maximum value of i_b as $I_{\max}$. The total variation in the current is evidently the difference between the two, and the peak value of the a.c. component is therefore

$$I_{\text{peak}} = \frac{I_{max} - I_{min}}{2} \tag{20}$$

The effective value is

$$I_{\text{eff}} = \frac{I_{\text{peak}}}{\sqrt{2}} = \frac{I_{\max} - I_{\min}}{2\sqrt{2}} \tag{21}$$

This current in flowing through R_L produces I^2R losses or, rather, power equal to

$$P_o = (I_{\text{eff}})^2 R_L = \frac{(I_{max} - I_{min})^2}{8} R_L \tag{22}$$

Equation (22) gives the a.c. power output of the tube in terms of R_L and the i_b variation. It is not necessary, however, to use the curves of Fig. 29 for $I_{\max}$ and $I_{\min}$. These can be obtained just as readily from Fig. 28, from which Fig. 29 was derived. Thus CI (Fig. 28) represents $I_{\max}$, and GH represents $I_{\min}$.

The power output can also be calculated from the fact that, if the a.c. component is practically a pure sine wave (of the same frequency as e_g), then the voltage across R_L will be practically a pure sine wave too. The maximum voltage across R_L is evidently BI (Fig. 28), and the minimum voltage BH, so that the voltage variation across R_L is their difference, or HI. But HI is also evidently equal to $OH - OI$, where OH is the maximum plate voltage (call it $E_{\max}$) and OI is the minimum plate voltage (call it $E_{\min}$). We have then that

$$HI = E_{\max} - E_{\min}$$

The peak value of the alternating voltage across R_L is evidently

$$E_{peak} = \frac{HI}{2} = \frac{E_{max} - E_{min}}{2}$$

and the effective value is

$$E_{eff} = \frac{E_{peak}}{\sqrt{2}} = \frac{E_{max} - E_{min}}{2\sqrt{2}} \tag{23}$$

The effective values of the current through and voltage across R_L are therefore known, and their product is evidently the power expended in R_L, that is, the power output of the tube. Hence, the latter may be written as

$$P_o = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8} \tag{24}$$

which is the more usual form in which the output power is written. Obviously, this equation must check Eq. (22).

It is thus seen that the graphical method makes it possible to derive the i_b-e_c or dynamic characteristic of the tube for any value of R_L from the static or i_b-e_b family of curves and also determine the power output if the dynamic characteristic is substantially linear. It can do more than this: it makes it possible to estimate the amount of distortion as well, if this is small.

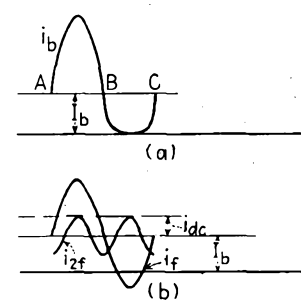


FIG. 30.—Analysis of distortion of the output wave of a tube.

17. Calculation of Second-harmonic Distortion.—If the dynamic characteristic does not depart too far from linearity, which is practically always the case for normal class A operation, then it is found, both experimentally and analytically, that the distortion consists practically entirely of the second harmonic.

A typically distorted output wave for a sinusoidal input is shown in Fig. 30A. Time AB equals time BC, and the rise in plate current i_b from the quiescent value I_b is greater than the drop in the next half cycle. This wave can be analyzed into a fundamental current i_f, a second-harmonic current i_{2f}, and a d.c. component i_{dc} in addition to the original value I_b (Fig. 30B). The latter component, i_{dc}, is due to the fact that the unequal increase and decrease in i_b take place in equal time intervals,

$AB = BC$. It is also to be noted that i_{dc} equals the peak value of i_{2f}.*

If the three components are added, it will be apparent from the figure that the positive peak of i_b (call it I_{max}) is

$$I_{max} = i_f + i_{2f} + i_{dc} + I_b \tag{25a}$$
$$I_{min} = i_b - i_f + i_{dc} + i_{2f} \tag{25b}$$

The sum and difference of Eqs. (25a) and (25b) give, respectively,

$$I_{max} + I_{min} = 2(i_{dc} + i_{2f} + I_b) \tag{26}$$
$$I_{max} - I_{min} = 2i_f \tag{27}$$

Since $i_{dc} = i_{2f}$, Eq. (26) becomes

$$I_{max} + I_{min} = 4i_{2f} + 2I_b \tag{28}$$

From Eqs. (28) and (27) the percentage second-harmonic distortion can be found.

$$\% \text{ 2d harmonic distortion} = \frac{i_{2f}}{i_f} \times 100$$
$$= \frac{(I_{max} + I_{min}) - 2I_b}{2(I_{max} - I_{min})} \times 100 \tag{29}$$

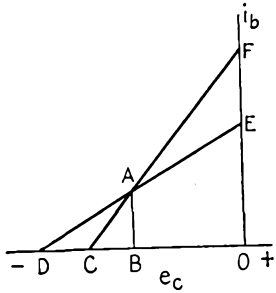


FIG. 31.—Power-output behavior for a linear tube.

18. Maximum Power Output—Linear Characteristic.—In the special case where the static characteristics are linear (which is a rather rough approximation for an actual tube) the values of E_c and R_L for maximum power output may be easily derived. Thus, suppose, as in Fig. 31, that CAF is the static characteristic, DAE is a dynamic characteristic for some value of R_L, and OB is the correct bias voltage for that value of R_L, that is OB is one-half of OD, the cutoff grid voltage. If the tube is assumed to have an amplification factor of value μ, a plate supply voltage of value E_{bb}, and an internal resistance of R_p, the equation for CAF may be written as

$$i_b = \frac{\mu e_c}{R_p} \tag{30}$$

* This is true only in resistance coupling. For choke feed, i_{dc} flows through the zero impedance of the choke and i_{2f} through R_L, so that i_{dc} exceeds the peak value of i_{2f}. The error in assuming them equal is not serious, however, and facilitates the analysis.

and for DAE as

$$i_b = \frac{\mu e_c}{R_p + R_L} \tag{31}$$

The dynamic cutoff voltage OD is given by Eq. (19) and the fact that plate voltages must be divided by μ to give their equivalent effect in the grid circuit. Thus

$$OD = \frac{-e_b}{\mu} = \frac{-E_{bb}}{\mu} - \frac{I_b R_L}{\mu} \tag{32}$$

Furthermore,

$$I_b = \frac{E_{bb} + \mu E_c}{R_p} \tag{33}$$

since the only impedance to it is the internal resistance of the tube, R_p, which, for linear characteristics, is the constant a.c. resistance of the tube. By Eqs. (32) and (33) we have

$$OD = -\frac{E_{bb}(R_p + R_L)}{\mu R_p} + \frac{E_c R_L}{R_p} \tag{34}$$

The grid bias $E_c (= OB)$ must be half of OD, so that finally we obtain

$$OB = E_c = \frac{-E_{bb}(R_p + R_L)}{\mu(2R_p + R_L)} \tag{35}$$

The peak signal voltage e_g must not exceed E_c. For this grid swing the plate current follows along the dynamic characteristic DAE, and its peak positive value is

$$I_{pm} = \frac{-\mu E_c}{R_p + R_L} = \frac{E_{bb}}{2R_p + R_L} \tag{36}$$

The power output P_o is the product of the square of the effective value of I_{pm} by R_L, so that

$$P_o = \frac{(E_{bb})^2 R_L}{2(2R_p + R_L)^2} \tag{37}$$

It can be shown by the methods of differential calculus that P_o is a maximum when

$$R_L = 2R_p \tag{38}$$

From Eq. (35) the corresponding value of the grid bias is found to be

$$E_c = -\frac{3}{4}\frac{E_{bb}}{\mu} \tag{39}$$

and the maximum power output to be

$$P_{o\max} = \frac{E_{bb}^2}{16R_p} \tag{40}$$

Equations (38) and (39) give the optimum values of R_L and E_c, respectively, for maximum output, for a tube having linear characteristics. If the grid swing is limited in an actual tube to the region where the characteristic is substantially straight, *i.e.*, where $I_{\min}$ is above the sharp bend near cutoff, results are obtained in fair agreement with the above equations. Analytically this means that the B voltage is reduced in effectiveness by an amount depending upon how much $I_{\min}$ exceeds zero. Let E_s denote the plate voltage for $e_c = 0$ and $i_b = I_{\min}$. Then the previous derivation can be modified by denoting the effective B voltage as $E_{bb} - E_s$ instead of E_{bb}, Eqs. (37), (39), and (40) become, respectively,

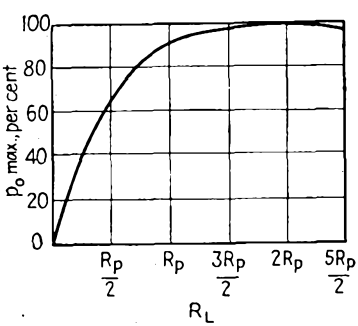


FIG. 32.—Power output vs. load resistance.

$$P_o = \frac{(E_{bb} - E_s)^2 R_L}{2(2R_p + R_L)^2} \tag{37a}$$

$$E_c = \frac{-\frac{3}{4}(E_{bb} - E_s)}{\mu} \tag{39a}$$

$$P_{o\max} = \frac{(E_{bb} - E_s)^2}{16R_p} \tag{40a}$$

while Eq. (38) remains unchanged, that is, $R_L = 2R_p$, where R_p is the internal plate resistance at the operating point. As a consequence, these optimum values are the ones usually chosen, tentatively, at least, for a single tube operating class A. It will be found, however, that R_L may be varied over quite a range without P_o changing materially. In Fig. 32, P_o is plotted against R_L, and it is to be noted that the peak of the curve is very flat. This result can be verified by substituting various values for R_L in Eq. (37). Thus, for $R_L = R_p$ instead of $2R_p$, the power output is decreased by 11.1 per cent, or approximately $\frac{1}{2}$db, an amount hardly noticeable to the ear.

The reader may be surprised to learn that P_o is a maximum when $R_L = 2R_p$, as it may possibly be his impression that maximum power output occurs when the load impedance is equal to the generator impedance. The reason for this apparent

contradiction is that, while the tube may be considered a generator developing an e.m.f., by the equivalent circuit theorem, there is nothing in this theorem that imposes any limit on the magnitude of the equivalent e.m.f. except that it be μ times the grid signal voltage. For an ordinary generator the e.m.f. is independent of the load impedance; in the case of the tube, the maximum signal voltage on the grid and hence the equivalent e.m.f. in the plate circuit vary in the manner described above with load impedance. Hence, as the latter is increased, the former is increased, but the a.c. component of the plate current through the impedance is decreased, so that these opposing considerations conspire to give maximum power output when the load impedance is twice the tube plate resistance, as has been shown above. If the grid swing is fixed, however, and thus the equivalent e.m.f. in the plate circuit, then maximum power output occurs when $R_L = R_p$.

19. Maximum Power Output, Parabolic Characteristics.—If the static characteristic is not linear but may be represented fairly accurately by the first- and second-order terms of a power series, then, by the principles of plane analytic geometry, the plot of the plate current vs. grid or plate voltage (static characteristic) is a parabola and is called a *parabolic characteristic.* The dynamic characteristics for various values of load resistance are also parabolas, flatter than that for the static characteristic.

In this case it can be shown that the maximum power output, the magnitude of the distortion products being disregarded, occurs when

$$R_L = 1.21R_p \tag{41}$$

$$E_c = -\frac{0.614E_{bb}}{\mu} \tag{42}$$

These results apply to the case where the load resistance R_L is paralleled by an ideal inductance, *i.e.*, inductance plate feed. It is interesting to note that here R_L exceeds R_p by only 21 per cent and that the grid bias is less than for the linear tube. It is evident, therefore, that for an actual tube the optimum values for R_L and E_c may be noticeably different from those given by Eqs. (38) and (39) or by Eqs. (41) and (42), respectively. Furthermore, the distortion products, especially in the latter case, may be excessive, and hence a higher value of R_L may be required to

reduce these. The actual values of R_L and E_c to be used will therefore depend upon the tube; but, once chosen, the graphical method will give a fairly accurate idea of the power output under these conditions. This will be discussed further in the following section.

20. Practical Application of Graphical Method.—The preceding derivation was for a linear tube, and optimum values of load resistance and bias were found for this type. The i_b-e_b characteristics for an actual tube, however, are seldom straight lines, and so the optimum values must be chosen with respect to a permissible amount of distortion rather than no distortion at all, as was the case above. In this discussion, a triode is assumed, and the problem resolves itself into three parts:

1. *Choice of the Quiescent Point.*—This is the point in the i_b-e_b family which determines the no-signal plate current I_b and the applied voltages. If no signal is applied to the grid, the plate current is d.c. (I_b) and flows through the choke rather than the parallel-load resistor R_L, so that there is no (a.c.) power output. The (d.c.) power input from the B supply is therefore all expended in heating the plate. This power, termed *plate dissipation*, W_{pd}, must not exceed the safe value permitted by the manufacturer of the tube.

A further consideration is the maximum B supply voltage E_{bb} that may be applied to the tube. On the negative half cycle of the grid swing, the plate voltage rises to a value

$$e_b = E_{bb} + I_b R_L$$

This value must not exceed the dielectric strength of the insulating members of the tube. This maximum value of e_b indirectly determines E_{bb}, and safe values of the latter are generally given by the manufacturer.

Since W_{pd} and E_{bb} are thus specified, I_b can be found from the relationship

$$I_b = \frac{W_{pd}}{E_{bb}} \tag{43}$$

This value of I_b is laid off vertically on the i_b-e_b family through the value of $e_b = E_{bb}$. If the actual choke (usually the primary of the output transformer) has an ohmic resistance of R_w, then the load line for R_w can be drawn through E_{bb} instead of the vertical

line. In either case, the intersection of the line with a curve of the family at a point I_b units above the e_b axis is the quiescent point, and the grid voltage corresponding to that curve is the value of grid bias, E_c, which is to be employed.

2. *Choice of* R_L.—The load line for R_L is to be drawn through the quiescent point mentioned above, at least as a first approximation. Clearly this line must not strike the e_b axis at a point to the left of the i_b-e_b curve whose grid voltage is $2E_c$, since then i_b would reach cutoff before the grid had reached its peak negative swing and the distortion would be excessive. (A peak swing of E_c is assumed, since this will just drive the grid to zero volts, with respect to the cathode, on the positive half cycle. It will thus give maximum power output for the particular value of R_L chosen under class A operating conditions.)

How close to the e_b axis the load line for R_L may cut the $2E_c$ curve depends upon the next factor to be considered, *viz.*, the permissible percentage distortion. The latter is assumed to be all second harmonic, which is not far from the truth for a triode. A typical value is 5 per cent; this means that the second-harmonic current or voltage component is 5 per cent of the corresponding fundamental component. The power ratio is therefore as the square, or 1:400. This may also be expressed as the second harmonic is 26 db down on the fundamental, since

$$10 \log \tfrac{1}{400} = -26 \text{ db}$$

A tentative load line is drawn subject to the obvious precaution mentioned previously, and $I_{\max}$ and $I_{\min}$ are noted. The percentage second-harmonic distortion can then be calculated from Eq. (29). If the value is more than is permissible, it can be reduced by tilting the load line up, although the line must still pass through the quiescent point. After a few trials, the correct tilt for the load line will be found. It must be borne in mind, however, that the position of the load line will have to be corrected owing to the self-rectification of the tube, and this in turn will change the percentage of distortion. This correction will be given in greater detail in Sec. 27.

There is a method of finding at once the proper slope of the load line by means of a "distortion rule"; but since this does not take into account the shifting of the load line due to self-rectification,

with the attendant change in the percentage distortion, it will not be described here.

3. *Calculation of Power Output.*—Once a suitable slope for the load line has been found, the corresponding value of R_L can be calculated, as by choosing any two points on it and dividing the difference in voltage ordinates by the difference in current ordinates. The power output is then given by Eq. (22),

$$P_o = \frac{(I_{\max} - I_{\min})^2 R_L}{8} \tag{22}$$

or, alternatively, by Eq. (24).

Several points of interest are to be noted. If the B supply voltage can be safely increased beyond the value chosen, then I_b [by Eq. (43)] will decrease for a fixed W_{pd}. This, in turn, means that the quiescent point moves to the right and downward; that is, I_b is less, and E_c is greater (to obtain the lower I_b with increased E_{bb}). As a result, R_L must be increased in order that cutoff be prolonged to twice this greater value of bias, particularly since the quiescent point has moved down as well as to the right.

The power output will at first increase because it tends to vary as E_{bb}^2 [provided that the bias can be adjusted close to the optimum value given approximately by Eq. (39)] and only slowly to decrease as R_L is increased. However, as E_{bb} is increased, E_c must be decreased to values considerably more negative than the corresponding optimum values; and since R_L must also be considerably increased, P_o begins to decrease. At the same time, the signal voltage required to obtain P_o increases with E_c.

The result is that, if a tube is correctly designed, it will be found that W_{pd}, R_p, μ, and the maximum value of E_{bb} are so coordinated that the optimum value of R_L is not much in excess of $2R_p$ for a nominal value of permissible distortion, say, 5 per cent. For example, a type 45 tube has a maximum recommended value of 275 volts for E_{bb}, a permissible plate dissipation of 10 watts, hence a value of I_b of about 36 ma., to which corresponds a bias of $E_c = -56$ volts. If the tube were linear, then the optimum value of E_c would be -58.9 volts for a μ of 3.5 [Eq. (39)]. The value of R_p at the quiescent point is 1,700 ohms, and the value of R_L for 5 per cent distortion is 4,600 ohms, which is not too far in excess of $2R_p = 3{,}400$ ohms, the optimum value for a linear tube. The power output is 2 watts.

If E_{bb} were materially increased above this value, R_L would have to be considerably greater than $2R_L$, the bias and hence maximum grid swing would be increased markedly, but the power output would be but little greater.

Another point to note is that the plate dissipation is greater for no signal than for maximum e_g permissible under class A operation. In short, the tube runs cooler when it is delivering a.c. power than when it is not. While this is contrary to the action of ordinary sources of power, it need occasion no surprise, since a vacuum tube is not a source of energy in the sense of being the seat of an active a.c. e.m.f., but rather (as mentioned previously) a variable resistor capable of converting some of the d.c. energy supplied to it by the B supply into a.c. output energy. When its resistance is not varied (e_c maintained constant at the value E_c), then it dissipates the d.c. energy as heat.

21. Tetrode and Pentode Tubes.—It is shown in Sec. 16 that the power output is

$$P_o = \frac{(E_{\max} - E_{\min})(I_{\max} - I_{\min})}{8} \tag{24}$$

For a linear tube, $I_{\min}$ can be zero, and evidently $I_{\max}$ is then $2I_b$, the d.c. component of the plate current. For a tube that is not linear, $I_{\min}$ is greater than zero, and hence $I_{\max}$ is less than $2I_b$. Since I_b is fixed by plate-dissipation considerations after E_{bb} has been chosen, it is evident that $I_{\max}$ is indirectly limited by these two factors, also. It will likewise be evident that, since the load line for R_L must pass through the $e_c = 0$ curve of the plate family at the value of $I_{\max}$, through the $e_c = E_c$ curve at the value I_b, and through the $e_c = 2E_c$ curve at the value $I_{\min}$, $E_{\max}$ is determined by the intersection of the load line with the $2E_c$ curve and $E_{\min}$ by the intersection with the $e_c = 0$ curve. Thus, all factors entering into the power-output formula are determined by the tube family of characteristics and the position of the quiescent point on the tube family, *i.e.*, the permissible plate dissipation and choice of E_{bb}.

From the inherent characteristics of a vacuum tube, it is evident that, for a given I_b, $I_{\max}$ cannot exceed $2I_b$, since, for equal current excursions about I_b, this would require a negative value for $I_{\min}$. The quantity $I_{\max} - I_{\min}$ may therefore be regarded as relatively fixed. If it is desired to increase P_o, then such increase

must be sought for in the quantity $E_{max} - E_{min}$. The limits here (like those for the current) are $E_{min} = 0$ and $E_{max} = 2E_{bb}$, but it will be found, as by examination of Fig. 28, the E_{min} is considerably greater than zero for a triode, owing to the appreciable slope of the $e_c = 0$ curve. Since, ignoring distortion,

$$E_{max} - E_{bb} = E_{bb} - E_{min}$$

it is evident that E_{max} is considerably less than $2E_{bb}$. Improvement in tube performance can therefore be obtained if E_{min} can be brought closer to zero.

A hypothetical tube whose characteristics would permit such increase in output power is shown in Fig. 33. It is not necessary that the i_b-e_b curves continue into the second quadrant (shown by broken lines in the figure). Instead, they can coalesce into a single line making a very acute angle with the i_b axis, as shown by the solid line in the figure, and yet the power output will be the same as for the former case. In this way, we do not depart too far from an actual tube in that we do not require plate current to flow when the plate voltage is negative.

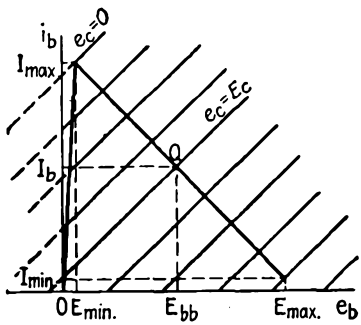


FIG. 33.—Hypothetical tube having high power output.

It is evident that, as I_{max} is shifted to the left (as in the case of this hypothetical tube), R_L will have to be increased, as compared with a linear tube whose plate family has the same slope as those of this tube but whose curve corresponding to $e_c = 0$ passes through the origin. This means that the load line for R_L will swing counterclockwise about Q, so that E_{max} will shift to the right, as well as E_{min} to the left. As a result, $E_{max} - E_{min}$ will increase, as will also, therefore, the power output.

In practice, much the same effect may be obtained by introducing into the tube structure a fourth electrode, called a *screen grid*, between the control grid and the plate and imposing a positive potential, usually less than that on the plate, upon this electrode.

The tube characteristics shown in Fig. 34 are considerably modified from those shown in Fig. 33, but the effect upon the output power is much the same. While the screen grid was possi-

bly introduced for another purpose, *viz.*, that of eliminating coupling between the output and input circuits through the plate-to-control-grid capacitance, we shall concern ourselves here with its effect upon output power, since this is a matter more pertinent to graphical analysis.

It will be noted that for the particular tetrode shown in Fig. 34 the i_b-e_b curves show only a tendency to coalesce to the left and that to the right they are all nearly horizontal. At a plate voltage denoted by a vertical line and marked as the limit of a region labeled "values unstable," the curves show a sudden change in

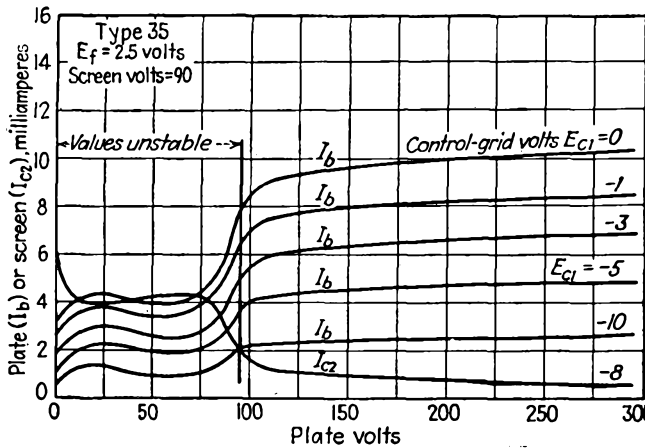


FIG. 34.—Actual tetrode characteristic.

curvature; this break has been called a "knee" in the characteristic. Obviously this knee should be as sharp and as close to the i_b axis as possible, in order that increased power output may be obtained. (This is the case for Fig. 33.)

In an actual tetrode the knee is considerably rounded, and the curves to the right of the knee are more nearly horizontal. The reason for the latter fact will be discussed first. The screen grid has much more effect upon the space current than the plate because of its closer proximity to the cathode. At the same time, the screen draws normally very little current because its area of interception of the cathode current is small, and thus most of the electrons pass through its interstices en route to the plate. The screen, however, tends to shield the cathode or space current from electrostatic action by the plate; and if the shielding were perfect,

the plate current would be independent of the plate voltage and due solely to the screen voltage. In this case, the curves would be truly horizontal, and the R_p of the tube would be infinite. In actual tubes, the electrostatic field of the plate penetrates into the cathode region through the interstices of the screen grid and thus affects the plate current to some extent. As a consequence, the curves rise slowly with plate voltage and thus produce a finite, though very high, value of R_p.

This same shielding action also produces a tube having a very high μ, where μ is defined as

$$\mu = \frac{de_p}{de_c} \quad (i_b \text{ and } e_{sg} \text{ maintained const.})$$

A value for de_p of 1,000 volts may just balance the effect of 1 volt on the control grid; *i.e.*, μ may be 1,000 or higher. The ratio $\mu/R_p = G_m$, however, is about the same as for a triode.

The various factors affecting rounding of the knee of each characteristic curve will now be discussed. The electrons projected by the screen grid through its interstices toward the plate enter a region where they are decelerated if the plate voltage is less than the screen voltage, which is usually the case during the positive half cycle of the control-grid swing. As a result of this, they accumulate in this region in the form of a space-charge cloud and tend to repel fresh oncoming electrons.

Many of the latter are thus repelled back to the screen, increasing the screen current at the expense of the plate current. Some are even repelled (through the interstices of the screen) all the way back to the cathode, thus even reducing the cathode current. The lower e_b, the more marked these effects.

However, another factor appears that is even more potent in rounding the knee, and this is secondary emission from the plate. If primary electrons of the cathode stream strike the plate with a velocity corresponding to twenty volts or more, they cause secondary electrons to be emitted from the plate in a manner akin to the action of heat in producing thermionic emission. In addition, some of the primary electrons are reflected from the plate back toward the cathode. The screen grid, being at a considerably higher potential, captures these electrons and thus augments its current at the expense of that flowing to the plate. In extreme cases, the plate-current loss of secondaries

may be so severe that the external plate-current flow may even reverse. The space-charge cloud between the screen and plate, however, tends to prevent this loss in plate current, since the secondaries are emitted at relatively low velocities corresponding to but a few volts and hence are readily repelled by the cloud back to the plate. They may, however, skirt around the cloud and thus reach the screen.

The characteristic curves shown in Fig. 34 are based on the premise that the screen voltage does not vary. In the more general case, where this is not so, the plot would involve four dimensions. However, if the screen voltage is maintained constant and then the control-grid potential is set at various values, i_b-e_b characteristics can be obtained that require only a plane for their portrayal. This is the case for Fig. 34, where the screen is shown maintained at a fixed potential of 90 volts.

As mentioned before, the tetrode was developed to prevent interaction between the output and input circuits, particularly at radio frequencies, for such interaction could result in regeneration and a tendency for the tube to oscillate. This matter can be well handled by analytical methods. The ability of a tetrode tube over that of a triode to give increased output at a given B supply voltage which has been studied here by graphical methods but which is not sufficiently emphasized in most texts, is another important feature of this tube.

The effects of secondary emission by the plate can be reduced in several ways. One is to treat the surface of the plate with a material, such as carbon, which is a poor secondary emitter. (Secondary emission depends upon the work function, just as does thermionic emission.) Another method is to rib the surface of the plate in such a manner as to produce strong local electrostatic fields in the crevices in such direction as to tend to prevent the loss of secondaries.

A third, and most important method, is to introduce a fifth electrode, called the *suppressor grid*, between the screen and plate, which makes the tube a pentode. The suppressor is normally maintained at cathode potential. Under these conditions, it does not interfere materially with the action of the screen grid in shifting the characteristic curves, as mentioned at the beginning of this section. Since, however, it is relatively very negative with respect to the plate (as well as the screen grid), it

repels secondaries from either direction back to the electrode of their origin, and particularly is effective in moving the knee in the characteristics to the left, *i.e.*, to lower plate potentials.

A fourth method is that embodied in the beam power tubes of the tetrode type. Here, by proper design, an equipotential surface of zero gradient with respect to the cathode is established in the region between the screen and plate. That such a surface is possible is evident from the consideration that in this region between two positive electrodes must be some points constituting, in general, a surface, where the electrostatic forces of the two electrodes upon a charge are balanced. This surface may vary in position depending, in part, upon the space-charge effects of the electron stream in this region, but by proper design its position over wide ranges of plate current can be maintained within a suitable interval of the region. Such an equipotential surface can function as a suppressor grid with even greater uniformity than a physical grid. The establishment of such a zero gradient surface is facilitated by the focusing of the space current in the form of beams through the interstices of the control grid. By aligning the helical coils of the screen grid with those of the control grid and by the use of "beam-confining" plates at cathode potential, the electron stream can be focused away from the screen-grid wires even if e_b is considerably lower than e_{sg}, so that the diversion of space current to the screen is minimized. At the same time, this focusing produces beams that overlap in the region between the plate and screen grid to form a surface of uniform electron density having zero gradient, and this surface acts as a uniform suppressor grid. The beam-confining plates prevent plate secondaries from flowing around the sides of the beams back to the screen grid or beyond (to the cathode).

In the ordinary pentode, the suppressor grid has a nonuniform action upon the electrons in the region between the screen grid and plate. As a consequence, the reduction of plate current as e_b is reduced is more gradual, particularly as the component of velocity normal to the plate varies from one electron to another, so that some are stopped and reversed in direction more easily than others. In the beam power tube the electron velocities are more uniform in direction. Hence, as the plate voltage is reduced, a fairly well defined value is reached at which the plate current suddenly begins to decrease; in short, the knee is

very sharp and well defined, which, in turn, means that E_{min} is quite definite in value and lower than it would be for the ordinary pentode tube. This, in turn, results in greater power output, particularly for a permissible amount of distortion.[2]

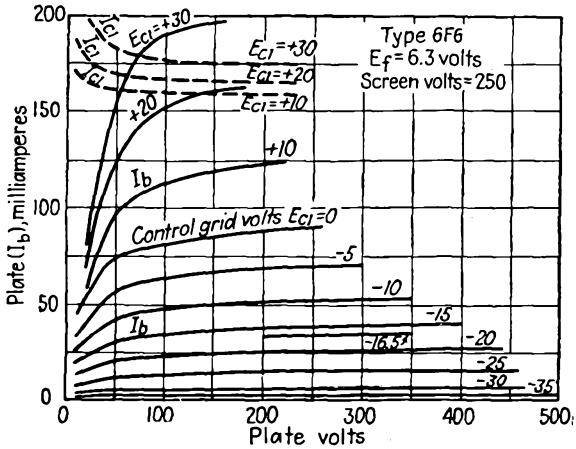


FIG. 35.—Pentode characteristic.

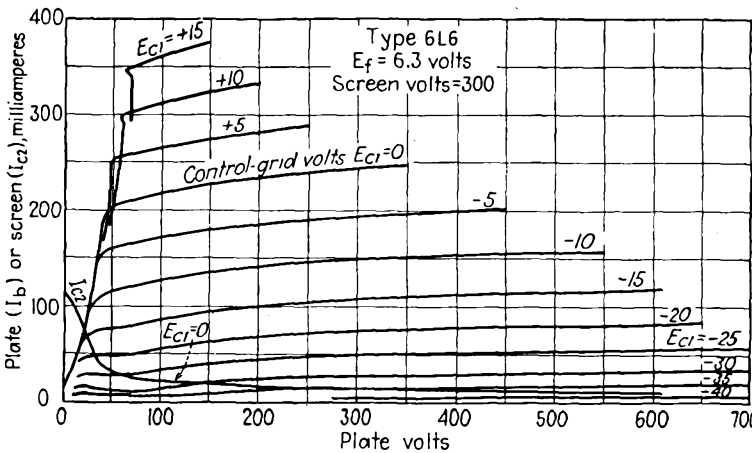


FIG. 36.—Beam power-tube characteristic.

In Figs. 35 and 36 are shown the characteristics for a pentode and beam power tube, respectively.

Distortion will obviously be excessive if the load line for R_L cuts through the i_b-e_b curves below the knee. The output wave shape will have a flattened peak on the positive half cycle of the

grid swing, or even possibly a subsidiary minimum, whereas the other half cycle may be a relatively faithful copy of the grid signal voltage. In any event, excessive distortion is indicated.

This distortion may be minimized in several ways:

1. By reducing R_L. This reduces the plate-voltage excursions to a region to the right of the knee in the characteristics. This is the mode of operation best suited to the development of maximum power output with permissible percentage distortion or for the extreme wide-band amplifiers, such as video amplifiers, where the load impedance is inherently low.

2. By reducing the grid swing e_g. This is usually feasible only in the low level or front end of an amplifier system; *i.e.*, R_L may be maintained at a high value, with resultant high circuit gain, only if the tube is relatively "oversize" compared with the grid swing it is called upon to handle.

3. By increasing E_{bb} to compensate for the large voltage drop in R_L. In this way, large voltage gains and levels can be had simultaneously. By increasing E_{bb} to 1,000 volts and adjusting E_c accordingly for a 57 tube, output voltages across a load resistor of 250,000 ohms have been obtained sufficient to swing the grid of an 845 tube to maximum power output. This method is rather obvious and not ordinarily feasible on account of the high power supply voltage required.

4. By reducing the screen potential. This is a common method employed in screen-grid resistance-coupled audio amplifiers and permits high values of R_L to be used, with the result of high circuit gain. On the other hand, the reduction in screen potential lowers the g_m of the tube, and this tends to reduce the circuit gain. The reduction in g_m is a result of the complex reactions that take place in a multigrid tube, as can also be noted from the shape of the i_b-e_b characteristics. Of interest in this connection is the unequal spacing of the i_b-e_b curves for equal increments of e_c. The principal effect of lowering e_{sg} is the shifting of the knee to the left, so that for relatively large grid swings and high values of R_L the plate voltage can dip to fairly low values without having the space current or plate secondaries diverted to the screen grid, *i.e.*, without reaching the knee of the characteristics and thus producing excessive distortion.

It might appear that choke feed would obviate the need for reducing e_{sg} under these conditions. However, a choke having

sufficient inductance (hence, reactance at the lowest frequency to be amplified) relative to the high value of R_p encountered is rather difficult to build, and the above condition is necessary for a practical choke to approximate an ideal choke premised for such operation. Hence, resistance coupling is normally employed, and the screen potential is reduced to about half the value possible if an ideal choke were available. The gain possible is much higher than that obtained by an ordinary triode but is far less than the tube μ, because this could be approached only if R_L were many times R_p. For power pentodes, however, R_L is relatively low, so that, by Thévenin's theorem, the choke feed sees a fairly low apparent source impedance. In this case, choke feed

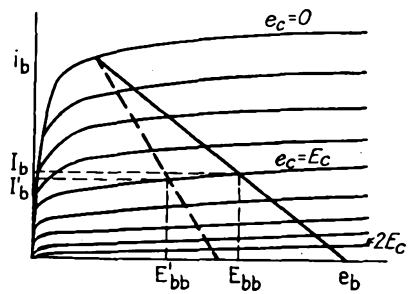


FIG. 37.—Graphical constructions for a tetrode or pentode tube.

is practicable and appears usually in the form of an output transformer.

22. Graphical Constructions for Tetrodes and Pentodes.—The graphical constructions for a tetrode or pentode are exactly the same as for a triode: the quiescent point is first located, and then the load line for R_L is drawn, as a first approximation, through the quiescent point. It must not reach cutoff for a grid swing equal to the value $e_c = 2E_c$, just as in the case of the triode. However, in the case of a triode, a grid swing in a positive direction to $e_c = 0$ occasions no difficulty, whereas, for a tetrode or pentode, such a swing for the load line required to prevent premature cutoff may carry the path of operation on the positive half cycle into the knee of the characteristic. In short, excessive distortion of either half of the cycle of i_p is possible.

In Fig. 37 are shown the i_b-e_b characteristics for a power pentode together with a load line (solid line) for a suitable R_L

(choke feed). For class A operation, the bias must be midway between e_c for cutoff and $e_c = 0$. To this value of E_c will correspond a certain value of E_{bb} and I_b and hence a certain value of $W_{pd} = I_b E_{bb}$ at no signal. If W_{pd} is not excessive, the value of R_L is satisfactory, and the power output can be calculated, by Eq. (22), just as in the case of a triode.

But if W_{pd} is excessive, a trial load line (broken line) corresponding to a lower value of R_L can be chosen. The cutoff point is not affected very materially because of the small effect of e_b upon i_b; but, for the same bias as before, the positive grid swing can be made to avoid the knee of the characteristics, and, as will also be noted, E_{bb} is reduced, so that W_{pd} is decreased too.

It will be apparent from Fig. 37 that cutoff is poorly defined, since the i_b-e_b curves for very negative values of e_c are very closely spaced. Consequently, only that region of operation where the curves have more open spacing is generally employed, and the calculation of distortion for a given R_L becomes much more involved than for a triode. This will be discussed in the next section.

The graphical constructions very clearly illustrate the considerations presented in this section. For example, it may be noted how R_L is limited to values far less than the R_p of the tube, for practicable operation and minimum distortion. The curves of the figure can apply to resistance coupling instead of choke feed, if E_{bb} is assumed to be the value where R_L cuts the e_b axis. Thus, Fig. 37 illustrates the conditions for voltage amplification as well as for power output.

23. Distortion Products in a Pentode.—Figure 37 indicates that the relation between i_b and e_b is quite involved when the screen is maintained at a fixed potential and that the curves do not obey a simple law such as the $\frac{3}{2}$ power relationship of Child. As a consequence, the distortion products involve more than the second harmonic, and an appreciable third harmonic is also generated. The amounts of second and third harmonics produced depend upon the value of load resistance chosen. A safe rule is to avoid having the load line cut the curves to the left of the knee of the characteristic. In the majority of circuits, the tubes are used in a push-pull or balanced amplifier circuit, and in this case the even harmonics are balanced out. In such cases the load line may be chosen to give a minimum of third harmonic,

as then the tubes in push-pull will generate a minimum of distortion products.

To examine this in more detail, it will be evident that the simple formula for second-harmonic distortion in a triode will hardly apply here, at least to the odd-harmonic terms, and a more extended analysis must be used. This may be done graphically by plotting the signal voltage (assumed a sine wave) against time and then plotting the corresponding current function of time. The method has been outlined in Sec. 13 of this chapter (see Fig. 26*A*, *B*, and *C*). The resultant current wave can then be analyzed by means of the Fourier analysis in terms of its fundamental and harmonic components. As a final check, the tube may be set up in an actual circuit, and an experimental run made, using either an oscillograph and the Fourier analysis or a harmonic analyzer. Several tubes should be tried, and the results averaged to minimize the discrepancies due to unavoidable variations between individual tubes.

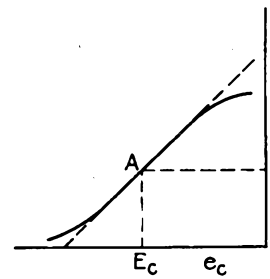


FIG. 38.—Dynamic characteristic for a pentode tube.

Another method, which is possibly more illuminating, is to find the dynamic characteristic graphically and express it then by a power series having a sufficient number of terms to fit the curve to the degree of accuracy desired. If the grid voltage is then assumed a sine wave, the distortion products in the plate current can be calculated. It will be found that the second-order term contributes to the d.c. component and the second harmonic.[3] It will be shown that an odd term—in particular, the third-order term—contributes to the third harmonic and the fundamental. Thus, the dynamic characteristic of a pentode for a somewhat high value of R_L is of the shape shown in Fig. 38. It has a kind of symmetry in that the top and bottom of the curve bend toward the horizontal, so that the positive and negative lobes of the plate a.c. component will be flattened, although not necessarily to the same degree, for a sinusoidal grid signal voltage. This, in turn, indicates odd harmonics. The power series representing this curve will require at least three terms to fit it to any degree of accuracy.

$$i_p = k_1 e_g + k_2 e_g^2 + k_3 e_g^3 \tag{44}$$

where i_p is the variation of the plate current about the d.c. component and e_g is the variation in grid voltage about the normal bias voltage E_c. In short, Eq. (44) expresses the variation of the plate current with grid voltage about the operating point A for some load resistance R_L. The second-order term $k_2e_g^2$ will contribute to the second-harmonic and d.c. components if

$$e_g = E \sin \omega t \tag{45}$$

The third-order term $k_3e_g^3$, when the value for e_g from Eq. (45) is substituted and expanded trigonometrically, becomes

$$k_3e_g^3 = k_3E^3 \sin^3 \omega t = k_3E^3(\tfrac{3}{4} \sin \omega t - \tfrac{1}{4} \sin 3\omega t) \tag{46}$$

It is thus seen that the third-order term contributes three times as much to the fundamental as it does to the third harmonic. Furthermore, for the shape shown for the dynamic characteristic, k_3 is inherently negative, so that the third-order contribution to the fundamental is subtractive from the first-order term $k_1E \sin \omega t$ and hence tends to reduce the fundamental power output. This is to be expected, since the curve shows that $I_{max} - I_{min}$ is less than for a linear characteristic, as shown by the broken line (Fig. 38).

If a lower value of R_L is chosen, the dynamic characteristic will have the curvature of the lower part of the curve shown in Fig. 38 but will tend to remain concave upward, rather than having the S-shaped characteristic shown in the figure. This is due to the fact that the steeper load line cuts the i_b-e_b curves in a region where they continually draw farther apart as the grid is made less negative.

The simplest power series approximating this characteristic at all satisfactorily is a parabola, *i.e.*, one having the form

$$i_p = k_1e_g + k_2e_g^2 \tag{47}$$

Such a characteristic gives rise to second-harmonic distortion for sinusoidal excitation. It is thus evident that, as R_L is increased from a low value, the second-harmonic distortion tends to decrease from some high value, possibly down to zero. As the load line begins to cut into the knee of the characteristic, third-harmonic distortion begins to appear. Further increase in R_L usually tends to augment this and also to produce increasing amounts of second-harmonic distortion, as well as a rich crop of higher

harmonics. The increase in second-harmonic distortion from the minimum (possibly zero value) results from asymmetry in the dynamic characteristic about the operating point.

It is therefore difficult to choose a load line so that a maximum power output can be obtained with a predetermined or permissible percentage distortion, since the latter cannot even approximately be assumed to be all second, as in the case of a triode. For a single-ended tube (not push-pull) a value of R_L is chosen, after a series of trials, such that no harmonic exceeds a certain permissible value. One rule is to specify that the square root of the sum of the squares of the percentages of the various harmonics does not exceed a certain value, such as 7 per cent. As mentioned previously in this section, for balanced-amplifier operation a value of R_L that gives a minimum of third harmonic, though considerable second, is usually chosen, since no even harmonics will appear in the output in any case. The 6L6 beam power tube, on the other hand, is so designed as to produce a minimum of third harmonic for a reasonable value of R_L, since an important application is in push-pull circuits.

It has therefore been seen that the main effect of the symmetrical curvature of the dynamic characteristic is the reduction in the fundamental output and that a secondary effect is the production of third-harmonic distortion. While the main effect reduces the plate efficiency of the tube, the latter is still higher than that of a triode. The third-harmonic distortion, as well as the second (produced by the second-order term), is, however, appreciably higher than that for a triode; nevertheless, the pentode has established a place for itself as a power-output tube. Owing to its inherently high μ, it has a high power sensitivity and can operate directly out of the power detector in a radio receiver without overloading the latter, yet with sufficient power output to the loud-speaker. (The power sensitivity is defined as the quotient of the power output by the required grid signal-voltage input.)

24. Effect of Variation in Load Impedance.—The discussion in Sec. 21 indicates that, in the process of increasing the power output from a tube by introducing a screen grid, R_L can be increased somewhat but that in actual tubes the R_p is increased enormously owing to the shielding action of the screen between the plate and cathode. This is on the basis of maximum power output with an

acceptable percentage distortion. Unlike the triode, however, it was shown that R_L cannot arbitrarily take on higher values without the distortion being increased inordinately. Hence, R_L is limited to values considerably less than the R_p of the tube.

If R_L is increased, the power output may at first increase owing to the mismatch in impedance being less, but the distortion products increase rapidly, since the load line slopes too far to the left (see, for example, Fig. 37) and cuts the i_b-e_b curves below the knee. This is also true if the load impedance is partly inductive as well as resistive, although the load line for this kind of impedance has not as yet been derived. The result is that the load impedance must not vary materially from its permissible value

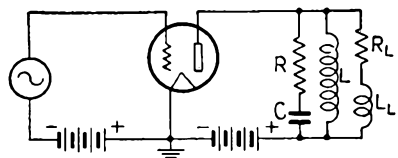


FIG. 39.—Loud-speaker compensating circuit, pentode output tube.

with change in frequency or any other cause. Where it is resistive in nature, no such difficulty will be experienced. Usually, however, the pentode feeds a loud-speaker, and the latter has some inductance as well as resistance. Its impedance will increase with frequency, and hence the distortion products may be excessive at the higher frequencies. For example, an electrodynamic loud-speaker may have a voice-coil impedance of 10 ohms at 30 cycles and 50 ohms at 5,000 cycles. It is evident that the loud-speaker will reflect back to the tube an excessive value of load impedance at the latter frequency and hence may produce excessive distortion products. It must be noted, however, that the second harmonic is 10,000 cycles, which is the limit of the usual audio range; but if a complex signal voltage having an h.f. and l.f. component is impressed, a different beat frequency in the audio range may be produced.

This effect may be compensated by the introduction of a resistance-capacity network in the plate circuit of the tube (Fig. 39). The condenser C draws a leading current through resistor R, which limits the magnitude of this leading current and hence the compensation for the lagging current drawn by the useful load, represented by a resistance R_L and an inductance L_L

in series. For perfect compensation (assuming ideal choke for L) the elements must have the following relation:

$$\sqrt{\frac{L_L}{C}} = R = R_L \tag{48}$$

In this case, the entire load would appear resistive at all frequencies and of value $R = R_L$. It would be found, however, that the l.f. components of the plate current would flow mainly through R_L and the h.f. components through R; and since R_L represents a loud-speaker or similar load and R an ordinary resistance, it can be seen that the acoustic output would be limited to the lower frequencies, while the higher frequencies would be dissipated solely as heat energy in R. For this reason, R is as large as possible consistent with maintaining the load impedance below a value that will give excessive distortion products and, together with C, is chosen of such value as to compensate sufficiently for the rise in impedance of R_L and L_L at the higher frequency to be reproduced. This is largely a matter of "cut and try," although a mathematical analysis is possible. The main difficulty is in determining R_L and L_L, which are composed of static and motional components.

Where two loud-speakers are used as h.f. and l.f. units, the compensation can be made more perfect, in that R can represent the impedance of the h.f. unit (which is usually practically resistive in nature); R_L, the resistive component of the l.f. unit; L_L, its inductance plus any additional amount placed in series with it; and C, the condenser in series with the h.f. unit. Equation (48) can then be used, although it gives no absolute values for C and L_L, but only their relation to R and R_L. Their absolute values can be determined, however, by the manner in which they are to divide the h.f. and l.f. components between themselves. This, however, depends upon the individual characteristics of the two loud-speaker units.

25. Plate Efficiency.—In analyzing the operation of power amplifier tubes, it has been found useful to rate them on the basis of plate efficiency. This is defined as the ratio of maximum a.c. power output to the d.c. input power supplied to the tube by the B supply. This can be formulated for a linear tube in terms of R_L and R_p by the use of Eqs. (33), (35), and (37). These equations give the values of I_b, E_c, and P_o in terms of μ, E_{bb}, R_p, and

R_L, so that for a given tube (having a certain μ and R_p) and a given E_{bb} the efficiency can be found for any value of R_L. (Note from the above equations that R_L determines E_c and hence I_b.) Thus, the power output is

$$P_o = \frac{E_{bb}^2 R_L}{2(2R_p + R_L)^2} \tag{37}$$

From Eqs. (33) and (35) we obtain

$$I_b = \frac{E_{bb}^2}{2R_p + R_L} \tag{48}$$

The d.c. input power is

$$E_{bb}I_b = \frac{E_{bb}^2}{2R_p + R_L} \tag{49}$$

and the plate efficiency

$$\text{Eff.}_p = \frac{E_{bb}^2 RL/2(2R_p + R_L)^2}{E_{bb}^2/2R_p + R_L} = \frac{R_L}{2(2R_p + R_L)} \tag{50}$$

This is a maximum when R_L is infinite in value, in which case

$$\text{Eff.}_{p\,\max} = 50 \text{ per cent}$$

This is the highest efficiency obtainable for a tube operating class A, but higher efficiencies are possible under class B and C operation. In these modes, the grid bias is normally set at cutoff and beyond, respectively, so that the plate current flows for but a portion of the cycle. The average value of the current or d.c. component is therefore small (hence, the power input) so that the plate efficiency is consequently high. Such methods of operation produce high distortion products, which can be filtered out in the case of a narrow frequency-range amplifier but are objectionable in the case of a broad frequency-range amplifier. However, where the latter is of the balanced-amplifier type, it may be operated class B without objectionable distortion, owing to the balancing out of even-order terms in the power series expressing the relation between a.c. signal input and a.c. output (see Chap. V, Balanced Amplifiers).

In the case of a linear triode operating under optimum conditions, R_L equals $2R_p$, and

$$\text{Eff.}_p = 25 \text{ per cent}$$

by Eq. (50). This means that, under maximum grid excitation, 75 per cent of the B power is still consumed in heating the plate.

In the case of a pentode the a.c. output for a given E_{bb} and I_b (hence B power) is greater than for the triode, owing to the greater value for $(E_{max} - E_{min})$ so that the plate efficiency can approach more nearly the theoretical maximum of 50 per cent. As a practical instance, the plate efficiency for a 45 triode is but 18.8 per cent and, for the 2A5 pentode, 35.3 per cent. The reason these are less than the theoretical values given previously is that the tubes are not linear, and hence the grid swing is restricted to that portion of the dynamic characteristic which is more nearly so, *i.e.*, that portion above the cutoff point and in addition, in the case of the pentode, to the right of the knee of the characteristics.

26. Space-charge and Coplanar Grid Tubes.—An interesting variation in the method of operation of the tetrode and pentode is to impress a positive potential upon the first grid and use the second as a control grid. The tube is then known as a *space-charge tube*. The positive first grid helps overcome the space charge and causes a large space current to flow, of which it diverts only a small portion to itself. The second grid, being negatively biased, slows up the electrons composing this space current, so that they form a cloud just before it. This cloud of electrons forms a second virtual cathode, of large surface, close to the control grid. As a result of the large surface, the plate resistance is low; as a result of the proximity of the cloud to the control grid, the μ is high; and thus high voltage amplification, and also large power output, are obtained at a low plate voltage. These tubes have been used in Europe to some extent, because the alternating power-supply voltages vary so much from one city to another that B batteries are employed to a much greater extent than in this country.

Another type of tetrode is the coplanar grid tube. This has two grids wound in the same plane; one is used as a space-charge grid by biasing it positively, and the other as a control grid. This type of tube is essentially a space-charge tube and has a high power output for a given B voltage. Unlike the screen-grid tetrode, it is not appreciably troubled by secondary emission from the plate, because the negatively biased control grid counteracts the attraction of the other positive grid for the plate secondaries. As a consequence, the plate characteristics resemble those of a pentode. The tube has not as yet been

marketed, although known for years, but shows considerable promise as a power amplifier.

27. Effects of Rectification in Plate Circuit.—According to Sec. 9 of this chapter, the dynamic characteristic can be represented by an equation of the form

$$i_p = k_1 e_g + k_2 e_g^2 + k_3 e_g^3 + \cdots$$

where i_p is the incremental component of the plate current i_b due to the incremental component e_g of the grid voltage e_c. In the absence of e_g, $i_p = 0$, and $i_b = I_b$, the initial direct plate current due to the direct grid-bias and plate voltages. If e_g is of the form

$$e_g = E_g \sin \omega t$$

and this is substituted in the previous equation, a term

$$k_2 E_g^2 \sin^2 \omega t$$

is obtained, which can be expanded into $E_g^2/2$, an additional d.c. component over and above I_b, as well as a second-harmonic component of the form $(E_g^2/2) \sin 2\omega t$. Higher even-order terms will also contribute positive or negative amounts to these two components. The additional d.c. component $E_g^2/2$, due to the nonlinearity of the dynamic characteristic, is attributed to a property of the tube known as *self-rectification*. This effect may be prorated as equivalent to a change in bias of the tube plus a suitably transformed a.c. grid-signal swing. In the case of inductive plate feed, it will also appear as an apparent increase in the power pack voltage E_{bb}. In any event, the operating point (about which the plate current varies to produce the various a.c. components) shifts away from the quiescent point, and a correction to the load line may have to be made.

If the load is a pure resistance (to both alternating and direct current), then it affects all components alike; *i.e.*, the load line for all components is the same. Thus, in Fig. 40, if A is the quiescent point and the grid voltage swings to $e_c = 0$ and $e_c = 2E_c$, then the plate current varies from A to B and to C, respectively. If the grid swing is sinusoidal in nature, then the plate current will have the wave shape shown to the right as D. This can be analyzed. Assume that it consists of an additional amount of d.c. component GI, a fundamental, and second

harmonic as shown at E. The additional d.c. component produces a voltage drop in the load resistance R_L of value FH, so that the fundamental and second harmonics (a.c. components) vary about the point G. Point G is called the *operating point* and corresponds to a lower bias voltage Ec'. Thus, although

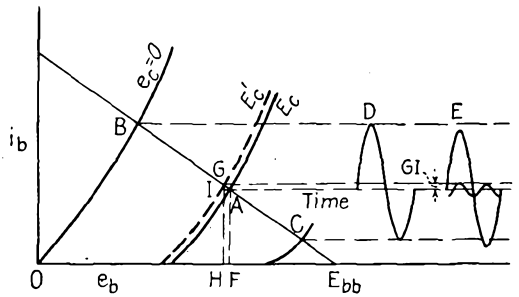


FIG. 40.—Shifting of operating point due to self-rectification, plate characteristic.

the operating point does not coincide with the quiescent point, it nevertheless lies on the load line BC.

The above can also be seen from the i_b-e_c curves (Fig. 41). The lettering corresponds with that for Fig. 39. The additional d.c. component that flows is GI, and for this the dynamic characteristic is the same as for the other components. Hence, it is projected over to the dynamic curve as point G.

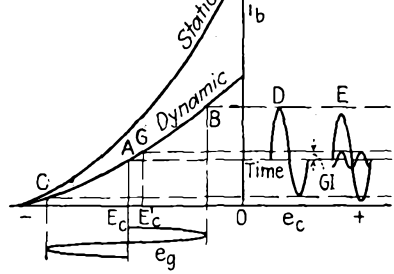


FIG. 41.—Shifting of operating point due to self-rectification, transfer characteristics.

The situation becomes more involved in the case of inductive plate feed, since now the additional direct current as well as the original direct current flows through the zero resistance choke, while the a.c. components must flow through R_L. As a consequence, the load line for the direct current is vertical, while that for the alternating current is the sloping one for R_L, and the operating point is moved off the originally assumed load line passing through the quiescent point.[4] The solution of this difficult case is one of successive approximations, whether per-

formed graphically or analytically. In Fig. 42 the solution is indicated to the first degree of approximation. BAC is the original load line drawn through the quiescent point A. From this the tentative wave shape can be found for a sine-wave signal voltage, as in Fig. 39, and this wave shape analyzed for the additional d.c. component. This is drawn vertically upward from A as AG, as the resistance to AG is that of the choke, *viz.*, zero. The load line for R_L is now drawn through G (which is the first approximation to the operating point) and is represented by KGL. The resultant plate-current wave shape is again analyzed for further additional d.c. component, and the operation repeated until no further increase is obtained. The increments are usually negligible after the first approximation (which is therefore sufficient) because, it will be noted, G represents a lower bias voltage than A, so that the

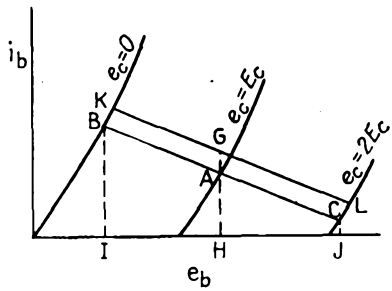


FIG. 42.—Correction to load line—inductive plate feed, plate characteristics.

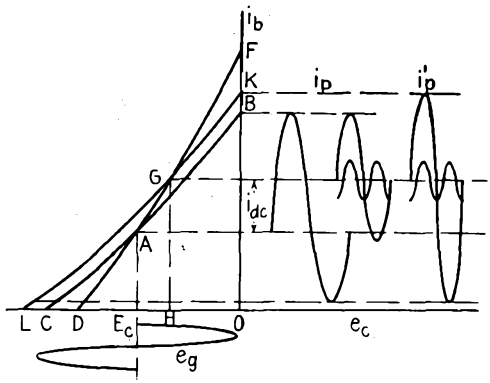


FIG. 43.—Correction to load line—inductive plate feed, transfer characteristics.

corrected grid swing about G is greater in the negative direction than in the positive direction (where G is the corrected a.c. axis), as a consequence of which the rise in plate current about G does not appreciably exceed the decrease in plate current in spite of the upward curvature of the family of characteristics.

In Figure 43 the same state of affairs is pictured on the i_b-e_c plane. The quiescent point A for $e_c = E_c$ is on the static curve DAF. The dynamic curve is CAB, and for a sine-wave grid signal voltage e_g the current wave is i_p. This contains a d.c. component i_{dc}, and this value is projected over to the static curve as G. The corrected dynamic characteristic is LGK, and G is the operating point, different from the quiescent point A. As stated before, for class A operation, i_{dc} (as well as the harmonic distortion) is small, and one approximation, as shown, is usually sufficient. In fact, for ordinary purposes this correction is unimportant, particularly for values of R_L comparable with R_p. This can be seen from the fact that, if R_L is not very high, the dynamic curve does not differ noticeably from the static curve over a small interval around the quiescent point A, so that G is nearly on the original dynamic curve CAB, and hence the corrected dynamic curve LGK is displaced from its original position CAB by only a small amount. As a result, the corrected plate current i_p' is nearly identical with the original shape i_p.

The correction for tube rectification is of some importance in class A work only in the case of pentodes, where the generation of additional d.c. component and harmonic distortion may be appreciable. In class B and C work (single-ended), where the grid swing is large and beyond the cutoff of the plate current, the correction may be considerable. However, in these cases, R_L may be chosen much below R_p in value (where the grid swings positive) in order to prevent the plate voltage from dropping to zero during the positive half cycle of the grid swing; and where this is the case, the dynamic curve does not differ much from the static curve, so that the operating point is nearly on the original dynamic curve and its correction almost coincides with it. In resistance coupling it has been noted that the operating point remains on the original dynamic curve, so that the maximum and minimum values of i_p' are the same as those for i_p and hence such items as power calculations remain unaffected. In passing, it may be noted that in a truly linear tube no rectification occurs during class A operation and the above correction is therefore unnecessary.

28. Load Line for a Reactance.—If an inductance L of finite value (in contradistinction to an ideal choke) be used as a plate load, the load line becomes an ellipse, for a linear tube. This is

a consequence of the fact that the voltage across L is out of phase with i_p, so that the load voltage is a doubled-valued function of i_p and the load line becomes a closed curve, *viz.*, an ellipse. To show this, assume that a sine-wave voltage is applied to the grid of a linear tube. The resulting variation in plate current will be sinusoidal in shape and can be expressed as

$$i_p = I_p \sin \omega t \tag{51}$$

This current, in flowing through the load impedance Z_L, consisting of inductance L and resistance R_L in series, sets up a load voltage that may be expressed in complex notation as

$$e_L = i_p Z_L = (I_p \sin \omega t)(R_L + j\omega L) \tag{52}$$

Multiplying through, we obtain

$$e_L = I_p R_L \sin \omega t + jI_p \omega L \sin \omega t \tag{53}$$

We may express

$$jI_p \omega L \sin \omega t = I_p \omega L \cos \omega t \tag{54}$$

Substituting Eq. (54) in Eq. (53),

$$e_L = I_p R_L \sin \omega t + I_p \omega L \cos \omega t \tag{55}$$

In the right-hand side of the equation, i_p is expressed in terms of its maximum value I_p and the sine and cosine of ωt. We wish i_p to be expressed directly in terms of I_p, and not with reference to ωt or t, as the load line is a relation between instantaneous current and voltage and depends only indirectly upon time in so far as this varies the impedance function that the load line represents. Hence, we shall substitute in Eq. (55) i_p for $I_p \sin \omega t$ and $\sqrt{I_p^2 - i_p^2}$ for $I_p \cos \omega t$ and obtain

$$e_L = R_L i_p + \omega L \sqrt{I_p^2 - i_p^2} \tag{56}$$

or

$$e_L - R_L i_p = \omega L \sqrt{I_p^2 - i_p^2} \tag{57}$$

Squaring, we obtain

$$e^2{}_L - 2R_L e_L i_p + R_L{}^2 i_p{}^2 = \omega^2 L^2 I_p{}^2 - \omega^2 L^2 i_p{}^2$$

or

$$e^2{}_L - 2R_L e_L i_p + i_p{}^2(R_L{}^2 + \omega^2 L^2) = \omega^2 L^2 I_p{}^2 \tag{58}$$

Since

$$R^2{}_L + \omega^2 L^2 = Z_L{}^2$$

we obtain finally

$$e_L^2 - 2R_L e_L i_p + i_p^2 Z_L^2 = \omega^2 L^2 I_p^2 \tag{59}$$

Equation (59) is that of an ellipse whose origin is at the quiescent point A (Fig. 44).

The following facts can be shown to hold:

1. The intersections of the maximum and minimum values of i_p with the ellipse at B and C, respectively, represent as abscissa the maximum and minimum values, respectively, of the inphase or resistive voltage drops, that is, $R_L I_p$ and $(-R_L I_p)$. These are shown as AO and AE, respectively.

2. Also, when i_p is zero (total plate current at that instant equals I_b), the maximum quadrature voltage drop across L is obtained. Thus, AF represents $I_p\omega L$, and AG represents $-I_p\omega L$.

3. Similarly, AH and AI represent $I_p Z_L$ and $-I_p Z_L$, respectively, which are the maximum positive and negative values of e_L.

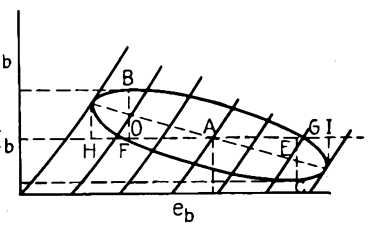


FIG. 44.—Reactive load line, plate characteristics.

4. Since the sum of the plate voltage, e_b, and e_L must at all times equal E_{bb}, we have that the change e_p in plate voltage from its normal value is given by the equation

$$e_p = -e_L \tag{60}$$

so that an equation similar to Eq. (59) can be obtained between e_p and i_p. Hence, the path of i_p, with respect to e_p, is the above ellipse.

5. For an inductive load, in which case we have $+j\omega L$, the ellipse is to be traversed in a clockwise direction; for a capacitive load $-j/\omega C$, in a counterclockwise direction.

6. In view of (1), (2), and (3) or from Eq. (59), it can be seen that, the greater R_L compared with ωL, the flatter the ellipse. Thus, as the inductive reactance is reduced the load ellipse flattens out to the familiar load line for R_L, *viz.*, BAC.

7. The derivation culminating in Eq. (59) is independent of the quiescent point. The latter may therefore be determined by any B supply voltage and any resistance R_{dc} of the load that would determine I_b and hence point A. Thus R_{dc} may be differ-

ent in value from R_L, the load resistance to alternating current. Also, for the linear tube, the ellipse is symmetrical, and hence the a.c. component is symmetrical about the quiescent point, so that the operating point coincides with the quiescent point.

8. The projection of the i_b-e_b curve (ellipse) upon the other coordinate planes—in particular, the i_b-e_c plane—is evidently an ellipse too; i.e., the dynamic characteristic is an ellipse instead of an open curve, as in the case of a resistance (see Fig. 45). A is an ellipse for a load impedance Z_L, in which the ratio of ωL to R_L is large; B is an ellipse for the same magnitude of ωL, but a smaller ratio of ωL to R_L; C is the line for R_L equal to Z_L and ωL equal to zero.

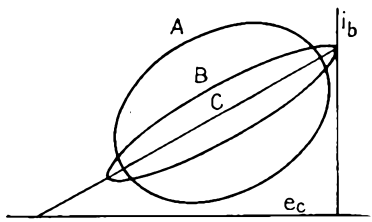


FIG. 45.—Reactive load line, grid characteristics.

9. For a nonlinear tube, Eq. (51) no longer holds, in that i_b contains harmonic-distortion products. The curve is still in general a closed one, but not elliptical in shape (although it may be described visually as an ellipse having curved axes). The derivation of this curve can be accomplished graphically and is illustrated in the next chapter. The most important feature here is that the elliptical load line dips farther down into the curved lower part of the plate family of curves than the one for a load resistance, and hence the distortion products are greater. In order to operate on the upper, more linear part of the curves, it is necessary to reduce the C bias or increase the load impedance, and this is more necessary as the ratio of L to R_L increases, owing to the wider ellipse obtained.

In practice, reactance plate loads are intentionally used in voltage amplifier stages, where the grid swing is usually comparatively low and Z_L very high, in order to obtain a high voltage gain. In this case, the slant of the ellipse and its size are such as generally to avoid the high distortion part of the characteristic. Conditions, however, may possibly arise in a power tube under which the leakage reactance of the output transformer, as well as the inductive reactance of the connected loud-speaker, will give an elliptical load line at the higher frequencies and hence more distortion than at the lower frequencies. These distortion products, however, will be mainly outside the frequency band,

although the difference beat frequency for a complex signal voltage may not be.

Reactive load impedances in the general case are treated by a method given in the following chapter. The derivation in this section is given mainly as an example of the complex reactions actually taking place even in a linear tube, but any unfavorable effects are minimized under usual operating conditions by the requirement of a high ratio of Z_L to R_p for high voltage gain.

BIBLIOGRAPHY

1. CARSON, J. R.: A Theoretical Study of the Three-element Vacuum Tube, *Proc. I.R.E.*, April, 1919.
2. SCHADE, O. H.: *Beam Power Tubes, Proc. I.R.E.*, March, 1938.
3. EASTMAN, A. V.: "Fundamentals of Vacuum Tubes," pp. 133–141, McGraw-Hill Book Company, Inc., 2d ed, 1941.
4. KILGOUR, C. E.: Graphical Analysis of Output Tube Performance, *Proc. I.R.E.*, **19** (1), 42, 1931.
5. CHAFFEE, E. L.: "Theory of Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., 1933.
6. KELLOGG, E. W.: Design of Non-distorting Power Amplifiers, *Jour. A.I.E.E.*, May, 1925.
7. WILLIS, F. C., and L. E. MELHUISH: Load-carrying Capacity of Amplifiers, *Bell System Tech. Jour.*, October, 1926.

CHAPTER IV

REACTIVE LOADS

1. Introduction.—Chapter III deals with nonlinear resistive circuits, special circuits involving ideal inductances paralleled by linear resistances, and a linear resistance (linear vacuum tube) in series with an inductance. This chapter deals with nonlinear resistances in series with reactive loads and voltages of any wave form, although the method it develops can also be employed in the case of nonlinear inductances or condensers. It is therefore general in scope and comparatively simple to apply. For want of a better name, it shall be called the *finite-operator method.*

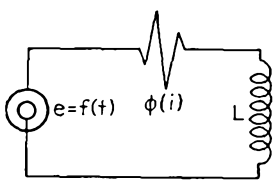


FIG. 46.—Simple *Lr*-series nonlinear circuit.

The finite-operator method is an adaptation of the point-by-point method of solving a differential equation and is analogous to that described by R. Usui,[1] but it is not confined to oscillatory circuits. Perhaps the best way to explain it is by means of illustrative examples.

2. Inductance and Nonlinear Resistance.—Suppose we have an inductance L, in series with a nonlinear resistance r, and a source of e.m.f. e, which is some known function of time $f(t)$. Let r be defined in terms of the voltage e_r and current i relations. Thus

$$e_r = \phi(i) \tag{1}$$

The equation for the circuit (shown in Fig. 46) is

$$e = f(t) = L\frac{di}{dt} + \phi(i) \tag{2}$$

This equation is not, in general, capable of solution, owing to the presence of the term $\phi(i)$. If r were a linear resistance, $\phi(i)$ would represent a constant, G, its conductance; and Eq. (2) would become an ordinary linear differential equation of first order and degree and could easily be solved. In the form shown above, it is solvable by means of a series expansion for

certain forms of $\phi(i)$. The point-by-point method, and particularly the graphical method to be given, is capable of solving it even though $\phi(i)$ is known, not analytically, but merely as an experimental plot of i vs. e_r, known as the *load line* or *terminal characteristic for r*.

In using the point-by-point method, we start with certain initial conditions of e and i. Usually i is assumed zero at the start, as the start is ordinarily the closing of a switch. For the initial value of e and i, a value Δi_1 is found, which, in a given time Δt_1, will satisfy Eq. (2). The current is now

$$i_2 = i_1 + \Delta i_1 \tag{3}$$

and the voltage is now

$$e_2 = e_1 + \Delta e \tag{4}$$

A new time interval Δt_2, preferably equal to Δt_1, is assumed, and with e_1 and i_2 as given above a new increment Δi_2 is found that satisfies Eq. (2). The process is repeated until as many points of the over-all load line (relation between over-all voltage and generator current i) are determined as required. The process is therefore an approximate method of determining some particular integral or solution of Eq. (2), depending upon the initial conditions assumed.

In this method, the following will be noted:

1. The voltage e at the start of an interval of time Δt is assumed to remain at that value during the interval and then instantly to jump to a value $e + \Delta e$ at the end of the interval. It then remains constant at this value and then jumps to the next value at the end of the next interval, etc. This means that the voltage time wave is being broken up into narrow rectangles of width Δt and of such length that they fall within the time-wave curve. It is possible, however, to assume that the voltage rises to the value $e + \Delta e$ at the beginning of the time interval. This is the method of breakup of the voltage time wave employed in the graphical method, to be described next, and corresponds geometrically to breaking up the wave into rectangles whose tops fall outside the curve. Finally, it may be assumed that the voltage rises linearly with time from the value e to $e + \Delta e$ so that its average value during the time interval Δt is $e + (\Delta e/2)$. This corresponds to breaking up the wave into narrow trapezoids and approximates the wave form more closely than the

other two methods. Any one of the three may be employed in either method, and for sufficiently small time intervals should yield results as nearly equal as desired.

2 The additional voltage drop in r due to Δi is not taken into account. To do so would require a series of trials, especially if r is nonlinear. The graphical method does take this factor into account.

To proceed with the graphical method, Eq. (2) is first rewritten in finite-increment form. It is to be remembered that the voltage at the start of and throughout the time interval Δt is assumed to be $e + \Delta e$ and that the current at the end of the time interval is $i + \Delta i$. Hence,

$$e + \Delta e = L \frac{\Delta i}{\Delta t} + \phi(i + \Delta i) \tag{5}$$

The fraction $L/\Delta t$ is to be called the *finite operator curve* and in this case is to be represented by a straight line (because L is assumed a linear inductance) at an angle to the voltage axis of

$$\theta = \cot^{-1} \frac{L}{\Delta t} \tag{6}$$

The graphical construction is as follows:

The voltage time wave is broken up into small rectangles of (preferably equal) width Δt. The voltage at time t will be e; throughout the interval $t + \Delta t$, $e + \Delta e$. In Fig. 47, suppose, for example, that, at $t = 0$, $e = 0$, $i = 0$. The first voltage increment Δe_1 is projected up as OC. From C a straight line at the angle θ [Eq. (6)] is drawn. It intersects the load line for r in A. Then AB is the first increment of current, Δi_1. It produces a voltage drop in r of OB equal to $\phi(\Delta i_1)$ and in the inductance L of value

$$BC = AB \cot \theta = \Delta i_1 \frac{L}{\Delta t_1} \tag{7}$$

Point D is the first point of the over-all load line. From point E, as shown, GE is drawn at the same angle θ. Then GH is Δi_2, the next increment of current in time Δt_2. It is evident that

$$OB + AH = \phi(\Delta i_1 + \Delta i_2) \tag{8}$$

and

$$HE = \Delta i_2 \frac{L}{\Delta t_2} \tag{9}$$

and that

$$OB + AH + HE = \Delta e_1 + \Delta e_2 \tag{10}$$

so that Eq. (5) is satisfied. The construction is continued, and further points on the over-all load line are thus obtained. In

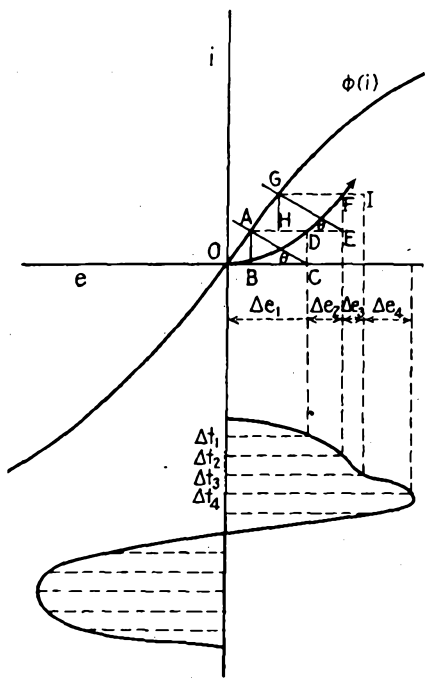


FIG. 47.—Graphical solution for *Lr* nonlinear circuit.

general, the load line spirals around until it finally closes, which means that steady-state conditions have been attained. The number of spirals necessary is in general theoretically infinite, but in practice the curve is essentially closed after a few turns. This convergence depends, among other things, upon the skill with which the initial conditions are chosen and the damping in the circuit. When the steady-state solution alone is desired, the transient spirals are a handicap and are due to the generality of the method.

3. Illustrative Example.—The above method may be illustrated by an example involving a purely linear circuit, for which the analytical solution is known. Some nonlinear examples will be given later. Consider a circuit having only inductance in series with a sinusoidal voltage e applied so that, when $t = 0$, $e = 0$.

Since the circuit has no resistance we can expect a transient that never dies out. The construction is shown in Fig. 48. We note that the figure closes in one cycle and that it is distorted in its elliptical form, also that it is entirely in the first and second

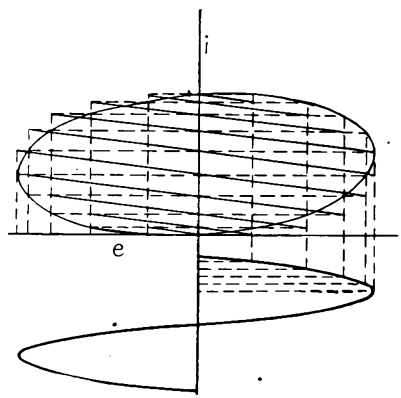


FIG. 48.—Displaced ellipse obtained for pure linear inductance.

quadrants. The distortion is due to finite increments used. The fact that the figure is in the first and second quadrants indicates that there is a d.c. component present in the current flow, of value (from the symmetry of the figure) of one-half the peak value. Its actual value, as well as the equation of the figure, can be calculated as follows:

Let the time intervals Δt be such that

$$\Delta t = \frac{t}{x} \tag{11}$$

where x is the number of current increments obtained in the time t. Let the equation of the voltage be

$$e = E_m \sin \omega t \tag{12}$$

From the figure it is evident that

$$\left.\begin{aligned}
\Delta i_1 &= \frac{E_m t}{xL} \sin \frac{\omega t}{x} \\
\Delta i_2 &= \frac{E_m t}{xL} \sin \frac{2\omega t}{x} \\
\Delta i_3 &= \frac{E_m t}{xL} \sin \frac{3\omega t}{x} \\
&\cdots\cdots\cdots \\
\Delta i_x &= \frac{E_m t}{xL} \sin \omega t
\end{aligned}\right\} \tag{13}$$

Summing up the equations we obtain

$$i_x = \frac{E_m t}{xL}\left(\sin \frac{\omega t}{x} + \sin \frac{2\omega t}{x} + \cdots + \sin \omega t\right) \tag{14}$$

The sum of the series of the angles is known from trigonometry, so that Eq. (14) becomes

$$i_x = \frac{E_m t}{xL}\left[\frac{\sin \frac{1}{2}(x+1)(\omega t/x) \times \sin (\omega t/2)}{\sin \frac{1}{2}(\omega t/x)}\right] \tag{15}$$

As x approaches infinity, this reduces to

$$i_x = \frac{E_m t}{xL\,(\omega t/2x)}\left(\sin^2 \frac{\omega t}{2}\right) = \frac{2E_m}{\omega L}(1 - \cos \omega t) \tag{16}$$

Equation (16) checks with the analytic expression for this circuit and is the equation of the curve if the time intervals Δt are made infinitesimal in value. The d.c. component mentioned previously is $E_m/\omega L$, and the peak value is evidently double this or $2E_m/\omega L$.

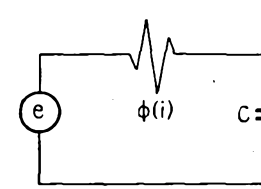


FIG. 49.—Simple rC-series nonlinear circuit.

It is thus seen that this graphical method checks with the analytical solution for a simple circuit and initial conditions. At the present time, it would appear that the proof given above is too complicated to use in establishing the correctness of the graphical method for more complicated circuits and initial conditions.

4. Capacity and Nonlinear Resistance.—We now take the case of a capacity C in series with a nonlinear resistance r and voltage

e (Fig. 49). Assume, for simplicity, that there is no initial charge in C and no current flow in the circuit. The equation, in differential form, is

$$e = f(t) = \frac{1}{C}\int i\,dt + \phi(i) \tag{17}$$

In finite-increment form this becomes

$$e + \Delta e = \frac{\Delta t}{C}\sum i + \frac{\Delta t}{C}\,i + \phi(i) \tag{18}$$

The finite operator curve will now be $\Delta t/C$. Note that $(\Delta t/C)\,\Sigma i$ is the total voltage E_c, built up in the condenser in the previous time intervals, and $(\Delta t/C)i$ is the additional voltage e_c built up in the condenser in the present time interval by the current flow. Equation (18) may therefore be rewritten as

$$e + \Delta e - E_c = \frac{\Delta t}{C}\,(i) + \phi(i) \tag{19}$$

The graphical construction is as follows:

The voltage time wave is broken up into small (preferably equal) time intervals Δt. In Fig. 50A, Δe_1 is the first voltage increment OC. At C, a line AC at the angle

$$\theta = \cot^{-1}\frac{\Delta t}{C} \tag{20}$$

is drawn, and it intersects the load line for r at A. Note that AC is the finite operator curve, straight because C was assumed a constant parameter. The first current increment Δi_1 is evidently AB, while OB is the voltage across r and BC is the voltage developed across C. The point D is the first one on the over-all load line. A length AB equal to Δi_1 is now laid off on the auxiliary diagram (Fig. 50 B), and AC is drawn at the angle θ to BC. Then BC is the total voltage E_c developed across the condenser, which in this first construction is the same as BC in Fig. 50A. The second voltage increment Δe_2 is laid off (point E), and a distance FE $(=BC)$ subtracted from it. Through F the finite-operator curve is drawn at the angle θ to the voltage axis; it intersects the load line for r in G. Then GH is the new value of current, and I is the second point on the over-all load line. The

distance HF represents $(\Delta t/C)i$, FE represents $E_c = (\Delta t/C)\Sigma i$, OH represents $\phi(i)$, and, clearly,

$$OH + HF + FE = e + \Delta e \tag{21}$$

The remaining points of the over-all load line are found in the same manner. Thus, GH is laid off on Fig. 50B as GC, and GF is drawn at the angle θ. Then BF is the new value of $E_c = (\Delta t/C)\Sigma i$. This is laid off from J (Fig. 50A) as JK, and

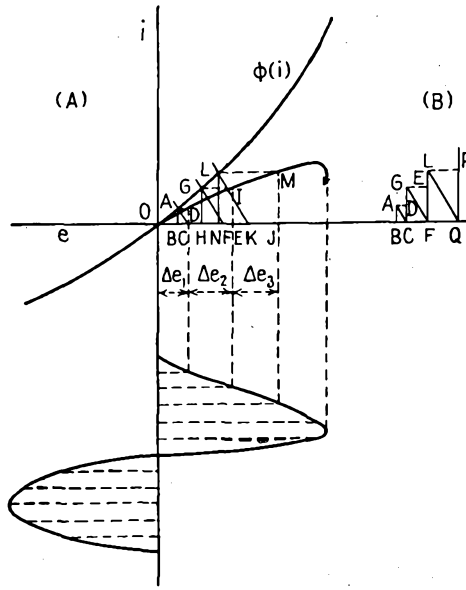


FIG. 50.—Graphical solution for rC nonlinear circuit.

from K the finite operator curve is drawn. It intersects the load line of r in L, and M is the third point in the over-all load line, while LN is the new value of current. This is laid off on the auxiliary diagram (Fig. 50B) as LF, and the finite operator curve LQ drawn. Then BQ is the next value of E_c, which is laid off from the end of Δe_4 on Fig. 50A, etc. Since E_c builds up quite rapidly as i continues to flow, it ultimately overtakes the impressed voltage e, so that the operator curve begins to move to the left and i begins to decrease even if e is still increasing. The result is that the over-all load line begins to fold over and spiral around clockwise, which is the proper direction for a capacitive load line to take.

The auxiliary diagram (Fig. 50*B*) has more than constructional utility. The points *D*, *E*, and *P* as shown are the points on the capacitive load line; *i.e.*, they are the locus of current vs. voltage across the condenser when the latter is in series with the given nonlinear resistance *r* and the voltage *e* impressed.

5. Illustrative Examples.—As before, an example illustrating the method outlined above will be chosen from a linear circuit for which the analytic solution is known.

Suppose a condenser *C* and linear resistance *R* are in series with a unit impulse voltage E_1. The graphical construction is shown in Fig. 51.

At the start ($t = 0$), Δt is zero (hence $\Delta t/C$) and therefore θ is 90 deg., so that the first operator curve AB is perpendicular to the

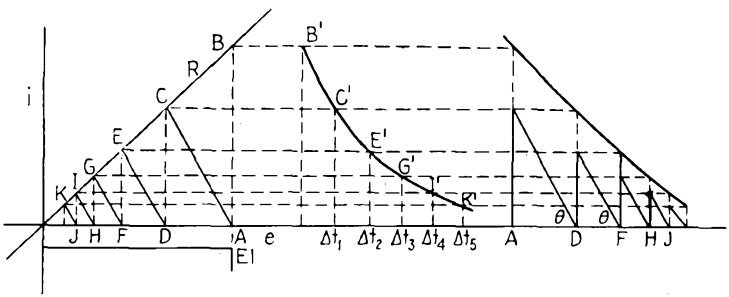


FIG. 51.—Exponential charging of linear *C* through linear *R* from d.c. voltage.

e axis, as shown. The first current flow is therefore AB. From then on, the finite operator curves AC, DE, FG, HI, JK, etc., make the angle θ given by Eq. (32) with the *e* axis. Since the impressed voltage E_1 is constant after $t = 0$, the voltages built up across the condenser due to the accumulating charge are subtracted in line from point A. Thus, E_{c_1} equals AD; E_{c_2} equals $AD + DF$; E_{c_3} equals $AD + DF + FH$; etc. An auxiliary diagram is therefore not required for constructional purposes in this special case but has been drawn in at the extreme right to give the load line for the condenser alone. The current may be plotted against time, as shown in the center diagram (this can be done in any case for a load line).

The relation between current i_x, during the interval Δt starting at any time *t*, and voltage E_{x-1} at the beginning of that interval may be formulated for the above example as follows:

$$i_{x-1} = E_{x-1} \frac{1}{R + (\Delta t/C)} \tag{22}$$

$$i_x = \frac{E_{x-1} - i_{x-1}(\Delta t/C)}{R + (\Delta t/C)} = \frac{E_{x-1}[1 - (\Delta t/C)/(R + \Delta t/C)]}{R + (\Delta t/C)} \tag{23}$$

so that

$$\frac{i_x}{i_{x-1}} = 1 - \frac{\Delta t}{C[R + (\Delta t/C)]} \tag{24}$$

and therefore

$$\frac{i_x}{i_1} = \left\{1 - \frac{\Delta t}{C[R + (\Delta t/C)]}\right\}^{(x-1)} \tag{25}$$

Since

$$i_1 = \frac{E}{R + (\Delta t/C)} \tag{26}$$

and

$$\Delta t = \frac{t}{x} \tag{27}$$

we obtain

$$i_x = \frac{E}{R + (t/xC)} \left[\frac{1}{1 + (t/xCR)}\right]^{(x-1)} = \frac{E}{R} \left[\frac{1}{1 + (t/xCR)}\right]^{x} \tag{28}$$

As x increases without limit ($\Delta t \to 0$),

$$i_x \to \frac{E}{R} e^{-t/RC} \tag{29}$$

which checks the analytic solution for an RC circuit in series with an impulse voltage E_1.

6. Inductance, Linear Resistance, and Nonlinear Resistance in Series.—We shall now proceed more briefly with several other circuit combinations. First we take the case of an inductance L in series with a linear resistance R, a nonlinear resistance r, and a source of e.m.f. $e = f(t)$.

It is possible first to combine R and r in series graphically as described in Chap. I. Then this new value of resistance can be solved with L as described in Sec. 2 above. However, it is desirable to perform the above construction in one diagram, if possible, particularly if in a more involved circuit it is necessary to do so in order to perform the entire construction at all.

The finite-increment equations are

$$e + \Delta e = \phi(i + \Delta i) + (i + \Delta i)R + \Delta i \frac{L}{\Delta t} \tag{30}$$

or

$$e + \Delta e - e_r = iR + \Delta i\left(R + \frac{L}{\Delta t}\right) \tag{31}$$

where

$$e_r = \phi(i + \Delta i)$$

as before.

The reader may have noticed by this time how similar are finite-operator expressions such as $R + (L/\Delta t)$ to operational and j-operator impedance expressions. Further, such similarities will be noted in succeeding examples.

The graphical construction is as follows:

The voltage wave is as usual broken up into finite time increments Δt and voltage increments Δe. Suppose, as in Fig. 52, at some instant the voltage is $e + \Delta e$ and the current at that time is i. From the end of $e + \Delta e$ a line R is drawn at an angle to the e axis of value

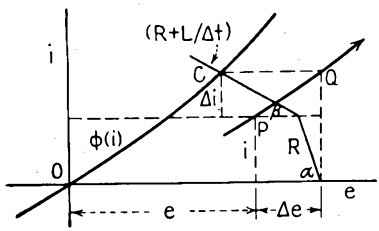


FIG. 52.—Series circuit involving linear L and R and nonlinear r.

$$\alpha = \cot^{-1} R \tag{32}$$

Where it intersects the abscissa through the tip of i, another line $R + (L/\Delta t)$ is drawn at the angle β such that

$$\beta = \cot^{-1}\left(R + \frac{L}{\Delta t}\right) \tag{33}$$

Where this second line intersects the load line for r in C determines Δi, as shown, and point Q is the next point of the load line. (P was the preceding point previously determined.) The method of construction is of course used from the very start, and Fig. 52 merely illustrates the method at some particular moment. The reader can check for himself that it conforms with Eqs. (30) and (31). Indeed, it will be noted that the construction is such that i is determined by the load lines for R and r, and Δi by $L/\Delta t$ in addition to r and R.

7. Capacitance, Linear Resistance, and Nonlinear Resistance in Series.—We next take the case of a capacitance C, a linear resistance R, and nonlinear resistance r, in series with an e.m.f. $e = f(t)$. The equations are

$$e + \Delta e = \phi(i) + iR + \frac{\Delta t}{C}\sum i + \frac{\Delta t}{C} i \tag{34}$$

$$e + \Delta e - e_r - E_c = i\left(R + \frac{\Delta t}{C}\right) \tag{35}$$

where $\phi(i) = e_r$ and $(\Delta t/C)\Sigma i = E_c$.

The voltage built up in the condenser, E_c, can be determined, starting with the initial current, in the manner described in Sec. 4. The only change is that, instead of using $\Delta t/C$ as the cot θ in Eq. (20) for the construction of Fig. 50*A*, the finite operator curve is at an angle

$$\theta' = \cot^{-1}\left(R + \frac{\Delta t}{C}\right) \tag{36}$$

To determine E_c, however, we use Fig. 50*B* unchanged; *i.e.*, angles *ACB*, *GFB*, etc., are equal to θ as given by Eq. (20). Again we note that R and r could be first combined in series into an equivalent resistance and then the constructions of Fig. 50 used as shown but that the method given above achieves the same result with one less diagram. Also, if R is zero, the construction given above becomes identical with that given in Sec. 4.

8. Inductance, Capacitance, Linear Resistance, and Nonlinear Resistance in Series.—In this circuit we have an L, C, R, r series circuit. The final equation for the graphical construction is

$$e + \Delta e - e_r - E_c = i\left(\frac{\Delta t}{C} + R\right) + \Delta i\left(\frac{\Delta t}{C} + \frac{L}{\Delta t} + R\right) \tag{37}$$

where $e_r = \phi(i + \Delta i)$ and $E_c = (\Delta t/C)\Sigma i$.

The graphical construction is given in Fig. 53. The angles are given by

$$\alpha = \cot^{-1}\left(\frac{\Delta t}{C} + R\right) \tag{38}$$

$$\theta = \cot^{-1}\left(\frac{\Delta t}{C} + \frac{L}{\Delta t} + R\right) \tag{39}$$

The voltage E_c across the condenser at the beginning of the time interval Δt is found by an auxiliary diagram exactly the same as that shown in Fig. 50*B*, and the angle θ is the same as there.

In Fig. 53, the current increment Δi is represented by AB and together with the current i at the beginning of the time interval forms the new value of current used to obtain the next value of E_c, as well as determining the next point of the over-all load line. Point C is the one just determined. If $R = 0$, angles α and θ can still be found from Eqs. (38) and (39), and the same construction used. The circuit is now reduced to an L, C, r series circuit.

9. Parallel Inductive Circuit.—In Fig. 54 is shown an e.m.f. in series with a nonlinear resistance r and a combination of an

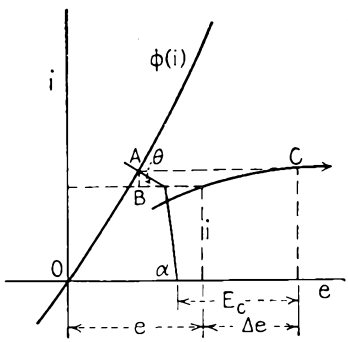


FIG. 53.—Series circuit having linear L, C, and R and nonlinear r.

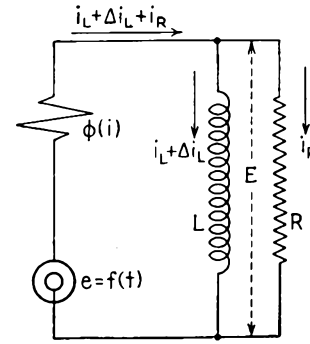


FIG. 54.—Parallel LR in series with nonlinear r.

inductance L and linear resistance R in parallel. Suppose the currents are as shown, that the impressed voltage during time interval Δt is $e + \Delta e$, and that the voltage across the LR parallel combination is E. We have, then, that

$$e + \Delta e - \phi(i_L + \Delta i_L + i_R) = E \tag{40}$$

$$E = \Delta i_L \frac{L}{\Delta t} = i_R R \tag{41}$$

$$\Delta i_L = E \frac{1}{L/\Delta t} \tag{42}$$

$$i_R = \frac{E}{R}$$

$$i_R + \Delta i_L = E\left(\frac{1}{L/\Delta t} + \frac{1}{R}\right) \tag{43}$$

$$E = (i_R + \Delta i_L)\left[\frac{1}{(\Delta t/L) + (1/R)}\right] \tag{44}$$

Substituting the value of E from Eq. (44) in Eq. (40), we obtain

$$e + \Delta e - \phi(i_L + \Delta i_L + i_R) = (i_R + \Delta i_L)\left[\frac{1}{(\Delta t/L) + (1/R)}\right] \tag{45}$$

Equation (45) forms the basis of the graphical construction shown in Fig. 55. At some point in the construction the voltage has attained the value $e + \Delta e_m$. The inductive current from the previous time interval has attained the value i_L, as shown. From A the finite-operator curve BA is drawn so that

$$\angle BAC = \cot^{-1}\frac{1}{(\Delta t/L) + (1/R)} \tag{46}$$

Its intersection B with the load line of r determines point F, a newly determined point on the over-all load line. The voltage across the parallel circuit is E, as shown. Then DA is drawn such that

$$\angle DAC = \cot^{-1}\frac{L}{\Delta t} \tag{47}$$

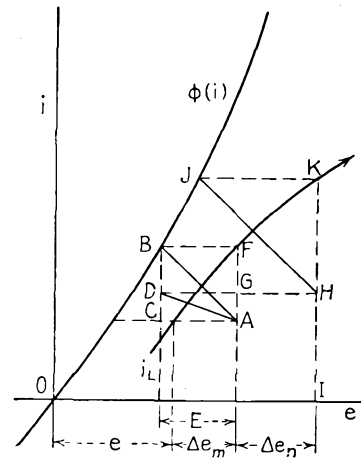


FIG. 55.—Graphical construction for parallel LR.

It is evident that $DC = \Delta i_L$, $BD = i_R$, and $\angle BGD = \cot^{-1} R$. To determine the next point of the load line, we use as our new current through the inductance the previous value i_L, or HI. Through H a line is drawn parallel to BA, and where it intersects the load line for r in J determines the next point K of the over-all load line. The construction is continued as just described, thus giving succeeding over-all load-line points. The start of the construction is of course determined by the initial conditions.

10. Parallel Capacitive Circuit.—Suppose, in Fig. 54, we replace the inductance L by a condenser C. Our equations are

$$e + \Delta e - \phi(i_C + i_R) = E \tag{48}$$

$$E = \frac{\Delta t}{C}\left(\sum i_C + i_C\right) = i_R R \tag{49}$$

from which we obtain

$$E = \frac{1}{(C/\Delta t) + (1/R)} \left(i_R + i_C + \sum i_C \right) \tag{50}$$

so that

$$e + \Delta e - \phi(i_c + i_R) - \frac{1}{(C/\Delta t) + (1/R)} \sum i_C = \frac{1}{(C/\Delta t) + (1/R)} (i_R + i_C) \tag{51}$$

Equation (51) forms the basis of the graphical construction which is similar to that in Fig. 50. The quantity

$$\frac{1}{(C/\Delta t) + (1/R)} \sum i_C$$

is found at any time by adding the condenser currents up to that time (as in Fig. 50*B*) but drawing through the tip of each current a finite-operator line not at the angle given by Eq. (20) but at an angle

$$\theta = \cot^{-1} \frac{1}{(C/\Delta t) + (1/R)} \tag{52}$$

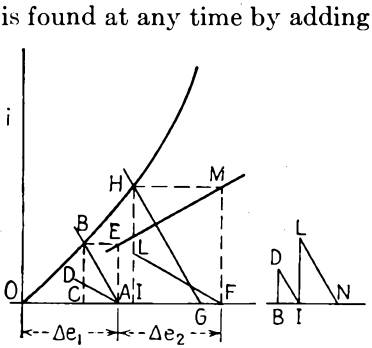


FIG. 56.—Graphical construction for parallel *RC* linear parameters.

The quantity is then subtracted from $e + \Delta e$ on the main diagram; and, from the end of the remainder starting on the e axis, finite-operator curves are drawn at the angle θ given by Eq. (52). Thus the first two successive points of the over-all load line are shown in Fig. 56. (Initial conditions are that current, charge, and voltage are all zero.) Starting from A, BA is drawn so that $\angle BAO = \theta$. Then DA is drawn so that $\angle DAO = \cot^{-1} R$. Then $DB = i_{c_1}$, and $DC = i_{R_1}$. On the auxiliary diagram to the right, DB is drawn equal to i_{c_1}, and DI so that $\angle DIB = \theta$. Then $BI = \frac{1}{(C/\Delta t) + (1/R)} \sum i_C$ for the first time interval, in which Σi_c is merely i_{c_1}, and BI is subtracted from $\Delta e_1 + \Delta e_2$ as FG on the left-hand main diagram. (Meanwhile, we note that E is the first point of the over-all load line). Through G, HG is drawn at

the angle θ, giving rise to M as the second over-all load-line point. If we draw FL parallel to DA, we obtain LI as i_R and HL as i_{c_2}. Then i_{c_2} is laid off on the auxiliary diagram to the right as LI, and LN drawn parallel to DI. Then BN represents the new value of $\frac{1}{(C/\Delta t) + (1/R)} \sum i_C$ that is to be subtracted from the voltage $\Delta e_1 + \Delta e_2 + \Delta e_3$ in the next time interval before the third over-all load-line point can be found. The procedure is continued until as much of the over-all load line is obtained as is desired.

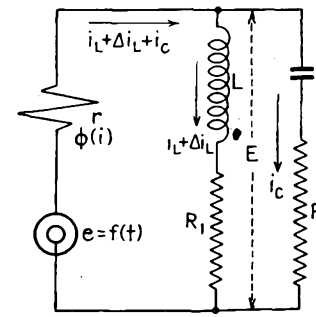


FIG. 57.—Parallel resonant nonlinear circuit

11. Parallel Resonant Circuit.—A parallel resonant circuit, shown in Fig. 57, has a nonlinear resistance r. We wish to solve for the incremental current in the case of L and present current in C and therefore may regard the previous currents i_L and Σi_C as known, since they are the algebraic sums of Δi_L and i_C, respectively, determined previously by the method by which the present Δi_L and i_C are to be found. With this in mind we write down the equations

$$e + \Delta e - \phi(i_L + \Delta i_L + i_C) = E \tag{53}$$

$$E = \frac{\Delta t}{C} \sum i_C + i_C \frac{\Delta t}{C} + i_C R_2 = i_L R_1 + \left(R_1 + \frac{L}{\Delta t}\right) \Delta i_L \tag{54}$$

Equation (54) may be written after some simple algebraic transformations as

$$E = (\Delta i_L + i_C) \left(\frac{Z_1 Z_2}{Z_1 + Z_2}\right) + i_L \left(\frac{R_1 Z_2}{Z_1 + Z_2}\right) + \left[\frac{\Delta t Z_1}{C(Z_1 + Z_2)}\right] \sum i_C \tag{55}$$

where $Z_1 = R_1 + (L/\Delta t)$ and $Z_2 = R_2 + (\Delta t/C)$.

Substituting the value of E from Eq. (55) in Eq. (53), we obtain

$$e + \Delta e - \phi(i_L + \Delta i_L + i_C) - i_L \left(\frac{R_1 Z_2}{Z_1 + Z_2}\right) - \left[\frac{\Delta t Z_1}{C(Z_1 + Z_2)}\right] \sum i_C = (\Delta i_L + i_C) \left(\frac{Z_1 Z_2}{Z_1 + Z_2}\right) \tag{56}$$

This last equation forms the basis of our graphical construction. We have managed to introduce the branch currents i_L and Σi_C (which, as stated above, are known) in such a manner that their voltage effects are both subtractive from $e + \Delta e$, instead of individually subtractive from E. The graphical construction is

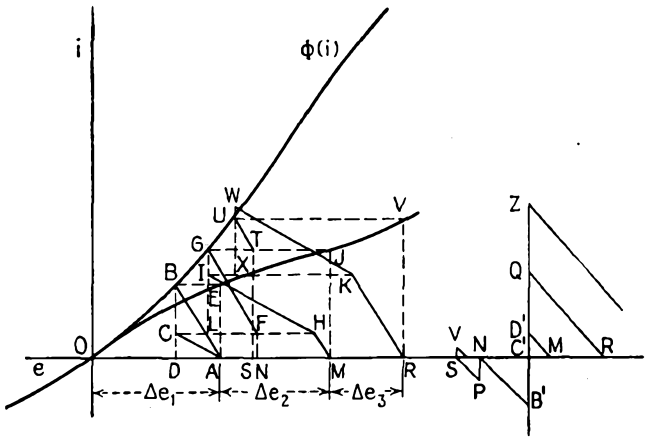


FIG. 58.—Graphical solution for parallel resonant circuit.

shown in Fig. 58. From A the finite operator curve BA is drawn so that

$$\angle BAO = \cot^{-1} \frac{Z_1 Z_2}{Z_1 + Z_2} \tag{57}$$

Then AC is drawn so that

$$\angle CAO = \cot^{-1} Z_1 \tag{58}$$

Then $CD = \Delta i_{L_1}$, and $BC = i_{c_1}$. These are then laid off on the auxiliary diagram on the right as $D'C'$ and $C'B'$, respectively. Then $D'M$ is drawn so that

$$\angle D'MC' = \cot^{-1} \frac{R_1 Z_2}{Z_1 + Z_2} \tag{59}$$

whereupon $C'M = i_L[R_1 Z_2/(Z_1 + Z_2)]$, where i_L is simply Δi_{L_1}. Also from B', NB' is drawn so that

$$\angle C'NB' = \cot^{-1} \frac{\Delta t Z_1}{C(Z_1 + Z_2)} \tag{60}$$

Then $C'N = [\Delta t Z_1/C(Z_1 + Z_2)]\Sigma i_C$, where Σi_c is simply i_{c_1} in this first time interval.

From the voltage during the next time interval, *viz.*, $\Delta e_1 + \Delta e_2$, both these quantities must be subtracted or a total distance MN (right-hand diagram) must be subtracted on the left-hand diagram. This length is shown in the latter diagram also as MN. From N we go up a distance Δi_{L_1} to F and draw FG parallel to AB. Then GL represents $i_C + \Delta i_L$, and J is the next point on the over-all load line. To get Δi_L and thus by subtraction from GL the quantity i_C, we proceed as follows: From M we draw HM so that $\angle HMO = \cot^{-1} R_1$. From H we draw HI parallel to CA. Then IL is Δi_L in this second time interval, and hence GI is i_C. On the auxiliary diagram to the right we draw $NP = GI$ and $QD' = IL$. Then PS is drawn parallel to $B'N$ and QR parallel to $D'M$. Then SR is subtracted from $\Delta e_1 + \Delta e_2 + \Delta e_3$ on the left-hand diagram, where it is represented by RS too. From T directly above S we draw TU to the load line of r, and V is the next point of the over-all load line. Then from R (as previously from M) we draw KR parallel to HM and then KW parallel to AC. The length WX represents the next value of Δi_L, which, when subtracted from UX, gives UW, a reverse, or discharge, current in the condenser. Thus, on the right-hand diagram, we draw $VS = WU$ and $QZ = WX$, and the construction continues in the manner described.

The construction is really simpler than the above detailed exposition would indicate. The actual operations are easy to perform, but the process is admittedly laborious. However, no other method will give the solution to this complicated circuit more quickly or easily. Indeed, if the load line for r is very irregular or broken, any other method is far more complicated or impossible to use, except, of course, a machine differential analyzer.

We note, in passing, that if C is infinite and $R_1 = 0$ the construction becomes identical with that given in Sec. 9, while, if $L = 0$ and $R_2 = 0$, it becomes identical with that given in Sec. 10.

12. Nonlinear Parallel Branch.—The illustrative examples that have been given should afford a fairly good insight into the general method of solving graphically reactive circuits. Other and more complicated examples can be worked out, but it must not be overlooked that the process becomes very involved, just as an operational expression or even ordinary integration becomes

very involved and almost impossible to evaluate. One other example will be given of practical and theoretical interest. The circuit is shown in Fig. 59. This may represent a linear driver tube of constant plate resistance R_p, driving the grid of a succeeding tube positive, the latter's very nonlinear resistance being represented by r_g. The equivalent voltage generated in the plate circuit of the driver tube is represented by $e = f(t)$. If this is of low frequency, then the driver transformer may be fairly accurately represented as shown, where R_{pw} represents the primary winding resistance, R_{sw} the secondary winding resistance, and L_m the open-circuit inductance (assumed linear). For the purpose of construction, it is convenient to lump R_p and R_{pw} into one equivalent resistance R_1 and R_{sw} and r_g into an equivalent non-linear resistance r.

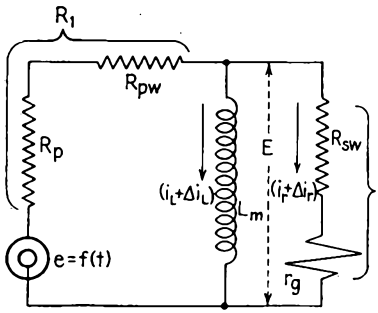


FIG. 59.—Circuit having nonlinear parallel resistor.

The equations are

$$e + \Delta e - (i_L + \Delta i_L + i_r)R_1 = E \tag{61}$$

$$E = \Delta i_L \frac{L}{\Delta t} = i_r r \tag{62}$$

from which we can obtain

$$E = (i_r + \Delta i_L) \frac{1}{(\Delta t/L) + (1/r)} \tag{63}$$

so that

$$e + \Delta e - (i_L + \Delta i_L + i_r)R_1 = (i_r + \Delta i_L) \frac{1}{(\Delta t/L) + (1/r)} \tag{64}$$

The expression $\dfrac{1}{(\Delta t/L) + (1/r)}$ is to be the finite operator curve that will be used to intersect R_1. It will be noted that, owing to r, this operator is truly curved, and not a straight line, as has been the case in the preceding examples. It must be determined before we can proceed with the remainder of the construction.

Referring to Fig. 60, we have the load line for r plotted. From any point A on the current axis, a line AB is drawn so that

$$\angle BAO = \cot^{-1}\frac{\Delta t}{L} \qquad (65)$$

The intersection B, projected upward to C in line with A, is a point on the finite-operator curve $\frac{1}{(\Delta t/L) + (1/r)}$. This means that, if the voltage were to change from zero to the value AC in a time interval Δt and L and r were in parallel, then the total current through the two branches would be AO, of which JO would be the portion through r and AJ the portion through L. Similarly, DE and GH, etc., are drawn parallel to AB and give rise to the respective points F and I, etc., on the finite operator curve. This curve we shall designate as Z.

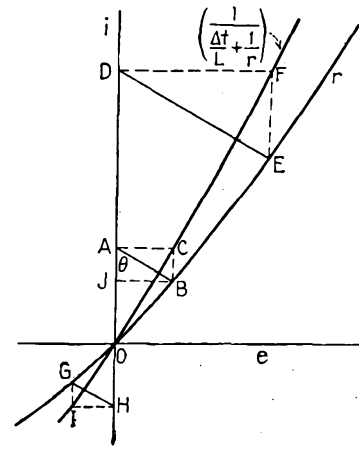


FIG. 60.—Parallel Lr finite operator—auxiliary graphical construction.

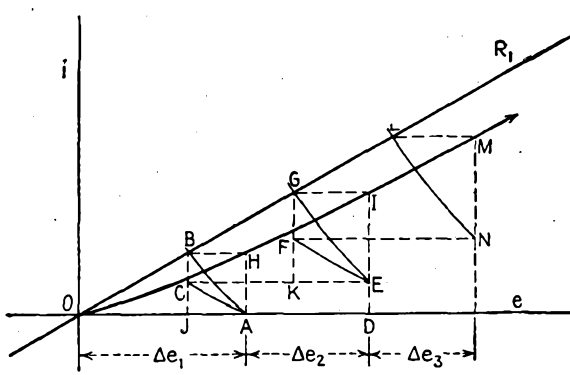


FIG. 61.—Graphical solution for nonlinear parallel resistor.

We can now proceed with the remainder of the construction, which is shown in Fig. 61. The load line for R_1 is plotted. While in this example it is assumed a straight line (R_1 linear), the method can be used even if R_1 is nonlinear as well as r. That is, the driver tube may be regarded as nonlinear too. From the end

of Δe_1 (point A) the finite operator curve AB, which was determined in Fig. 60 as OCF, is drawn. Its intersection with the load line of R_1 in B is projected over to H, which constitutes the first point of the over-all load line. Then CA is drawn so that

$$\angle CAO = \cot^{-1} \frac{L}{\Delta t} \tag{66}$$

whereupon $CJ = \Delta i_{L_1}$ and $BC = i_{R_1}$. Then C is projected over horizontally to E directly over D, where $OD = \Delta e_1 + \Delta e_2$. The finite operator curve is shifted over so that it passes through E and intersects the load line for R_1 in G. Then FE is drawn parallel to CA, FK is Δi_{L_2}, and FG represents i_{R_2}. Also, point I is the second point of the over-all load line. The finite operator curve is now shifted over so that it passes through point N, and its intersection L with the load line for R determines the next point M of the over-all load line. In this way the successive points of the latter can be determined.

If R_1 is a linear resistance, a modification in the graphical construction can be employed that will eliminate the need for shifting the curved finite operator curve but will shift the straight load line for R_1. However, it is just as simple to cut out a template for the curved finite operator curve, and therefore the alternative construction will be omitted.

13. Application of Graphical Constructions to Triode.—The above constructions can be applied to a triode (or multigrid tube if all but one grid are at constant potentials), with the only restriction, of course, that the connected circuits can pass direct current. In the case of a triode we note that the impressed voltage e equals E_{bb}, the generated plate supply voltage, and is therefore constant. The instantaneous resistance of the tube, however, is a function of the grid voltage. If the latter is a known function of time, then the approximate value of the instantaneous resistance is known during any time interval Δt, and the graphical construction is therefore possible. To illustrate the application, let the triode characteristics be those shown in Fig. 62. Suppose an inductance L and resistance R are in series with the plate and the B voltage E_{bb}. There are really two transients, (1) when the plate supply switch is first closed (establishment of initial d.c. component), and (2) when signal voltage $e_s = f(t)$ is first impressed in the grid circuit in addition

to the normal bias voltage E_c. Ordinarily we are not interested in the transient set up in (1); hence we determine the initial d.c. component by merely drawing the load line for the resistance R of the load, in the well-known manner. The load line for (2) is the one we now wish to determine. We proceed as follows:

From the plate voltage E_{bb} we draw the load line for R as shown in the figure, line AE_{bb}. Its intersection B with the bias voltage curve E_c determines the initial d.c. component.

The instantaneous signal voltage e_s is now determined for each time interval Δt, and the corresponding curve of the plate family

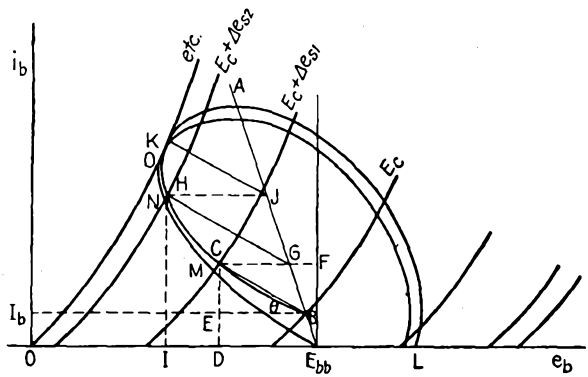


FIG. 62.—Path of operation for triode with inductive load.

used. As shown in Fig. 62, e_s is assumed to be sinusoidal and to start from its zero value. Only these curves of the plate family have been plotted, as shown, and in the figure large time intervals Δt have been taken in order to make the figure clearer. If the grid voltage in the first time interval Δt changes from E_c to $E_c + \Delta e_{s_1}$, the current rises along BC, where

$$\theta = \cot^{-1}\left(\frac{L}{\Delta t} + R\right) \tag{67}$$

and CE represents Δi_1 and CD the total current i at that instant. The procedure is exactly the same as that followed in Sec. 6. The point C is projected over to the E_{bb} ordinate as F, and the latter is really the point on the over-all load line, all points of which will be along this E_{bb} ordinate. However, in the case of the triode we prefer to call such points as C, H, K, etc., the over-all load line, in which case our meaning is "plot of functional relation of plate current vs. grid voltage," rather than "actual energy-supplying

voltage E_{bb} impressed in the circuit." From the bottom of F, or point E_{bb}, a line is drawn at the angle whose cotangent is R—*viz.*, $E_{bb}A$. This line intersects the current abscissa through C and F in G. Through G we draw GH parallel to BC, and HI is the next instantaneous value of current.

In the case of a triode, the procedure just outlined is unnecessary because the impressed voltage E_{bb} is constant, whereas in Sec. 6 every step is necessary. Here we need merely project the points over to the load line for R and then draw the finite operator curve through the projected point over to the next value of the grid parameter. Thus H is projected over to J and JK drawn parallel to CB. Then K is the next point of the load line, and the rest are found in the same manner.

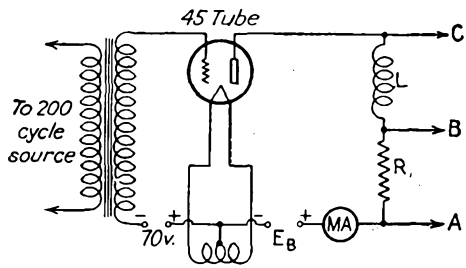


FIG. 63.—Experimental setup to check graphical construction of type 45 triode with inductive load.

It will be noted from Fig. 62 that if cutoff of the plate current occurs during the negative portion of the grid swing, at point L, the curve will continue to the left along the e_b axis to the point E_{bb} and will rise from there for every cycle of grid swing thereafter. Hence, if the steady-state solution is desired and it is foreseen that cutoff will occur, the construction can be started at E_{bb} and the initial loop $BCHK$ ignored. This is a fortunate reduction in the amount of work necessary in a very nonlinear case, that of operation beyond cutoff.

In analyzing the theory of this application, we may regard the triode as a resistance that varies both with current and time in a determinable manner, while the impressed voltage remains constant at the value E_{bb}. On the other hand, we may regard the construction as the discrete projections on the e_b-i_b plane of the intersection of a shifting finite-operator surface with the tube surface.

14. Experimental Verification.—Experimental setups have been made to check some of the circuits described in the preceding

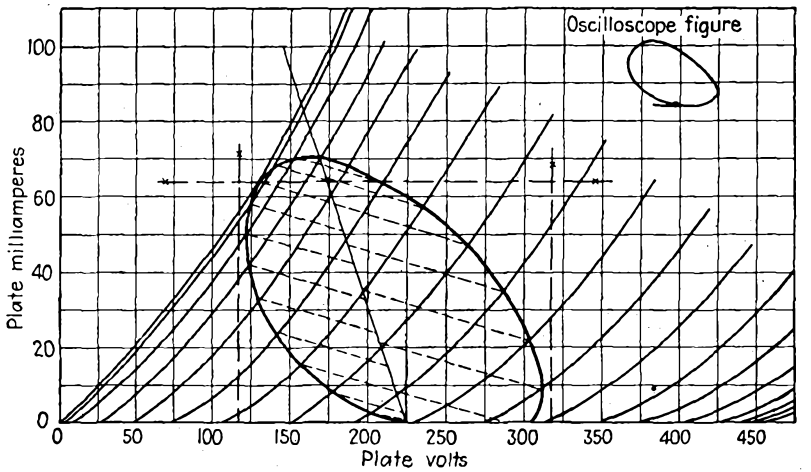


FIG. 64.—Path of operation for type 45 triode—graphically and experimentally—225-volt plate supply.

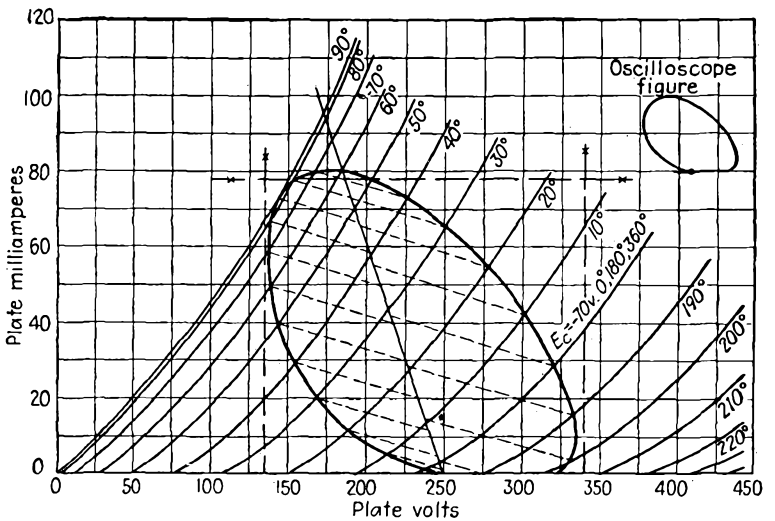


FIG. 65.—Path of operation for type 45 triode—250-volt plate supply.

sections. One such is shown in Fig. 63. A 45 tube was measured for its plate-family characteristics. Thus calibrated, it was connected to an inductance L, of 1 henry, and a resistance R as shown.

The total resistance of the load circuit was adjusted to 800 ohms. The grid was connected to a beat-frequency oscillator set to 200 cycles and the grid signal voltage, of sinusoidal wave shape, adjusted so that its peak value was 70 volts and therefore equal to the bias. In this way, as large a grid swing as possible was obtained without the grid being driven positive and thus contributing distortion products to the plate-current wave shape. The plate voltage E_{bb} was adjusted to various values as shown on the

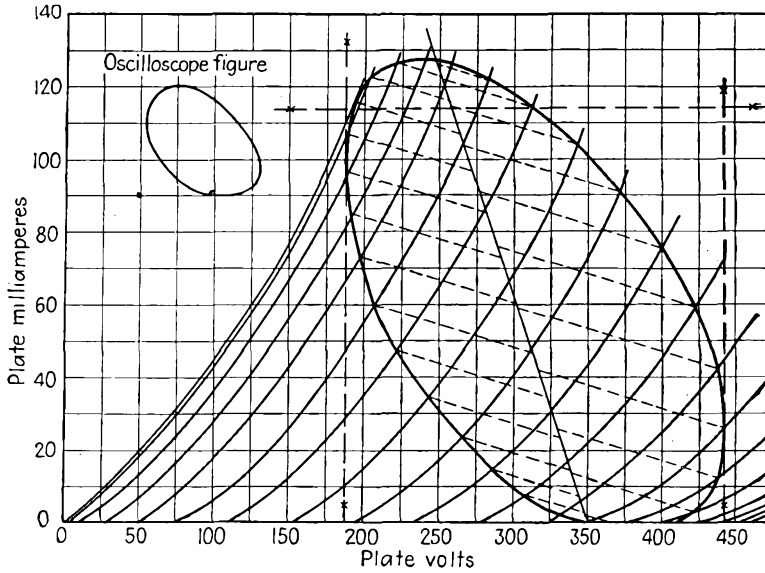


FIG. 66.—Path of operation for type 45 triode—350-volt plate supply.

accompanying graphical diagrams (Figs. 64 to 66), and wires A and B were connected to the vertical deflection plates of a cathode-ray oscilloscope, while A and C were connected to the horizontal deflection plates. In this way, a figure was obtained on the screen that would correspond to the graphical construction for these operating conditions.

The results are shown in Figs. 64 to 66. The small figure on each graph represents a copy of the oscilloscope trace. It will be noted that qualitatively the two correspond, although the trace shows a transient that the author feels originated in the oscilloscope. Quantitative checks were not possible, and investigation showed that the small figure of thick line on the oscilloscope could

hardly serve for accurate comparison. Moreover, it was found that the deflection plates were not quite at right angles to one another. Hence a crest voltmeter was used to measure minimum and maximum plate voltages. These checked fairly well with the graphical construction, and the values are shown on the diagram by crosses.

A brief summary of the graphical method will be given. The procedure is as in Sec. 6, with the difference that a triode is used here. The grid voltage wave is $e_s = 70 \sin 400\pi t$.

This wave is broken up into 10-deg. intervals, so that

$$\Delta t = 1/7{,}200 \text{ sec.}$$

Then $L/\Delta t = 7{,}200$, and $R + (L/\Delta t) = 8{,}000$. In drawing a finite operator curve, it must be remembered that the cotangent of the angle of inclination is 8,000 only if the units of current and voltage are amperes and volts, respectively, and the scale divisions are equal. In the graphs shown, the rise is 10 divisions for every 32 divisions horizontally for $R + (L/\Delta t)$ and 50 divisions up for 16 divisions horizontally for $R(= 800$ ohms). In these examples, the current cuts off during a portion of the cycle; hence, steady-state over-all load lines can immediately be drawn by starting the construction at $i_b = 0$ and $e_b = E_{bb}$ as shown.

In comparing the results, it was found that the maximum deviation between the construction and the peak voltmeter readings was approximately 4 per cent for plate voltage. A study of the figures will show, however, that the readings are probably in error, particularly the peak current readings. Thus, we know from theoretical considerations that the load line must start on the plate-voltage axis at E_{bb}; that is, $i_b = 0$, $e_b = E_{bb}$ is one point on the curve, since cutoff occurs. Also, the load line must be tangent to the zero grid-voltage curve. If it is attempted to draw a load line through these two points and the peak current and voltages are experimentally determined, it will be found that the load line does not at all resemble the figure of the oscilloscope, which at least qualitatively checks the graphical construction.

Another check for the latter is as follows: If the load line is replotted as a plate current–time wave and analyzed for its d.c. component, it will be found that it checks the d.c. plate milliammeter very closely. It is felt that this check, together with

the fairly close maximum and minimum plate-voltage peak voltmeter readings, constitutes sufficient proof of the correctness of the method and the construction. Of course, some error is to be expected in the latter since finite rather than infinitesimal time intervals are used.

The second example is that of a dynatron circuit. The construction of Sec. 11 is to be used. The circuit and value of parameters employed are shown in Fig. 67. The inductance had an ohmic resistance of 500 ohms, so that the total resistance of that branch was 800 ohms. The screen voltage was maintained constant at 130 volts, and the plate supply voltage at each of three values, 45, 22.5, and 13.5 volts.

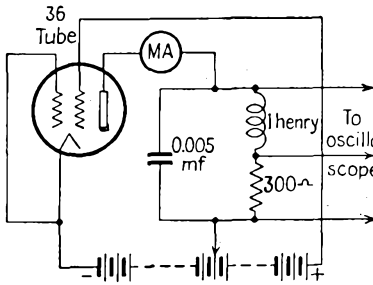


FIG. 67.—Dynatron oscillator circuit using type 36 tetrode.

A simplification of the construction (Fig. 68) of Sec. 11 is possible because of (1) zero resistance in the capacitive branch and (2) constant applied voltage at one of the three given values. The dynatron characteristic is represented by r and the plate supply voltage by E_0. Assume that this can be applied in such manner

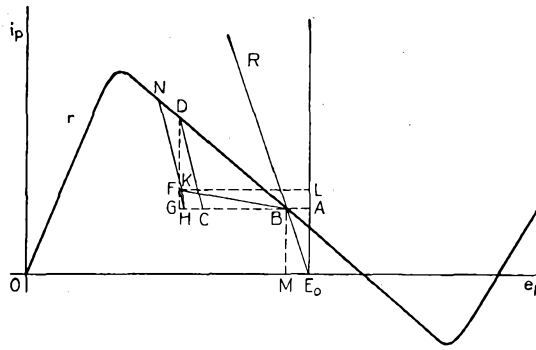


FIG. 68.—Graphical construction for dynatron oscillator circuit.

that no oscillations start, so that a steady current BM flows as determined in the well-known manner by R. This is our initial condition. Then BA represents the voltage drop in R due to BM. There are two equal voltages present in the parallel branches of the circuit, each of value $AB = (BM)R$. By the methods out-

lined previously, they may be replaced by two equivalent voltages appearing in series with E_0, *viz.*, $AB(Z_t/Z_L)$ and $AB(Z_t/Z_c)$. The former represents the series voltage drop equivalent to the actual voltage across R and the latter the series voltage drop equivalent to the actual voltage across C, due to its stored charge. If these two equivalent voltage drops are added algebraically to E_0, we obtain $E_0 - AB[(Z_t/Z_c) + (Z_t/Z_L)] = E_0 - AB$. Thus, if we attempt to proceed with a graphical construction from our initial condition by proceeding as in Sec. 11 and draw the finite operator curve for Z_t from $E_0 - AB$ to r, we find that Z_t starts from point B on r, and thus our construction does not move from its initial position B. This means that the circuit is in equilibrium and must receive a shock impulse in order to start oscillating. Suppose this is accomplished by imparting an additional charge on C and thus changing the voltage across the latter by an amount E_c. The equivalent voltage in series with E_0 is $E_c(Z_t/Z_c)$ and is represented (Fig. 68) by BC. We can now proceed. From C, CD is drawn to represent Z_t. Then DG represents $i_c + \Delta i_L$, and GA represents the voltage E across the parallel resonant tank circuit. From B, BF is drawn to represent Z_L. Then FG is Δi_L and FD is i_c; and it is noted that, owing to the negative resistance characteristic of r, FD represents a charging current into the condenser C. Two equivalent series voltages must now be added algebraically to C, *viz.*, $\Delta i_L(RZ_t/Z_L)$ and i_cZ_t. This can be done by drawing FH parallel to DC and then HK to represent the finite operator curve RZ_t/Z_L. Then K is the next point from which Z_t can be drawn to r to get the next values of i_c and Δi_L.

The construction can be further simplified in two ways:

1. From F a finite operator curve of value $Z_t[1 - (R/Z_L)]$ can be drawn to GA and the intersection projected up to FL to obtain point K. This replaces two finite operator curves FH and HK with one. Thus, for each point, three finite operator curves Z_t, Z_L, and $Z_t[1 - (R/Z_L)]$ are required.

2. In the examples to be presented, Z_L comes out such a high value that its operator curve is nearly parallel to the voltage axis. Hence, Δi_L (or FG) and FK were computed directly upon a slide rule to obtain greater accuracy. Thus, after CD is drawn, voltage GB can be read off from the graph. Then this voltage can be divided by Z_L and the quotient Δi_L laid off graphically as GF. Furthermore, Δi_L is multiplied by $Z_t[1 - (R/Z_L)]$ and the

product laid off graphically as FK. Then from K a line KN, parallel to CD and thus representing Z_t, is again drawn to r, and the process is repeated. The construction thus becomes one graphical and two slide-rule manipulations for each point and represents a welcome saving in construction lines in the solution. We thus use the graphical construction where it is indispensable—to find $i_C + \Delta i_L$ or the intersection of the finite operator curve Z_t with the irregular curve representing the load line or characteristic for r.

The characteristic to be plotted here is the current in the inductive branch vs. the tank voltage E. Point F in Fig. 68 represents one such point.

In the example of Fig. 67, the finite operator values were calculated for a Δt equal to 1/72,000 sec. Then

$$Z_L = \frac{1}{1/72{,}000} + 800 = 72{,}800$$

$$Z_C = \frac{1/72{,}000}{5 \times 10^{-9}} = 2{,}777$$

$$Z_t = \frac{72{,}800 \times 2{,}777}{72{,}800 + 2{,}777} = 2{,}675$$

and

$$Z_t\left(1 - \frac{R}{Z_L}\right) = 2{,}675\left(1 - \frac{800}{72{,}800}\right) = 2{,}650$$

These values are then divided by 1,000 to obtain the proper cotangents when the current scale is in milliamperes.

The characteristic for a plate voltage of 45 volts is shown in Fig. 69. It will be noted that there is a close similarity between the graphical construction and the figure obtained on the oscilloscope. In the former, the initial charge assumed was insufficient to give a closed loop in one cycle, but it was obtained in the second excursion around the dynatron characteristic. The number of points in the closed loop indicate the frequency in that they are 1/72,000 sec. apart and total to one period. Thus there are 35 points, and the corresponding frequency is

$$\left[\frac{1}{35(1/72{,}000)}\right] = 2{,}060 \text{ c.p.s.}$$

The experimentally determined value was 2,100 c.p.s. and is in good agreement.

From the points of the characteristic curve (load line for inductive branch) a time wave can be drawn. This was done and analyzed for its d.c. component. The value obtained was 0.79 ma. The experimental value read on the d.c. milliammeter varied from 0.75 to 0.8 ma. during the test. In reference to this it is to be noted that the dynatron characteristic varied over any

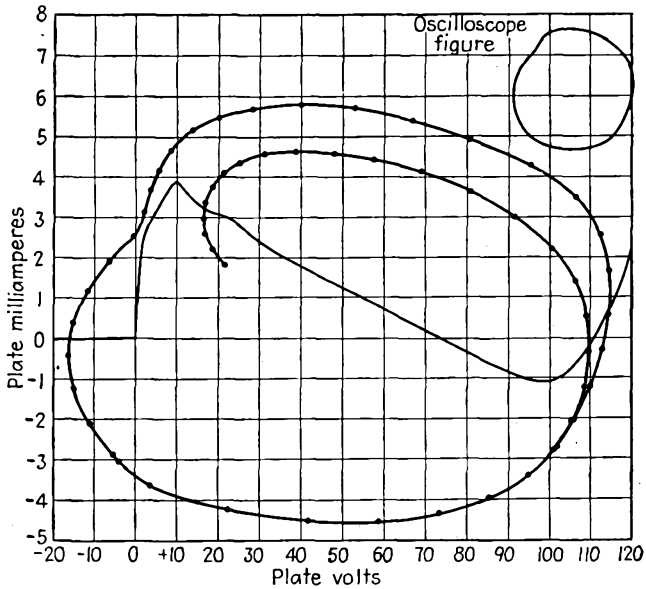


FIG. 69.—Path of operation for dynatron oscillator—45-volt plate supply.

appreciable period of time and was due apparently to decrease in secondary emission with time. Hence it was felt that only moderate agreement could be expected between the graphical and experimental values. The graphical figure was checked quantitatively against the oscilloscope figure after the latter was calibrated; and although good agreement was obtained, it was decided that the precision of measurement was too low to warrant this check as a means of confirmation of the graphical method.

In Fig. 70 is shown the characteristic for a plate supply voltage of 22.5 volts. There is a close similarity between the graphical and oscilloscope figures, particularly the cutin at the left-hand side. The frequency (graphical) was 1,750 c.p.s. (corresponding to 41 points), and the experimental value was the same. The d.c.

component was 1.42 ma. (graphical) as compared with 1.3 ma. average (experimental).

In Fig. 71 is shown the characteristic for a plate supply voltage of 13.5 volts. The oscillations are comparatively feeble, and

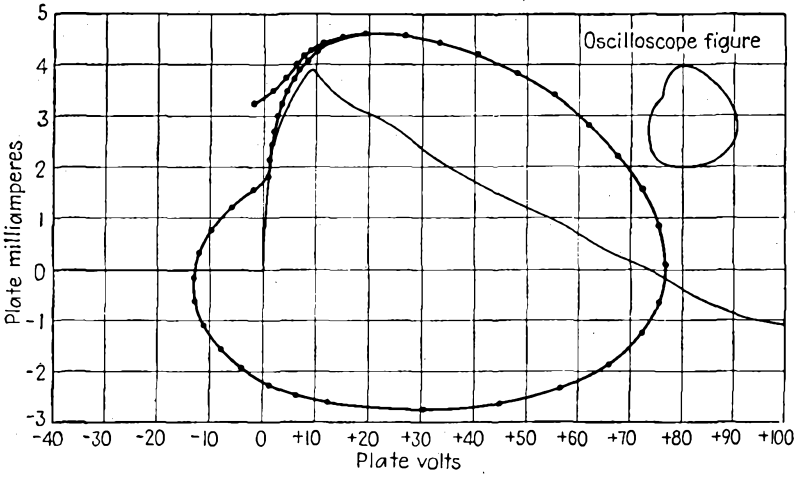


FIG. 70.—Path of operation for dynatron oscillator—22.5-volt plate supply.

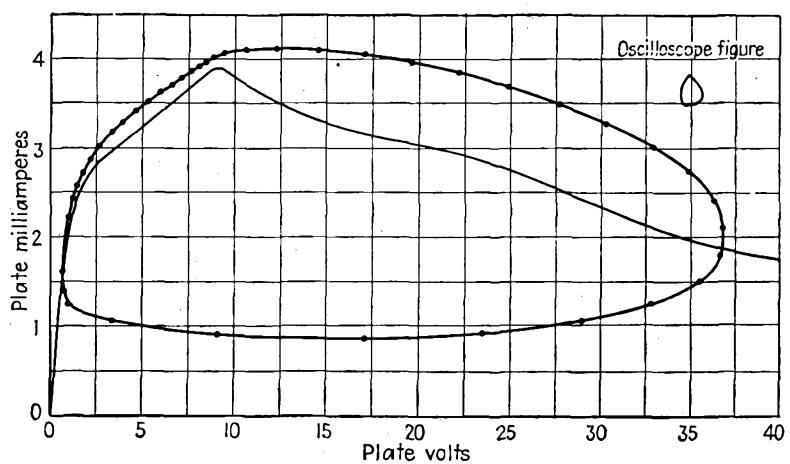


FIG. 71.—Path of operation for dynatron oscillator—13.5-volt plate supply.

hence it was found advisable to enlarge the scale of the graph. The frequency was 1,567 c.p.s. (graphical), corresponding to 46 points, as compared with 1,580 c.p.s. (experimental). The d.c. component was 2.79 ma. (graphical) as compared with 3.04 to

3.28 ma. (experimental). Although the agreement is poor, it is to be expected in view of the weak oscillations. No attempt was made to use a crest voltmeter in these runs as it was found that the capacitance of the leads and meter had an appreciable effect upon the characteristic. In the main, it was felt that good experimental verification of the graphical method had been obtained, particularly for the larger amplitudes of oscillation.

Interesting and very simple graphical constructions and results can be obtained from the method just outlined if some of the circuit parameters are allowed to approach zero or infinity. Thus, if $L = 0$, we have a resistance R by-passed by a condenser C; and if the nonlinear resistance r is that of a triode, which varies in known manner with time (signal voltage on grid known), then a transrectification diagram can be made if the triode be adjusted to act as a grid-bias detector. Lack of space precludes any further discussion of these matters; it may merely be noted that some of the results mentioned by R. Usui[1] can be verified by the construction given above.

15. Conclusions.—By breaking up a derivative ratio into two parts and using one part as a finite operator curve, a graphical method of construction has been developed of wide scope and comparatively simple manipulation. In contrast to the usual method of isoclines,[9] each lineal element (starting with the initial point) helps to determine the next one, so that no visual judgment is required in choosing these to blend into a smooth curve.

Questions can be raised as to the smallness of time intervals required, the possibility of cumulative error in proceeding from point to point, and the effects of discontinuous variations in the nonlinear elements. With regard to the latter, it is to be noted that the time interval can be decreased in the neighborhood of a discontinuity and increased again in the more uniform portions of the characteristic.

BIBLIOGRAPHY

1. Usui, R.: A Fundamental Concept for Oscillators, *Nippon Elec. Communication Engineering*, September, 1935. In English. Reference to other papers is made in this article.
2. Chaffee, E. L.: "Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., 1933.
3. Maloff, I. G.: New Methods of Solution of Vacuum Tube Problems, *Broadcast News*, November, 1933.

4. WEAVER, K. S.: *QST*, November, 1929.
5. SMYTH, C. N.: Electro-acoustical Problems, *Wireless Engineer*, October, 1935.
6. MOUROMTSEFF, I. E., and KOZANOWSKI, H. N.: Analysis of the Operation of Vacuum Tubes as Class C Amplifiers, *Proc. I.R.E.*, July, 1935.
7. DOHERTY, R. E., and E. G. KELLER: "Mathematics of Modern Engineering," vol. I, pp. 163–188.
8. LIENARD, A.: Etude des oscillations entretenue, *Rév. Gén. elec.*, Tome **23**, 901–946, 1928.
9. KIRSCHSTEIN, F.: Uber ein Verfahren zur graphischen Behandlung elektrischen Schwingungsvorgange, *Arch. Elek.*, **24**, 731, 1930.

CHAPTER V

BALANCED AMPLIFIERS

1. Introduction.—The *balanced amplifier*, or (as it is more popularly known) *push-pull circuit*, was invented by Colpitts, but the use of this circuit dates back to the development of the double-button carbon microphone. At this time its ability to cancel out the even-order modulation products of two nonlinear resistances was first recognized.

It was first developed for class A operation of vacuum tubes and, in inverse connection, as a frequency doubler. Later on, its suitability for class B and AB operation in audio work began to be appreciated. Today this circuit is finding wide application, not only in these fields, but also in the fields of detection, modulation, etc.

The correct analysis did not appear until a considerable period of time had elapsed after the introduction of the circuit. Two of the earliest expositions were given by Kilgour[4] and Thompson.[5] The physical analysis in Sec. 2 is based on an unpublished report by Dr. C. J. Travis, dated Aug. 18, 1932.* Much new material, however, has been added.

The balanced circuit may be resistance or transformer coupled. In the former case, it functions as two single-side amplifiers connected in series cumulative for odd-order modulation products and series opposition for even-order products. In the case of the transformer-coupled type, however, an additional factor enters in—the coupling between the plate circuits of the two tubes through the mutual inductance existing between the two halves of the primary winding. The analysis is therefore more involved and will be treated in detail here.

2. Physical Analysis.—Figure 72 shows the push-pull type of circuit but with an output choke instead of an output transformer. This may be considered a 1:2 autotransformer and for the present

* The author is indebted to the RCA License Division for permission to include this material.

may be regarded as an ideal transformer, *i.e.*, as having unity coupling, zero ohmic resistance, negligible magnetizing current, and no distributed capacity. An actual transformer approximates the ideal transformer over a frequency range depending upon the excellence of its design. The load resistance R_L is called the *plate-to-plate resistance* and is the reflected value that an actual load across an actual output-transformer secondary presents to the two tubes.

When a signal voltage e_g' is impressed across the primary of the input transformer, it induces equal and opposite voltages e_g and $-e_g$ in the two halves of the secondary. Assume the signal

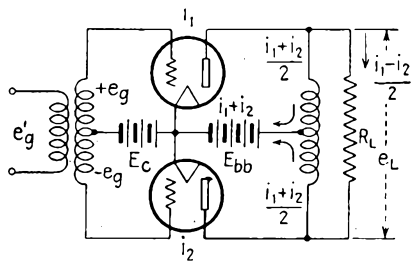


FIG. 72.—Idealized push-pull circuit.

voltage acting in the grid circuit of the top tube is in a positive direction and thus opposite to the bias E_c, while that in the grid circuit of the bottom tube is in a negative direction and thus additive to E_c. As a result, the top-tube current i_1 increases, while the bottom-tube current i_2 decreases.

Previously, when only the bias and power-supply voltages, E_c and E_{bb}, respectively, were acting, the two currents i_1 and i_2 were equal and flowing in opposite directions through the output choke. Now, when i_1 and i_2 are varying in opposite directions from their initially equal values, the output choke will allow the currents in its two half windings to vary only if they are equal to each other at all times. It is clear from the figure that the power supply must carry the sum of the two currents $(i_1 + i_2)$ at all times; hence, each half of the output choke carries half of this current, or $(i_1 + i_2)/2$, as shown. If i_1 exceeds this amount, the difference, or

$$i_1 - \frac{i_1 + i_2}{2} = \frac{i_1 - i_2}{2} \qquad (1)$$

flows through R_L to the bottom of the choke and thence combines with i_2 to flow up through the lower half of the choke to the power supply and thence back to the two tubes. If we add these two components, we obtain

$$i_2 + \frac{i_1 - i_2}{2} = \frac{i_1 + i_2}{2} \tag{2}$$

which satisfies our initial assumption of equality of the two currents through the two halves of the choke.

The current through R_L is the load current and was found to be $(i_1 - i_2)/2$. It sets up a voltage e_L across R_L of value

$$e_L = \frac{i_1 - i_2}{2} R_L \tag{3}$$

From the figure it is evident that half this voltage appears in the plate circuit of the top tube in a direction opposing E_{bb}, while the other half appears in the plate circuit of the bottom tube as additive to E_{bb}. The two plate voltages therefore vary in opposite directions from their normal value E_{bb}.

If we regard the arrangement as a four-terminal network, we can draw the following conclusions concerning its performance:

1. An alternating signal voltage is impressed across its input terminals, whose magnitude may be considered as $2e_g$, a grid-to-grid voltage.

2. An alternating output current $(i_1 - i_2)/2$ flows through the load R_L connected to its output terminals. This current may be considered a plate-to-plate current.

3. An alternating output voltage e_L appears across its output terminals, and this voltage may be considered a plate-to-plate voltage.

Thus the input and output voltages and current are alternating in character, and the direct voltages E_{bb} and E_c are not apparent externally. Unfortunately, our primary information is that concerning the two tubes individually and is in the form of a family of curves for each, which, for similar tubes, are identical. In addition, the foregoing analysis has provided this further information:

1. The grid voltages vary oppositely from their common bias value, and to equal degree, when an input signal is impressed.

2. The plate currents vary oppositely from their common d.c. value, and in the same direction as the respective grid voltages, but not necessarily to equal degree.

3. The plate voltages vary oppositely from their common value E_{bb}, to equal degree, but in directions opposite to their respective grid voltages.

3. Graphical Application.—The foregoing is sufficient for a graphical analysis of the performance of the circuit. Thus, as shown in Fig. 73, we lay off our d.c. components at the power-supply voltage E_{bb}, just as in the case of a single-tube stage. The two values are represented by I_b. Now assume that a signal voltage $2e_g$ is impressed (grid-to-grid). One grid changes to a value $E_c + e_g$, the other to $E_c - e_g$ (where E_c is almost invariably inherently negative). For some value of R_L, the plate voltage for the tube whose grid voltage is $E_c + e_g$ will drop by an amount Δe_p, while for the other it will rise by the same amount Δe_p. Lines projected up from the axis to the respective grid voltages at these two points represent i_1 and i_2, respectively. Then $i_1 - i_2$ represents twice the load current flowing through this particular value of R_L. The latter's magnitude can be found from the fact that

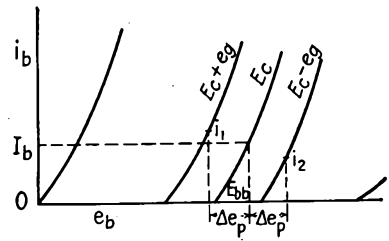


FIG. 73.—Graphical construction on plate family of tube characteristics.

$$\frac{R_L}{2} = \frac{2\,\Delta e_p}{i_1 - i_2} \tag{4}$$

For some other value of R_L, Δe_p will be different (hence also i_1 and i_2). In particular, for a higher value of R_L, Δe_p will be greater, and—as can be found by trial from the figure—$i_1 - i_2$ will be smaller. For R_L infinite, Δe_p will have some finite value, and $i_1 - i_2$ will be zero, which checks with Eq. (4). For R_L zero, $i_1 - i_2$ will have some finite value, but Δe_p will be zero. If Δe_p is chosen greater than the value for R_L infinite, $i_1 - i_2$ will come out negative. This means that R_L is now negative, *i.e.*, a source of energy, and is of no practical importance. In this way, successive values of $i_1 - i_2$ may be had for corresponding values of $2\,\Delta e_p$, while $2e_g$ is the parameter.

If we now make the latter assume some other values, so that half of it represents a departure from E_c in one direction and

the other half represents an equal departure from E_c in the opposite direction, we can repeat the process outlined in the preceding paragraph and obtain a new set of values for $i_1 - i_2$ and $2\ \Delta e_p$. This can be repeated until the desired range of $2e_g$ is covered.

A point worthy of note is that, for small values of Δe_p (low values of R_L) and large values of the parameter $2e_g$, the i_b-e_b curve for $E_c - e_g$ will strike the e_b axis at a point higher than $E_{bb} + \Delta e_p$ so that apparently i_2 cannot be determined. It must be remembered, however, that no curve of the family ends on the e_b axis; each continues from that point to the left along the axis. Hence we see that the above values of $2\ \Delta e_p$ and $2e_g$ merely mean that i_2 is zero under those conditions. This brings out the fact that, if the grid swing is great enough or R_L sufficiently small, each tube cuts off during alternate half cycles. This mode of operation may be defined as *class AB* (sometimes called *class A prime*). If $2e_g$ is not excessive or if R_L is sufficiently great, neither tube's current cuts off, and this may be considered class A. In addition, we must note that cutoff also depends upon the operating point, which we shall for the moment assume is determined by I_b.

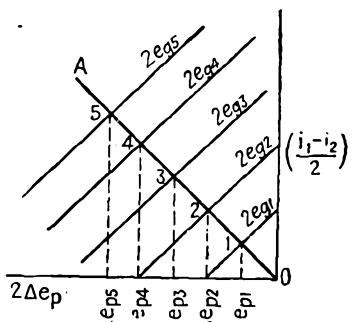


FIG. 74.—Push-pull family of characteristics.

The sets of values of $i_1 - i_2$ and $2\ \Delta e_p$ for various values of the parameter $2e_g$ may now be plotted on a separate sheet of paper, as shown in Fig. 74. These give rise to a family of curves that may be considered the characteristics for the balanced circuit and that corresponds to the families for the individual tubes. The difference, as noted previously, is that this family is for alternating voltages $2e_g$ and $2\ \Delta e_p$ and alternating currents $(i_1 - i_2)/2$, since the balanced circuit or four-terminal network is responsive to these only. We see that, the larger $2e_g$, the greater both $2\ \Delta e_p$ and $(i_1 - i_2)/2$ for a given load resistance R_L. It is also to be noted that for each set of operating values E_c and E_{bb} a different family is obtained. As shown in the figure, the curves occupy the second quadrant, which is the only one of practical interest in the case of a passive resistive load. For reactive loads

all four quadrants are involved owing to the energy storage and discharge in and from this type of load during different portions of the cycle of grid swing (assumed sinusoidal). The latter type of load, however, presents too many difficulties to be discussed here.

In a subsequent section it is shown that, for parabolic tube characteristics, the curves of the above family are straight lines. For actual tubes, however, they may not be; but if they are essentially parallel to one another, at least over a certain range, and equidistant, the distortion products for a suitable value of R_L will be small. For class A operation, the optimum value of R_L is

$$R_L = 2R_p \tag{5}$$

where R_p is the plate resistance at the operating point of either tube, the latter being assumed to have parabolic characteristics. For class AB operation a good value is

$$R_L = 4R_p \tag{6}$$

where R_p is the plate resistance of either tube at peak positive grid swing and for a value of tube current in the neighborhood of that to which it will rise for this value of R_L. Although this means that we have defined R_L in terms of itself, it will be found that R_p in this range does not vary much; moreover, the power output is not materially changed with nominal variation in R_L from the value set by Eq. (6).

In case the positive grid swing is such as actually to drive the grid positive with respect to the cathode (sometimes called *class* AB_2), then a new consideration enters in: the minimum plate voltage at that instant should be from two to three times the amount by which the grid is positive with respect to the cathode. Since the minimum plate voltage is given by

$$e_{b\min} = E_{bb} - E_p \tag{7}$$

and the positive grid voltage e by the algebraic sum

$$e = E_c + e_g \tag{8}$$

we can determine for $2e_g$ what $2\,\Delta e_p$ should be. This in turn will determine R_L.

Equation (4) indicates that the relation between $(i_1 - i_2)/2$ and $2\,\Delta e_p$ is linear. Hence, the load line for R_L is a straight line

on Fig. 74 and passes through the origin. It may be determined in exactly the same manner as that for single-tube operation and is shown in Fig. 74 as OA for the desired value of R_L. The intersections of AO with the various curves give the load current $(i_1 - i_2)/2$ and voltage $(2\ \Delta e_p = E_L)$ for various values of grid swing $2e_g$ and for the chosen value of R_L. We can now plot $(i_1 - i_2)/2$ vs. $2e_g$ and obtain the push-pull dynamic characteristic, as shown in Fig. 75. Corresponding points here are labeled as in Fig. 74. If the tubes are suited for this mode of operation and the operating point is satisfactory, the characteristic obtained will be straight, or nearly so. Actually, it occupies both the first and third quadrants, although it is evident that the plot in the third quadrant is the same as that shown in the first quadrant, only inverted, and hence has been omitted in the figure. As a result, only odd-order terms are present in the power series for the characteristic, and thus the circuit has eliminated the even-order terms in its output, which for a sinusoidal signal voltage means suppression of the even harmonics. The plot $(i_1 - i_2)/2$ vs. time may now be made if the wave shape of $2e_g$ is known or assumed and the former wave shape then analyzed for its sinusoidal components.

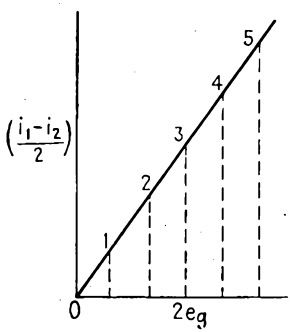


FIG. 75.—Transfer, or output, characteristic for balanced amplifier.

We can now plot the relation between the individual tube currents and their respective grid swings on Fig. 73 and thus determine the load line of R_L on each tube characteristic. Referring to Fig. 74, we see that when the grid-to-grid voltage is $2e_{g_1}$ the plate-to-plate voltage is $2\ \Delta e_{p_1}$, when the grid-to-grid voltage is $2e_{g_2}$ the plate-to-plate voltage is $2\ \Delta e_{p_2}$, etc. Half of each plate-to-plate voltage is to be associated with each tube in the proper direction. Thus in Fig. 76 we repeat Fig. 73 and on it show half of $2\ \Delta e_{p_1}$, $2\ \Delta e_{p_2}$, etc., laid off on the e_b axis on either side of the quiescent voltage E_{bb}. From these points, lines are projected vertically to the curves having corresponding values for the grid parameter. In this way, we obtain points 1, 1, 2, 2, 3, 3, etc. These represent at any instant the magnitudes of i_1 and i_2 in the respective tubes or the values of the current in either tube for

corresponding moments in the alternate half cycles. According to the latter viewpoint, if we join these points by a smooth curve, we have the relation between the current in either tube and its grid voltage, *i.e.*, the load line of R_L for each tube.

In general, this load line will be curved, rather than straight, as is the case for a single tube, and (as shown in the figure) may cut off on the right-hand side of E_{bb} (just beyond point 3). For parabolic characteristics, the load line will be a parabola too, up to the point of cutoff. The reason for the curvature of the load line is that the tubes may be regarded as two generators

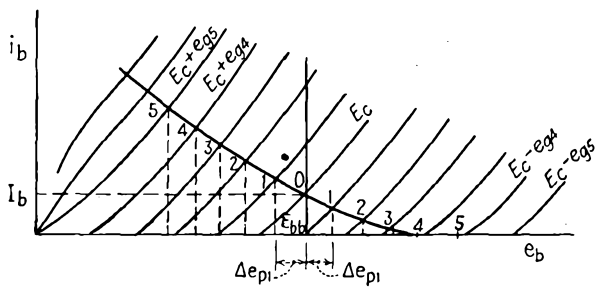


FIG. 76.—Apparent load-resistance line as viewed by either tube.

connected in parallel to the load R_L through the output choke. The tubes may be regarded as generating equal voltages, and in the same direction for the above equivalent circuit, but as having internal resistances variable throughout the grid cycle, and in opposite directions. As a consequence, the division of load throughout the cycle will be unequal (except at the operating point, where they have equal internal resistances), and hence the impedance R_L reflected to either will be variable. Specifically, at the operating point, R_L appears as $R_L/2$ to either; beyond cutoff of either tube, R_L appears infinite to that tube and $R_L/4$ to the other.

As indicated by Eq. (4), the lines joining points 1, 1, 2, 2, etc., of Fig. 76 all make the same angle θ with respect to the E_p axis, of value

$$\theta = \cot^{-1} \frac{R_L}{2} \tag{9}$$

Moreover, it will be evident from the geometry of the figure that these lines will all be bisected by the E_{bb} ordinate, that is, by the ordinate through the operating point, and projections ($2\ \Delta e_p$)

will be bisected by this ordinate too. Accordingly, Kilgour[4] has suggested sliding a rule at the angle given by Eq. (9) along the paper so that the segment intercepted by equal grid curves to either side of E_c will be bisected by the operating ordinate. The intersection of the rule under this condition with each pair of curves gives the value of i_1 and i_2 immediately on the original tube family of curves, from which all other relationships may be plotted. This method has much to recommend it when R_L is known, as it eliminates much labor; it will be used here for further work.

4. Dynamic Characteristics.—The load line of Fig. 76 may be replotted so as to exhibit the relationship between i_b and e_c.

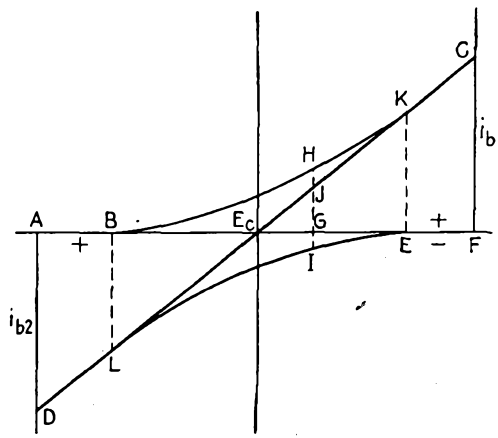


FIG. 77.—Individual and composite dynamic characteristics.

Such a plot is evidently the dynamic characteristic of either tube. This curve is shown as *ABC* in Fig. 77 for one tube. If this be then rotated first about a vertical line through E_c and then about the e_b axis, a second dynamic characteristic *DEF* for the other tube is obtained. The two curves will then exhibit the operation of the stage very clearly. A positive grid swing for one tube is a negative grid swing for the other. Thus, a positive grid swing E_cG for the top tube is an equal but negative grid swing for the bottom tube, and the corresponding plate currents are *GH* and *GI*, respectively. If the latter is subtracted from the former, *GJ* is obtained; and since this represents $i_1 - i_2$ at that instant, it also represents $2i_L$. In similar manner, other values of $2i_L$ can be obtained. For a grid swing per tube of

E_cE, the plate current of the top tube is KE, and that of the bottom tube is zero (cutoff), so that $2i_L$ becomes identical with the former tube's current from that point on. Thus, the peak value of i_L is $CF/2$. The same holds true for the next half cycle, in which BL replaces KE and AD replaces CF. The plot of $2i_L$ vs. e_c is thus DLE_cJKC and corresponds to the curve in Fig. 75 (which is drawn for the first quadrant only, however).

It will be noted that two curved dynamic characteristics can combine to give a fairly linear output characteristic, at least over the range where they overlap. Even where either tube cuts off, the departure from linearity may not be excessive, and hence the distortion products will be fairly small. In addition, if line LE_cJK blends smoothly into the portion KC and LD (is tangent to them), then the higher order distortion products will be relatively small. Specifically, this means that for a sinusoidal input e_g the higher harmonics of i_L will be small in amplitude. The concave portions KC and LD represent portions of the cycle where $2i_L$ becomes identical with one or other of the two tubes. These portions may have the curvature shown, whereupon i_L has an upward departure from linearity, or "overshoot," or the curvature may be in the opposite direction, whereupon i_L may be said to exhibit an "undershoot." An analysis of this feature will be given later.

5. Modes of Operation.—The amount of overlap of the two characteristics affords a means of defining the mode of operation of the stage. It will depend upon the position of the quiescent point of each tube and the value of the plate-to-plate load resistance R_L, *i.e.*, upon how soon cutoff of either i_b is reached during alternate half cycles. In spite of the fact that R_L appears as a nonlinear load resistance to either tube (curved load line), the position of the cutoff point on the e_c axis is determined by the same considerations as that of a single-ended tube (see Chap. III, Sec. 15). If R_L is small, the individual load lines are steeper, cutoff is earlier in the cycle, and the overlap is less. If the quiescent point is lowered, either by decreasing E_{bb} or increasing E_c, or both, the overlap is decreased. On the other hand, a change in the quiescent point may be balanced by the proper contrary change in the magnitude of R_L so as to leave the overlap unchanged. Finally, if the grid swing is sufficiently great, the overlap will not cover the peak-to-peak grid swing.

The important point about Fig. 77 is that the tubes may cut off alternately early in their respective negative half cycles of grid swing without the output current $2i_L$ having much distortion. This is in marked contrast to single-ended operation, where premature cutoff produces strong distortion products, *i.e.*, plate current must flow during the entire grid cycle. The removal of this exacting requirement in the case of balanced amplifiers permits E_c, R_L, and E_{bb} to be selected so as to obtain increased power output for a permissible amount of distortion. The operating conditions may be selected so as to produce the following modes of operation:

1. *Class A.*—If R_L, E_c, and E_{bb} are so selected that plate current for either tube flows during the entire grid-voltage cycle, and the grids are not driven positive, then the stage is said *to operate class A*. The overlap of the two dynamic characteristics is then complete; and if these characteristics have the proper curvature, i_L will be linearly related to the grid swing e_g and the distortion will be zero. In practice, i_L is more linear for this mode of operation than for the others to be described. For maximum output in the case of triode tubes, R_L should be in the neighborhood of $2R_p$, where R_p is the plate resistance of each tube at the quiescent point. (This will be shown later.) In this case, the overlap can be complete only if the quiescent point is sufficiently high, *i.e.*, if I_b is of sufficient magnitude. For a given E_{bb}, this value of I_b determines the bias E_c. On the other hand, $E_{bb}I_b$ must not exceed the permissible plate dissipation W_{pd}; hence, for class A operation, there is an upper limit to E_{bb}. This, in turn, limits the amount of power output possible from the given tubes.

2. *Class AB.*—As has been shown, fairly distortionless output may be obtained even if the overlap is not complete (Fig. 77). In this case, plate current for either tube flows for less than 360 deg. but for more than 180 deg. of the grid-voltage cycle. This mode is known as *class AB operation.* If this can be obtained with a grid swing that does not drive the grid positive, then it is called *class* AB_1. Usually, in any case, R_L is chosen to give maximum power output. Hence cutoff may be varied by the other factor, the position of the quiescent point on the tube i_b-e_b family of characteristics, and this position can be such that class AB_1 operation is obtained by making E_c sufficiently

large relative to E_{bb}, *i.e.*, by overbiasing. Stated in inverse fashion, E_{bb} can be increased to much higher values than is permissible in class A operation, since I_b can be kept down by overbiasing to a value that will not cause $E_{bb}I_b$ to exceed W_{pd}. The peak value of i_L will then be greater, and consequently the power output as well.

To summarize, class AB operation permits R_L to be chosen without reference to cutoff considerations and E_{bb} to be chosen without limitation as to plate dissipation. As a consequence, the power output will be greater than for a single-ended stage required to operate class A and subject to limitations as to the value of R_L and E_{bb}.

As an example, a single 2A3, which must therefore operate class A, can have a maximum value of E_{bb} of 250 volts. To keep within the permissible W_{pd} (at no signal) of 15 watts, I_b must be kept down to 60 ma., which in turn means a bias of -45 volts. The maximum power output employing an $R_L = 2{,}500$ ohms and for 6 per cent distortion is then 3.5 watts. Two 2A3 tubes in push-pull can have a value of $E_{bb} = 300$ volts, $I_b = 40$ ma., $E_c = -62$ volts, R_L (plate-to-plate) = 3,000 ohms. The maximum power output is 15 watts, and the distortion is 2.5 per cent. In this particular case, it will be found that the overlap is practically 100 per cent, so that the balanced stage may also be regarded as operating class A. However, it will be found that the dynamic characteristics are so curved that single-ended operation for the operating conditions chosen would result in objectionable distortion, and hence the preceding values must be used. Push-pull operation thus permits the very extreme of class A operation to be attained for this tube, and it will be observed that the power output is considerably more than double that for comparable single-ended operation. If the limit of insulation strength on the negative half cycle of either tube is not exceeded, then even higher values of E_{bb} and true class AB_1 operation with even greater output are possible.

Another subclassification of class AB is class AB_2, which refers to that mode of operation in which the grid swing is sufficient not only to produce early cutoff of the plate currents but also to drive the grids positive as well. More will be said about this later; at present it will be noted that even greater output is possible in this case than for class AB_1.

3. *Class B.*—If the two dynamic characteristics just fail to overlap, a mode of operation known as *class B* is obtained. This mode is produced by biasing the tubes to cutoff and is the limit case of class *AB*. It is evident that the plate current for each tube flows for only 180 deg. of the cycle in alternate sequence. It is also clear that $2i_L$ is identical first with the plate current of one tube during one half cycle and then with that of the other during the next half cycle. Hence, if the two dynamic characteristics are curved, distortion will be present in the output. Consequently, distortionless operation for true class *B* is possible only if the two tubes are linear, for only in this case will the two dynamic characteristics be straight lines.

In actual practice, the dynamic characteristics are fairly straight except near cutoff. The output current i_L can then be made fairly linear with e_g if a slight overlap of the two characteristics is permitted, so that their curvatures will cancel out. This is shown in Fig. 78. For a suitable bias E_c, corresponding to point *B*, a small direct plate current flows through either tube, *HB* and *BI*, respectively, and the respective cutoffs are *F* and *G*. Within this interval of overlap, $2i_L$ is given by *ABC*, and it may be fairly straight. Beyond a grid swing of *BG* or *FB*, $2i_L$ becomes identical with *CD* ($= i_1$) or *AE* ($= i_2$), respectively. If *CD* and *EA* are fairly straight and tangent to *ABC*, then higher order as well as lower order distortion products are relatively small.

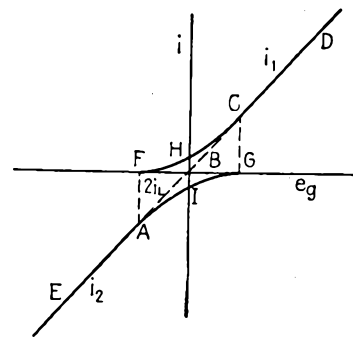


FIG. 78.—Dynamic characteristics for extreme class *AB* (class *B*) operation.

Within the interval *FG*, the dynamic characteristics *FHC* and *AIG* are the result of push-pull operation, and the sliding-rule method of determining these is theoretically necessary. Without this interval, however, one tube or the other is inoperative, so that R_L appears as $R_L/4$ to the active tube (the output choke acts as a 2:1 step-down transformer from R_L to either tube). Hence, over that portion of the cycle where only one tube is operative, the load line, as drawn on the i_b-e_b family, is straight and corresponds to $R_L/4$. If the overlap is small, then little

error is entailed if either dynamic characteristic is assumed to pass through point B, which means that although the operation is, strictly speaking, class AB_1 it may be regarded as practically class B.

Most so-called "class B stages" are really of the extreme class AB type shown in Fig. 78, but designating them as class B involves little error in describing their practical operation. An interesting example is that of a pair of type 46 or 59 tubes in a class B stage. These tubes are so designed that the plate current even at zero bias is very small; and, for the normal power outputs possible from these tubes, the grids are swung through such large voltages into the positive regions (from the above-zero-bias operating point) that cutoff occurs very early in the half cycle and true class B operation is closely approximated.

A further point is that, if $2i_L$ in the overlap region does not blend smoothly into the individual tube characteristics, then the distortion products may be an excessive percentage of the fundamental for grid swings which just carry $2i_L$ beyond the overlap region and in addition, as mentioned above, the higher order terms will be undesirably prominent.

6. Self-rectification.—Just as in the case of a single-ended stage, the d.c. component of the plate current, I_b, of each tube of a balanced amplifier generally increases with impress of grid signal voltage e_g. This effect is known as *self-rectification* and represents the manner in which the stage conforms with the *principle of the conservation of energy*, since, for the large grid swings possible, the a.c. power output would greatly exceed the original d.c. energy input; hence, the stage automatically draws a greater value of $2I_b$ and consequently d.c. energy input to cover the power output, plus an amount that represents the plate dissipation of the two tubes at full signal.

The mechanism by which the tubes draw additional d.c. component has been discussed in Chap. I (Sec. 25). If the increase in plate current during the positive half cycle exceeds the decrease during the negative half cycle, then the average or d.c. component will exceed the quiescent value; *i.e.*, self-rectification will result. In the case of the balanced amplifier this can easily occur during large grid swings, since either tube's current is driven to very high values during the positive half cycle, but only to zero on the negative half. Furthermore, low values of

R_L are possible and desirable, and these in turn produce cutoff of i_b early in the negative half cycle of e_g, which also helps increase the self-rectification. For these reasons, class AB_2 operation should exhibit self-rectification to the greatest degree.

The increase in I_b does not bear any simple relationship to e_g, and therefore the increase in d.c. input with e_g may be at a different rate from that of the a.c. power output. Hence the difference between the two, which is the plate dissipation, may be greater at full signal than it is at no signal, and a tube that is correctly biased to keep within the limits of plate dissipation at no signal may overheat at full signal. It is therefore necessary to calculate the amount of self-rectification.

While the graphical method gives the individual tube load lines and these in turn enable the d.c. component of the plate current to be determined in exactly the same manner as for single-ended operation, a method will be described that probably is quicker in obtaining the desired result. The quantity $i_1 + i_2$ represents the mid-branch current that flows through the power supply. As will be shown subsequently, it contains all the even harmonics, including the zero-frequency component, or d.c. It therefore pulsates above $2I_b$ (the quiescent d.c. components of both tubes) at a fundamental rate equal to twice the frequency of e_g. If $2I_b$ be subtracted from it and the difference plotted against time and analyzed, its d.c. component can be ascertained. This is then to be added to $2I_b$, giving the total $2I_b'$, which when multiplied by E_{bb} gives the d.c. energy input at full signal. The a.c. power output P_o can be determined from R_L and the peak value of i_1 and corresponding minimum value of i_2. Thus

$$P_o = \left[\frac{i_L(\text{peak})}{\sqrt{2}}\right]^2 R_L = \left(\frac{i_1 - i_2}{2\sqrt{2}}\right)^2 R_L$$
$$= \frac{(i_1 - i_2)^2}{8} R_L \tag{10}$$

(For other than class A operation, i_2 is zero, and it may just reach zero at peak swing even for class A operation.) The plate dissipation (per tube) at full signal is then evidently

$$W_{pd} = I_b' E_{bb} - \frac{P_o}{2} \tag{11}$$

and this, too, must not exceed the manufacturer's rating.

The actual procedure is probably best illustrated by an example, given in the following section.

7. Application to 6F6 Tube.—The construction described above will be applied to a pair of 6F6 tubes in push-pull. The operating conditions are as follows:

$$E_{bb} = 350 \text{ volts}$$
$$E_c = -38 \text{ volts}$$
$$I_b = 22.5 \text{ ma.}$$
$$R_L = 6{,}000 \text{ ohms (plate-to-plate)}$$
$$e_g = 63.8 \text{ volts (peak per tube)}$$

For this grid swing the grids are each driven 25.8 volts positive, but a zero-impedance driver (preceding) stage will be assumed.

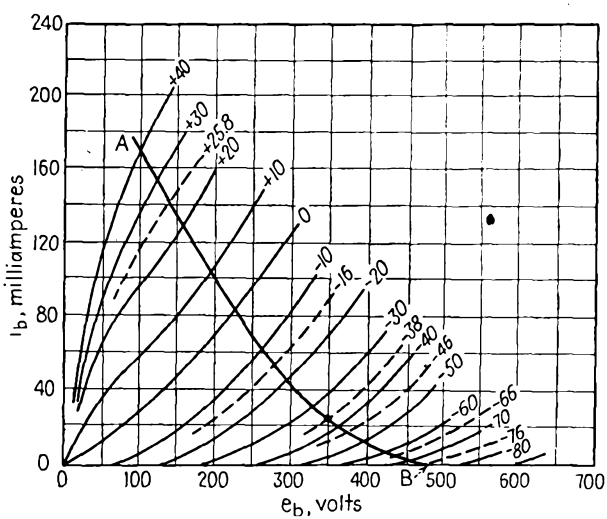


FIG. 79.—Push-pull load line for 6F6 tube—sliding-rule method.

The characteristic curves are shown in Fig. 79, and AB is the load line for either tube, obtained by sliding the rule at an angle corresponding to 3,000 ohms and bisected by the 350-volt ordinate. From this curve, simultaneous values of i_1 and i_2 can be obtained for the corresponding grid swings. Thus, if e_g is 8 volts per grid, i_1 is 30 ma. and i_2 is 15 ma. Then $(i_1 + i_2)/2 = 23$ ma., which is 0.5 ma. greater than the quiescent value 22.5. This increase of 0.5 ma. will be called i_{add}. The values for grid swings up to 63.8 volts have been set down in

Table I, page 155. The quantity i_{add} represents half the amount by which the power-supply current pulsates above its quiescent value.

This is then plotted against e_g and curve CD (Fig. 80) obtained. The purpose of this graph is to allow i_{add} to be interpolated for values of e_g not shown in Fig. 79 but corresponding to uniform increments of θ.

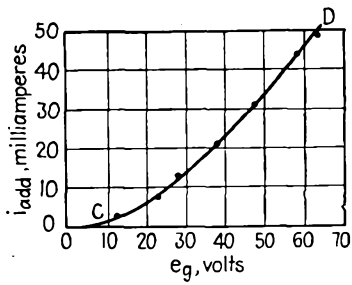


FIG. 80.—Interpolation curve—additional direct current vs. grid-signal swing.

If we now assume e_g is sinusoidal and of peak amplitude 63.8 volts, its equation is

$$e_g = 63.8 \sin \omega t$$

We can then calculate the instantaneous values of e_g for every 10 deg. of the cycle, or any other interval θ, and then find the corresponding values of i_{add} from Fig. 80. These are tabulated in Table II, page 155. From this table, i_{add} is plotted against θ and curve EF (Fig. 81) obtained. The area underneath this curve is then found in any convenient way, such as by adding all the small squares of the graph paper below it and dividing by the base. The quotient, to the proper scale, represents the d.c. component of i_{add}; in this case, it comes out to be 26.4 ma.

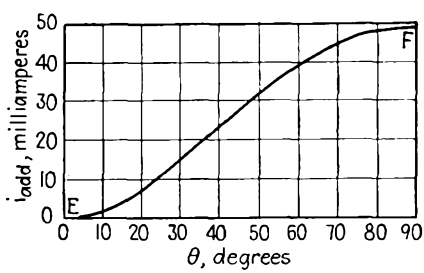


FIG. 81.—Wave shape for additional direct current.

This is added to the quiescent value of 22.5 ma., giving a total of 48.9 ma. $= I_b$. The d.c. power input is therefore

$$I_b' E_{bb} = 2 \times 0.0489 \times 350 = 34.2 \text{ watts}$$

The a.c. power output P_o is

$$\frac{(0.143)^2}{8} \times 6{,}000 = 15.3 \text{ watts}$$

The plate dissipation per tube is

$$W_{pd} = \frac{34.2 - 15.3}{2} = 9.95 \text{ watts}$$

The plate efficiency is

$$\frac{P_o}{I_b' E_{bb}} = \frac{15.3}{34.2} = 44.8 \text{ per cent}$$

Finally, in Fig. 82, is shown the output characteristic, *viz.*, i_L vs. e_g. Its linearity indicates that the stage output is practically directly proportional to the input, even though the individual tube currents are not. Hence, while the output of either tube is badly distorted, the output of the stage into the load resistance is not.

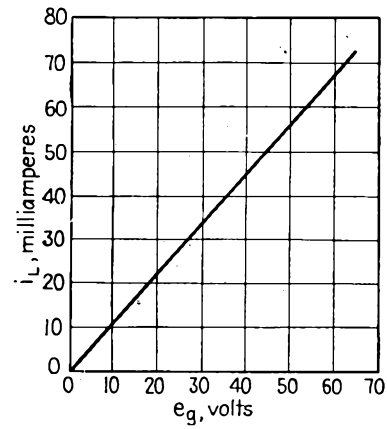


FIG. 82.—Output characteristic for 6F6 balanced amplifier.

8. Approximate Method for Determining the D.C. Component. The procedure described previously for obtaining the additional d.c. component is rather laborious, and a simplified method is greatly to be desired. A fairly good estimate of the total d.c. drawn under full signal drive can be made very simply from the following considerations:

If two linear tubes were operated class *B*, the bias of either would be at cutoff, and during no-signal periods their d.c. drain would be zero. Under full-signal drive, each would draw a half sine-wave current, and in the mid-branch the current would be a series of half sine waves similar to that furnished by a full-wave rectifier. If i_1 is the peak current of either, then the average current, or d.c. component, in the mid-branch would be $(2/\pi)i_1$, or 63.7 per cent of i_1.

In ordinary practice, however, the tubes employed are not linear, and only an approach to class *B* operation can be made, *i.e.*, extreme class *AB*. The curvature of their characteristics causes the tubes to draw current even during no-signal periods

TABLE I

e_g	i_1	i_2	$\frac{i_1 + i_2}{2}$	i_{add}
8.0	31	15	23.0	0.5
22.0	53	7	30.0	7.5
28.0	63	5	34.0	11.5
38.0	85	2	43.5	21.0
48.0	107	0	53.5	31.0
58.0	131	0	65.5	43.0
63.8	143	0	71.5	49.0

TABLE II

θ	sin θ	e_g	i_{add}
10	0.1736	11.07	1.5
20	0.3420	21.80	7.3
30	0.5000	31.90	15.2
40	0.6430	41.00	24.0
50	0.7660	48.90	32.0
60	0.8660	55.20	39.5
70	0.9400	59.90	45.3
80	0.9850	62.80	48.3
90	1.0000	63.80	49.0

and, when signal is applied, to draw a mid-branch current that looks more like a direct current with a double-frequency sine-wave current superimposed than a series of half sine waves. This is borne out, for example, by the current wave shape shown in Fig. 81. For such a wave shape, the d.c. component per tube is the original (no-signal) d.c. component per tube plus one-half of half the difference between the crest and trough values of the mid-branch current. This is the same as saying it is the average between the crest and trough values, *i.e.*, half their sum.

The crest value per tube is evidently half the sum of the two tube currents at peak grid swing, or $(i_1 + i_2)/2$. If i_2 is zero for this swing, then the crest value is simply $i_1/2$. The trough value is evidently the original d.c. component per tube. Hence we may write the full-signal current as being

$$I_{dc}' = \frac{1}{2}\left(\frac{i_1 + i_2}{2} + I_{dc}\right) = \frac{1}{2}\left(\frac{i_1}{2} + I_{dc}\right) \qquad (12)$$

where the latter expression is for $i_2 = 0$ and in either case I_{dc} is the original direct current per tube. If Eq. (12) were to be applied to the case of two linear tubes in class B operation (in which case $I_{dc} = 0$), we should have $I_{dc}' = 50$ per cent of $i_1/2$ instead of 63.7 per cent of $i_1/2$, or 13.7 per cent too low.

For class AB operation of actual tubes, a value of I_{dc}' intermediate in value between that given by Eq. (12) and that given by class B operation for two linear tubes will be found to flow. A fairly good approximation for a wide range of operation is to increase the value of I_{dc}' given by Eq. (12) by about 5 per cent. We thus have, finally,

$$I_{dc}' = \frac{1.05}{2}\left(\frac{i_1}{2} + I_{dc}\right) \tag{13}$$

If we apply Eq. (13) to the previous example of the 6F6 tubes, we obtain

$$I_{dc}' = \frac{1.05}{2}\left(\frac{142}{2} + 22.5\right) = 49.1 \text{ ma.}$$

By the more exact method of the preceding section, I_{dc}' was found to be 48.9 ma., which is very close to the above value.

As another example, take the case of an 807 tube, whose characteristics are shown in Fig. 85 and whose operating voltages are $E_{bb} = 400$ volts, $E_{sg} = 300$ volts, and $E_c = -25$ volts. From the figure, it is found that I_b is 50 ma. per tube. For a grid swing of $+15$ volts positive and a plate-to-plate resistance of 3,800 ohms, it will be found that $i_1 = 350$ ma. and $i_2 = 0$. Then, by Eq. (13),

$$I_{dc}' = \frac{1.05}{2}\left(\frac{350}{2} + 50\right) = 118.1 \text{ ma.}$$

By the more exact method, I_b' comes out to be 120 ma., which is again in good agreement with the former value.

Equation (13) in conjunction with the principles of Sec. 10 will be found to be particularly useful for quick calculations of the peak grid swing, optimum load resistance, power output, input power, plate dissipation, and plate efficiency. The more exact method need be used only where a more careful check of such matters is required or possibly for nearly pure class B operation.

9. Further Remarks on Self-rectification.—In Chap. I (Sec. 25), the need for correcting the load line in the case of choke feed was discussed. As was shown, the operating point moves off the load line to a position corresponding to a lower value of bias (if I_b increases when grid signal is impressed) in such manner that the locus of the d.c. component of the plate current is along the load line of the d.c. resistance of the choke, whereas the locus of the a.c. component is along the load line of R_L.

It would appear reasonable that a similar correction should be made in the case of a balanced amplifier. However, such is not

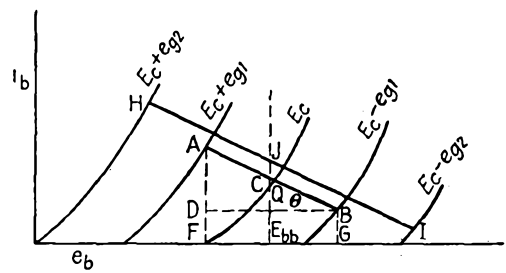


FIG. 83.—Sliding-rule method showing paths of operation for the output and mid-branch currents.

the case. A rigorous proof is not available, but the reasonableness of the above statement will be shown from two viewpoints.

1. In Fig. 83 is shown the position AB of the sliding rule for an instantaneous grid swing $2e_g$. As demonstrated in Sec. 3,

$$\cot \theta = \frac{R_L}{2}$$

and $AF = i_1$, $BG = i_2$, $AC = CB$, Q = quiescent point, and $FE_{bb} = E_{bb}G$. It is also evident that $AD = i_1 - i_2 = 2i_L$.

It is further apparent from the geometry of the figure that CE_{bb}, the bisector of the two sides of the trapezoid $FABG$, is the average between AF and BG, *i.e.*, that

$$CE_{bb} = \frac{AF + BG}{2} = \frac{i_1 + i_2}{2} \tag{14}$$

CE_{bb} is therefore equal to half the mid-branch current flowing through the B supply, CQ is equal to half the rise in this current from its quiescent value, $2QE_{bb}(= 2I_b)$, and hence $CQ = i_{add}$ (as mentioned in Sec. 7). For a larger grid swing $2e_{g2}$, i_{add} is

evidently equal to JQ. It will be noted that the locus of i_{add} is a vertical line through E_{bb}, or the load line for i_{add} corresponds to a zero impedance power supply and output choke, as assumed in the derivation of this graphical construction.

Since i_{add} contains all the even harmonics including the additional d.c. component due to self-rectification, it is evident that the graphical construction as given satisfies the requirements for the locus of $(i_1 + i_2)/2 = i_{add}$ as well as for $(i_1 - i_2)/2 = i_L$, and hence no correction for the load line is necessary as in the case of a single-ended amplifier. To summarize, the locus for i_1 and i_2 is $HAQBI$ and for i_{add} is $E_{bb}QCJ$, a vertical straight line, and both loci satisfy the respective requirements that $i_1 - i_2$ be determined by a plate-to-plate resistance of value R_L and that $i_1 + i_2$ be determined by a zero mid-branch impedance—that of the power supply and windings of the two halves of the output choke.

2. From a physical viewpoint it will be apparent that no correction to the load line is necessary. In the case of a single-ended amplifier, the quiescent d.c. value of the plate current, I_b, as well as the d.c. component at full signal, I_b', flows through the output choke paralleling R_L. In an actual circuit this choke has some finite inductance L, instead of infinite inductance, as assumed in the ideal case of Sec. 27, Chap. I. The magnetic energy stored in the choke is equal to $\frac{1}{2}LI_b^2$ at no signal and $\frac{1}{2}LI_b'^2$ at full signal. The difference in energy stored accounts for the correction of the load line (the shift of the operating point away from the quiescent point). If, as is usually the case, I_b' exceeds I_b, the load line is shifted upward; if I_b' is less than I_b, as may possibly be the case for a pentode, then the load line will have to be shifted downward.

In the case of a balanced amplifier, the d.c. components of the two tubes always balance one another in the two windings of the output choke, and thus no d.c. magnetic flux is established and no energy is stored. There is thus lacking the mechanism by which the load line may be shifted, and hence no correction is necessary.

10. Optimum Value of Load Resistance.—The fact that no correction is necessary is fortunate in that it enables in most cases an estimate or determination of the optimum value of R_L and e_g to be made by a simple graphical procedure.

Whether the tube is operated class A, AB, or B, cutoff of either tube current occurs in practical cases at least at the peaks of the grid swing, so that at those points, if not earlier in the cycle, $2i_L$ becomes identical with the individual tube currents. Specifically, in class AB operation, $2i_L$ becomes identical with the current of the tube whose grid is being driven positive over an appreciable portion of that half cycle. During this portion, the other tube is inoperative, and hence the first tube feeds R_L directly and thus sees its reflected value of $R_L/4$.

In Fig. 84 is shown the load line for R_L as it appears to either tube. Cutoff occurs for one tube at C, for a grid swing equal to $-e_{g1}$; for the other, undergoing a grid swing of $+e_{g1}$, the current

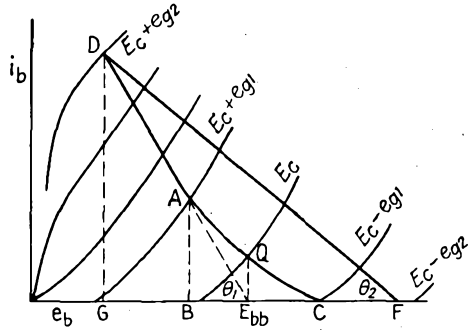


FIG. 84.—Method of determining optimum value of the load resistance.

has reached the value of AB. Its load line (for increased grid swing) continues along the straight line AD, corresponding to a value of $R_L/4$. Thus, for a grid swing e_{g2}, point D is reached and the current is DG, while, for the other tube, point F is reached and its current is zero.

From previous considerations it is clear that DF corresponds to $R_L/2$; *i.e.*,

$$\cot \theta_2 = \frac{FG}{GD} = \frac{R_L}{2} \tag{15}$$

But

$$GE_{bb} = E_{bb}F$$

$$\frac{GE_{bb}}{DG} = \cot \theta_1 = \frac{R_L}{4} \tag{16}$$

But DA has the slope corresponding to $R_L/4$. Hence DA when prolonged passes through E_{bb}. This important fact enables an optimum value of R_L to be found for maximum power output.

In linear-circuit theory, the maximum-power theorem states that maximum power output is obtained when the load resistance equals the generator internal resistance. In the case of a triode push-pull stage, the R_p of each tube varies from infinity (beyond cutoff) to a certain finite value during the grid-voltage cycle, and hence the above theorem cannot be applied directly to this circuit. However, it will be evident that most of the power output during one half cycle comes from the tube whose grid swing is in a positive direction and occurs mainly during the peak of the half cycle. The same is true for the other tube during the next half cycle. Hence, if R_L be matched to the value of the R_p of the tubes at the peak positive grid swing, maximum power output may be expected.

In this region of the i_b-e_b characteristic, the value of R_p is fairly constant, for it will be observed in general that the i_b-e_b curves are fairly straight and parallel to one another. However, it will also be noted that for positive grid voltages the curves exhibit a knee just as for pentode tubes, and for the same reason—a positive electrode (here the control grid) between the plate and cathode. It is evident that the load line per tube must avoid this knee. The higher the grid voltage, the farther to the right does the knee occur, since space current can be diverted to the control grid even at a higher plate voltage.

With these facts in mind, we can proceed with the determination of the optimum value of R_L. A line is drawn through the point E_{bb} on the e_b axis, at a slope corresponding to the average value of R_p in the upper left-hand region of the tube characteristics, such as $E_{bb}AD$ (Fig. 84). The optimum value of R_L is simply

$$R_L = 4R_p \tag{17}$$

Where this line strikes the knee of an i_b-e_b curve indicates the maximum grid swing possible without excessive distortion. It is possible, however, to lower R_L (tilt the above line more to the vertical) and thus obtain a greater grid swing before the knee of an i_b-e_b curve is reached and hence even more power output.

However, another factor must be taken into account, and that is the plate dissipation at full signal. The steeper the load line, the greater the self-rectification and hence the d.c. power input at full signal. Consequently, while a lower value of R_L may permit

a greater grid swing and hence more power output, the plate dissipation may rise to an excessive value.

Indeed, even if R_L is chosen according to Eq. (17), the plate dissipation may be excessive. In such an event, either a higher R_L or a smaller grid swing or both must be employed. Equation (17) is therefore a guide for, rather than an absolute determination of, R_L; but any tube whose design is properly coordinated will require a value of R_L not far from the value given by Eq. (17). For example, a pair of 2A3 tubes in class AB_1 operation require a plate-to-plate load resistance of 3,000 ohms. It will be found that this value is approximately $4R_p$ for either tube, where the R_p is measured at, for example, $e_c = 0$ and $e_b = 100$ volts, whereas R_p at the quiescent point is 800 ohms.

The above discussion is concerned with the triode tube. In the case of the pentode, it will obviously be impossible to choose R_L according to Eq. (17). In this case, a line is drawn through E_{bb} at a sufficient slope to avoid the knee of the characteristics. The slope of this line corresponds to $R_L/4$ and thus indicates the maximum value of R_L that can be used. Just as in the case of a single-ended pentode, R_L must be much smaller than R_p to avoid excessive distortion products, so that the mismatch in impedance is very great. Hence, the maximum permissible value of R_L is the value that gives the maximum power output, since it gives the least mismatch.

11. Typical Example.—As an example of the above, Fig. 85 gives the plate family for an 807 beam power tetrode. For class AB_1 operation, the maximum slope is given by AB and corresponds to a resistance of 1,650 ohms. The value of R_L is therefore 6,600 ohms. The distortion will be found to be 2 per cent, mainly third order, at a power output of 34 watts.

If class AB_2 operation is desired, it will be found—after a series of trials—that a positive grid voltage of +15 volts can be attained. Several factors enter into this determination and will be discussed presently. If this value be chosen, then the line CA must be used instead of BA, since the latter would obviously cut the knee of the characteristics for a grid swing into the positive region. For line CA, R_L comes out to be 3,800 ohms. The power output is approximately 60 watts, with a distortion content that depends upon the design of the preceding driver stage. The dissipation at full signal is approximately 24 watts (plate

and screen grid), which is within the limits permitted by the manufacturer.

If a higher positive grid swing is desired, a lower value of R_L will be required. While the power output will be increased, the dissipation will be increased to a value higher than permitted, so that the above-15-volt positive grid swing is the maximum permissible. This fact can be discovered only after trying different grid swings and corresponding values of R_L and analyzing the results in the manner described in Sec. 6 or 8.

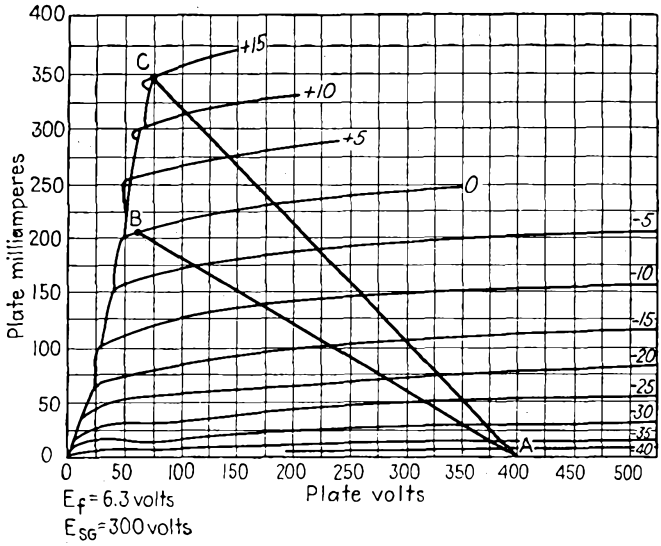


FIG. 85.—Push-pull characteristics for a 6L6 tube.

Other factors determining the peak grid swing are the driver power required for the grids (to be discussed later), the permissible grid dissipation, and, finally, instantaneous emission of the cathode (particularly in the case of oxide-coated cathodes). Unfortunately, little information concerning these is normally furnished by the manufacturer.

12. Correction for Mid-branch Impedance.—In actual balanced-amplifier circuits, the B and C sources often have appreciable impedance. These are called *mid-branch impedances* and usually the B supply impedance is the more important. We shall analyze the effects of two types of this impedance here:

1. The case where the impedance is a pure resistance R_b to all frequencies. The resistance is usually that of the various windings in the power-supply circuit, plus that of the rectifier tube.

2. The case where the impedance is a resistance R_b to direct current and zero to alternating current. This is approximately the case when the filter section is terminated in a condenser of such size that its reactance to twice the lowest frequency to be amplified is negligible.

We shall take Case (1) first. We note that the resistance to $(i_1 - i_2)/2$ is R_L as before but that the resistance to $2I_b$ (for

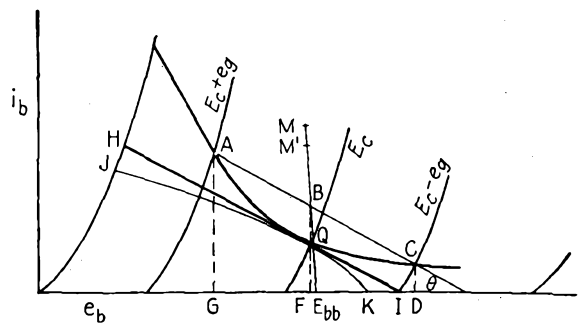


FIG. 86.—Graphical construction for the case of a mid-branch resistance.

$e_g = 0$), and to $i_1 + i_2$ (when $e_g \neq 0$) is R_b. This is equivalent to a current $(i_1 + i_2)/2$ flowing through a resistance $2R_b$. To depict these items graphically, we proceed as follows:

From the point E_{bb} (Fig. 86) (corresponding to the open-circuit or generated voltage of the B power supply) a line $E_{bb}B$ is drawn at an angle ψ such that

$$\psi = \cot^{-1} 2R_b \tag{18}$$

Its intersection with the curve labeled E_c gives the no-signal d.c. component of each tube. For equal grid excursions $E_c + e_g$ and $E_c - e_g$ about E_c, the simultaneous currents of the two tubes are given by AG and CD, where θ is determined by Eq. (9) and B is the mid-point of AC. The position of AC can be found by sliding a rule, at the angle θ to the e_b axis, up or down along the curves.

Since BF is always along the load line BE_{bb} of $2R_b$ and $i_1 - i_2$ along the load line of $R_L/2$, we have satisfied the conditions stated above. Points A and C are thus two points of the load line for the tube. Other points can be found in the same manner by

sliding the rule along other pairs of tube curves so that the midpoint of the segment of the rule is always on BE_{bb}.

Some interesting results can be obtained for various values of R_b. In general, the load line is flatter and not so steep as for R_b equal to zero. If R_b equals $R_L/4$, the load line becomes a straight line whose slope is that for $R_L/2$ (line HI, Fig. 86). For $R_b > R_L/4$, the load line is concave downward (line JK). In all cases, the power output is evidently decreased since $(i_1 - i_2)/2$ is smaller.

Case (2) is more difficult, although more common in practice, but it can be solved by a series of approximations. The difficulty lies in the fact that the d.c. component of i_{add} must lie along $2R_b$, whereas the a.c. components (the other even harmonics) lie on a vertical line, since the impedance to their flow is now zero.

As before, I_b is determined by $2R_b$, that is, line $E_{bb}B$ of Fig. 86. The sliding rule must now be bisected by the ordinate through Q instead of BE_{bb}. When the load line is drawn, the current i_{add} can be found, as in Sec. 6 or 8, and analyzed for its d.c. component. This is added to I_b, and where this value intersects BE_{bb}, say, at point M, is the new revised value of direct plate voltage of the tubes. That is, the increase in the d.c. component due to self-rectification causes a larger drop in R_b than $2I_b$ (the value of both tubes at no signal) and hence changes the operating point.

The sliding rule must now be bisected by the ordinate through M. A new load line is obtained and the new i_{add} analyzed for its d.c. component. This will be less, for the revised load line will lie below the first, although cutoff occurs earlier in the cycle. The revised d.c. component of i_{add}, when added to I_b, willl locate a lower point M' and hence a new ordinate between the second and the first (the one through Q). The process is continued until no further revision is necessary. The end result will be that the power output will be less than for the case where R_b is zero.

13. Correction for Winding Resistance.—In actual practice, a two-winding transformer instead of a tapped choke is used. The two halves of the transformer may each have appreciable resistance; call it R_{pw}. The secondary winding resistance R_{sw} may be combined with the actual load resistance r_L and reflected by the square of the turns ratio n to give the plate-to-plate resistance R_L. Thus,

$$R_L = \frac{r_L + R_{sw}}{n^2} \tag{19}$$

The primary resistances, however, cannot be disposed of so readily. The circuit of Fig. 72 may be modified to take care of this condition by the well-known transformer theory, as shown in Fig. 87. As can be seen, the primary winding resistances can be associated with the two tubes instead of with the transformer so that we may regard the tubes as generators having higher internal resistances. The situation, however, is not quite the

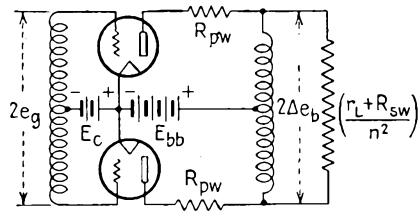


FIG. 87.—Balanced-amplifier circuit having output transformer with winding resistance.

same as that discussed in Sec. 12, for there R_b acted as a common, or coupling, resistance between the two tubes, whereas here each R_{pw} carries only its own tube current.

This case can be handled by first replotting all the tube curves so as to include R_{pw} as part of the equivalent tube's resistance. This is a laborious process, and the resulting family of curves would hold only for the particular value of R_{pw} used. Another

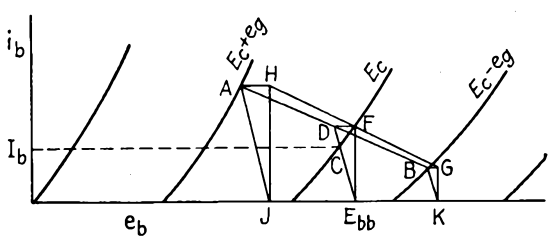


FIG. 88.—Proof of construction for case of winding resistances.

output transformer with a different primary winding resistance would require a new set of curves. Fortunately, the construction can be simplified so as to apply to the original tube family of curves.

In Fig. 88 is shown the method of construction which is also that of the sliding rule. Line DE_{bb} is the load line for R_{pw}. AB is the segment of the rule between curves of equal grid swings e_g.

The slope of AB is such that

$$\theta = \cot^{-1}\left(R_{pw} + \frac{R_L}{2}\right) \tag{20}$$

where R_L is given by Eq. (19). Furthermore, D is the mid-point of AB, and C is the normal d.c. component. Points A, C, and B are on the load line, and further points can be determined in exactly the way that A and B were. From this, further facts and constructions clearly follow.

The proof is as follows: Suppose the correct values of i_1 and i_2 were known beforehand for the grid swing e_g and are represented by H and G, respectively. Owing to R_{pw}, the instantaneous plate voltages are determined where the load lines for R_{pw} through A and B strike the e_b axis in J and K, respectively. If i_1 and i_2 are correct, then

$$JE_{bb} = E_{bb}K$$

and JK is the voltage across R_L.

Points H and G are the projections of A and B, over to the ordinates through J and K, respectively. The line joining H and G will make the proper slope corresponding to $\cot^{-1}(R_L/2)$. The mid-point of HG is represented by F and its vertical projection on CE_{bb} as D. We note that $FE_{bb} = (HJ + GK)/2$ and that triangles AHJ, DFE_{bb}, and BGK are similar, so that

$$DE_{bb} = \frac{AJ + BK}{2}.$$

This in turn means that figure $AJKB$ is a trapezoid, and therefore ADB is a straight line whose mid-point is D.

We finally note that the angle of slope of AB is given by

$$\theta = \cot^{-1}\frac{AH + JK - BG}{HJ - GK} = \cot^{-1}\frac{(HJ)R_{pw} + JK - (GK)R_{pw}}{HJ - GK}$$

$$= R_{pw} + \frac{JK}{HJ - GK} = R_{pw} + \frac{R_L}{2} \tag{21}$$

This establishes the correctness of the construction.

This construction can be combined with that given in Sec. 12. For the case of R_b, a resistance to all frequencies, DE_{bb} is drawn to represent the load line for $2R_b + R_{pw}$, and AB so that Eq. (20) is still valid. For the case of R_b, a resistance to direct current only,

we draw DE_{bb} to represent $2R_b + R_{pw}$, and where it intersects the E_c curve at C is the normal value of d.c. component. Through C a line representing R_{pw} is drawn, and the rule is slid so that AB is bisected by this line rather than by the first one representing $R_{pw} + 2R_b$.

14. Analytical Treatment.—The literature on balanced amplifiers is not so extensive as that on single-ended amplifiers; hence, an analytical treatment will now be given to supplement the graphical analysis. The most general case will be treated first—two tubes in push-pull with an actual three-winding (output) transformer, together with mid-branch impedances.

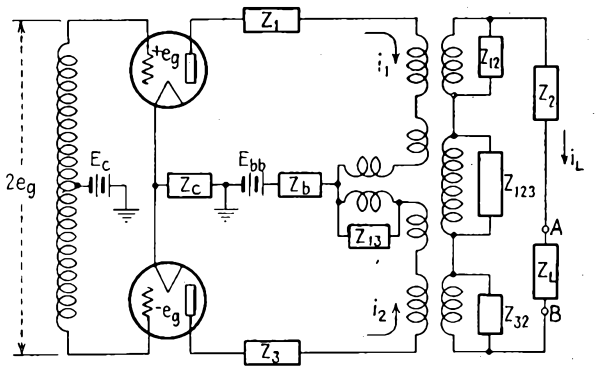


FIG. 89.—Generalized balanced-amplifier circuits.

Although there are possibly simpler equivalent circuits for the three-winding transformer, the one shown in Fig. 89 will suffice for the analysis. Here the three-winding transformer is represented by three ideal two-winding transformers, one ideal three-winding transformer, and suitable impedances. Thus, Z_{123} represents a mutual impedance common to the three windings and due to flux that links all three windings regardless of in which winding or windings the corresponding magnetomotive force or forces are located. Similarly, Z_{12} represents the mutual impedance exclusive to primary winding 1 and the secondary winding 2; Z_{32}, the mutual between primary winding 3 and secondary 2; Z_{13}, the mutual between primary windings 1 and 3. Furthermore, Z_1 represents an impedance to be found only in the mesh of primary 1. It can consist of an inductance due to flux that links only winding 1, winding resistance, and any other impedance deliberately inserted in series with the top tube. Similar interpretations

are to be given to Z_2 and Z_3, while Z_L represents the load impedance and Z_b and Z_c are mid-branch impedances, of which Z_c is also in the mid-branch grid circuit. Normally, windings 1 and 3 have an equal number of turns, while winding 2 may have a different number (usually less). However, another two-winding ideal-transformer of the corresponding turns ratio can be inserted at terminals A and B and the secondary connected to the actual value of load impedance. Hence, no loss in generality is incurred in assuming equality of turns of all three windings and a value of secondary load impedance that is the reflected value of the actual load impedance employed.

For perfectly balanced operation it is necessary that the tubes be identical, that impedance Z_1 equal Z_3 and Z_{12} equal Z_{32}. It will now be shown that the output load current contains no even-order modulation products* of the input signal voltage $2e_g$. First, suppose that $2e_g$ is symmetrical about the time axis, *i.e.*, has no even harmonics. From the symmetry of the circuit it is evident without further proof that the load current i_L will be symmetrical about the time axis and will therefore contain only odd harmonics too. Hence, for a symmetrical input voltage, no even-order modulation products appear in the output.

Consider next an input voltage that is asymmetrical with respect to the time axis. Specifically, assume

$$2e_g = E_1 \sin \omega_1 t + E_2 \sin \omega_2 t \tag{22}$$

* A modulation product is a sinusoid appearing in the output of a non-linear circuit, which is of a frequency different from those of the voltages impressed upon the circuit. If a single-frequency voltage is impressed, then the modulation product is a harmonic of the impressed voltage. If the impressed voltage consists of several sinusoids, then the modulation products are of two kinds, harmonics, and summation and difference beat frequencies. The latter are often called *cross-modulation products*, as they represent the interaction of the impressed voltages and their harmonics upon one another within the circuit. Let m represent the mth harmonic of one of two impressed voltages and n that of the other. The cross modulation produces summation and difference beat frequencies of value $m + n$ and $m - n$ respectively. If m and n are both odd or both even, then $m + n$ and $m - n$ are both even. These are then called *even-order modulation products.* If one is odd and the other even, $m + n$ and $m - n$ are called *odd-order modulation products.* This concept can be extended to three or more impressed voltages, and the harmonics themselves can be regarded as a special kind of cross modulation —that of the impressed voltage with itself, or one of the other harmonics that it evokes.

The output load current may be expressed as

$$i_L = I_1 \sin(\omega_1 t + \phi_1) + I_2 \sin(\omega_2 t + \phi_2) + I_3 \sin(2\omega_1 t + \phi_3) + I_4 \sin(2\omega_2 t + \phi_4) + I_5 \sin[(\omega_1 + \omega_2)t + \phi_5] + I_6 \sin[(\omega_1 - \omega_2)t + \phi_6] + \cdots \quad (23)$$

In Eq. (23) only the first- and second-order modulation products have been indicated. The form of the third and higher order is apparent. It will now be shown that the amplitudes of the even-order terms are zero.

Suppose that the connections of the input source were reversed. Then Eqs. (22) and (23) would become, respectively,

$$\begin{aligned} 2e_g' &= E_1 \sin(\omega_1 t + \pi) + E_2 \sin(\omega_2 t + \pi) \\ &= -E_1 \sin \omega_1 t - E_2 \sin \omega_2 t = -2e_g \end{aligned} \quad (24)$$

and

$$\begin{aligned} i_L' &= I_1' \sin(\omega_1 t + \pi + \phi_2') + I_2' \sin(\omega_2 t + \pi + \phi_2') \\ &\quad + I_3' \sin(2\omega_1 t + 2\pi + \phi_3') + I_4' \sin(2\omega_2 t + 2\pi + \phi_4') \\ &\quad + I_5 \sin[(\omega_1 + \omega_2)t + 2\pi + \phi_5'] \\ &\quad + I_6' \sin[(\omega_1 - \omega_2)t + 2\pi + \phi_6'] + \cdots \\ &= -I_1' \sin(\omega_1 t + \phi_1') - I_2' \sin(\omega_2 t + \phi_2') \\ &\quad + I_3' \sin(2\omega_1 t + \phi_3') + I_4' \sin(2\omega_2 t + \phi_4') \\ &\quad + I_5' \sin[(\omega_1 + \omega_2)t + \phi_5'] \\ &\quad + I_6' \sin[(\omega_1 - \omega_2)t + \phi_6'] + \cdots \end{aligned} \quad (25)$$

From the symmetry of the circuit it can be seen that

$$i_L' = -i_L$$

since the currents must have the same form but must flow in opposite directions through Z_L. This can be the case only if the even-order modulation products (which do not change sign upon reversal of the input voltage) are zero; *i.e.*, I_3, I_3', I_4, I_4', I_5, I_5', I_6, I_6', etc. are zero, and in addition the other I's in Eqs. (23) and (25) are equal, and also the ϕ's. The above results can be expressed more compactly as

$$i_L = \sum_0^n \sum_0^m I_{nm} \sin[(\omega_n \pm \omega_m)t + \phi_{mn}] \quad (26)$$

$$i_L' = -\sum_0^n \sum_0^m I_{nm}' \sin[(\omega_n \pm \omega_m)t + \phi_{mn}'] \quad (27)$$

where $I_{nm} = I_{nm}'$, $\phi_{mn} = \phi_{mn}'$, and $I_{nm} = I_{nm}' = 0$ if $n + m$ is even. Hence the output load current $i_L(t)$ and the output load voltage $e_L(t) = i_L(t)Z_L(\omega)$ contain only odd-order modulation products of $2e_g(t)$, where the t indicates that these quantities are functions of time and the ω indicates that Z_L is a function of frequency.

15. Mid-branch Current.—The mid-branch current is evidently $i_1 + i_2$ from Fig. 89 and is evidently unchanged in sign when $2e_g$ is reversed. Therefore, by a process of reasoning similar to the above, it cannot contain any odd-order modulation products and must contain only even-order terms. Hence the statement that the odd harmonics flow in series, and the even harmonics in parallel, through the balanced-amplifier circuit. The odd harmonics encounter the impedances, such as Z_L, in series around the circuit, whereas the even harmonics encounter only the mid-branch impedances. (Impedances such as Z_1 and Z_3 are effectively in parallel for these, whereas Z_{13} is effectively in series for the odd harmonics.) In the case of an ideal three-winding output transformer and zero branch impedances, the even harmonics encounter no impedances in the external circuit. Specifically, in flowing through the infinite impedances of the two primary windings in opposite directions, they encounter no net impedance because they cancel each other's induced voltages there, since the windings are assumed to have unity coupling.

16. Effect of Mid-branch Impedance.—The preceding analysis has revealed some interesting points.

1. Mid-branch impedances, such as Z_b and Z_c, can produce no even-order terms in the output of a perfectly balanced amplifier.

2. The voltages across them can, however, cross modulate with the signal voltage $2e_g$ to produce odd-order terms. For an interesting application to vacuum-tube voltmeters see Turner and MacNamara.[2]

3. If such odd-order terms have the right phase, they may reduce similar odd-order terms produced by the tubes themselves. This, however, depends upon the particular tubes; in general, mid-branch impedances should be avoided, as by the use of sufficient by-pass condenser capacity.

4. Contrary to statements sometimes made, mid-branch impedances do not produce even-order terms in a perfectly balanced amplifier.

5. The impedance of each half of the primary circuit must be the same; otherwise i_1 and i_2 will not interchange in value during the two halves of the cycle of $2e_g$, when the latter is symmetrical with respect to the time axis. In other words, unbalance here will produce even harmonics.

6. If the tubes are not identical, even-order modulation products will result.

17. Effect of Mid-branch Voltages.—We have seen that mid-branch impedances produce no even-order terms but that the voltages across them may cross modulate with e_g to produce odd-order terms. Since we have placed no restriction upon the form of Z_b and Z_c, they can be such as to have any kind of voltages developed across them by $i_1 + i_2$. Hence, what is true for Z_b and Z_c is true for any mid-branch voltages e_1 and e_2 inserted in the grid and plate mid-branches, respectively. e_1 and e_2 may be, for example, hum voltages.

If $2e_g$ is zero and only e_1 and e_2 are present, it is evident that the latter produce equal current flows in the two tubes, so that the voltages induced in the secondary cancel one another, and hence i_L is zero. That is, in a perfectly balanced amplifier, with no signal impressed, hum voltages produce no hum output. However, when a signal voltage is impressed, the hum voltages cross modulate with it to produce odd-order output current components in addition to that which is a copy of the signal voltage. It is therefore advisable to filter adequately the plate and grid-bias supplies, although the filtering in practice may be less than for a single-side amplifier because of the fact that the hum is not apparent during silent moments (when $2e_g = 0$) and the cross-modulation products are partly masked by the louder signal components in the output when $2e_g$ is not zero. For best results and least distortion, however, the filtering should be as complete as possible.

18. Ideal Push-pull Tubes.—It will now be of interest to find what the tube characteristics should be in order that the output current and voltage are faithful copies of $2e_g$. Since mid-branch impedances produce odd-order cross-modulation products in the output, it can be seen that these impedances should be zero, which in turn also implies that the output transformer should be ideal and the load impedance resistive—call it R_L.

It will be assumed that the tube characteristics can be expressed by a suitable power series. No restriction on the variability of the

μ of either tube (both identical) will be made; hence, a double power series will be required (for triode tubes). The expansion will be made from the origin—$e_b = e_c = i_b = 0$; hence, it will be of the Maclaurin form. Where operation beyond cutoff of either tube is desired, suitable modifications will be made. The application of the methods that follow to four- and five-element tubes should be self-evident to the reader and are omitted because the triple and quadruple power series required reveal nothing different from the double series. Indeed, for constant screen-grid and suppressor-grid voltages, the multielement tubes are equivalent to triodes. Also, as shown by Peterson and Evans,[3] a single power series in terms of the control-grid voltage as the independent variable may be written (the ordinary dynamic characteristic), but the more general double power series will be used here because it reveals the circuit parameters more definitely.

The series may be written as follows:

$$\begin{aligned} i_b &= A_{10}e_b + A_{01}e_c + A_{11}e_be_c + A_{20}e_b^2 \\ &\qquad + A_{02}e_c^2 + A_{21}e_b^2e_c + A_{12}e_be_c^2 + \cdots \\ &= \sum_0^m \sum_0^n A_{hk}e_c^h e_b^k \end{aligned} \tag{28}$$

From the calculus it is known that

$$A_{hk} = \frac{1}{h!k!}\frac{\partial^{h+k} i_b}{\partial e_c^h \partial e_b^k} \tag{29}$$

This is the law by which a Maclaurin expansion in two variables is accomplished. The extension to three or more variables should be self-evident to the reader.

The series is in terms of the tube electrode voltages; *i.e.*, if these are known, the tube current i_b can be found from Eq. (28). However, as is usually the case in tube problems, it is not these voltages which are known, but those applied to the tube electrodes through load impedances, in which occur voltage drops due to the as yet unknown current i_b. This is particularly the case for the plate circuit but can also be true for the grid circuit if grid current flows or if there is feedback from the plate into the grid circuit. In the balanced-amplifier circuit a further complication arises: the plate voltage of either tube depends, in part, upon the voltage induced from the other tube's action.

Fortunately, however, these effects all occur in the external load circuit, which is linear, so that they may be superimposed one upon the other. As mentioned at the beginning of this section, in the following analysis an ideal three-winding transformer is assumed, or its equivalent, an ideal two-winding autotransformer with unity coupling between its two windings, across which R_L is connected. (This was also assumed in the discussion at the beginning of this chapter.) To either tube, R_L appears as $R_L/4$. Thus, to tube I the voltage drop in the winding to which it is connected is $i_1R_L/4$. However, R_L appears as $R_L/4$ to tube II, and this tube experiences a voltage drop in its winding of $i_2R_L/4$. The latter voltage drop induces an opposite voltage in the first winding, and similarly for the voltage drop $i_1R_L/4$ in the second winding. As a consequence, the total voltage drop to Tube I is $(i_1 - i_2)(R_L/4)$; to Tube II,

$$(i_2 - i_1)(R_L/4) = -(i_1 - i_2)(R_L/4).$$

We may therefore write for tubes I and II, respectively,

$$\begin{aligned} e_{b_1} &= E_{bb} - (i_1 - i_2)\frac{R_L}{4} \\ e_{b_2} &= E_{bb} + (i_1 - i_2)\frac{R_L}{4} \end{aligned} \tag{30}$$

The corresponding grid voltages are

$$\begin{aligned} e_{c_1} &= E_c + e_g \\ e_{c_2} &= E_c - e_g \end{aligned} \tag{31}$$

Substituting these values from Eqs. (30) and (31) in Eq. (28), we obtain

$$\begin{aligned} i_1 &= \sum_0^m \sum_0^n A_{hk}(E_c + e_g)^h \left[E_{bb} - (i_1 - i_2)\frac{R_L}{4}\right]^k \\ i_2 &= \sum_0^m \sum_0^n A_{hk}(E_c - e_g)^h \left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right]^k \end{aligned} \tag{32}$$

The load current is

$$i_L = \frac{i_1 - i_2}{2} \tag{33}$$

For distortionless amplification,

$$i_L = B_1 e_g = \frac{i_1 - i_2}{2} \tag{34}$$

where B_1 is a constant. But from Eq. (32) we have

$$\frac{i_1 - i_2}{2} = \sum_0^m \sum_0^n \frac{A_{hk}}{2} \left\{ (E_c - e_g)^h \left[E_{bb} - (i_1 - i_2)\frac{R_L}{4} \right]^k - (E_c + e_g)^h \left[E_{bb} + (i_1 - i_2)\frac{R_L}{4} \right]^k \right\} \tag{35}$$

The individual factors can be expanded by the binomial theorem, so that each term of the function represented by Eq. (35) may in itself be written as the difference in the products of two functions. Thus,

$$(E_c + e_g)^h = \sum_{q=0}^{h} \frac{h(h-1)\cdots(h-q+1)}{q!} E_c{}^{(h-q)} e_g{}^q = \sum_{q=0}^{h} r_{q+1} E_c{}^{(h-q)} e_g{}^q$$

$$(E_c - e_g)^h = \sum_{q=0}^{h} (-1)^q \frac{h(h-1)\cdots(h-q+1)}{p!} E_c{}^{h-q} e_g{}^q = \sum_{q=0}^{h} (-1)^q r_{q+1} E_c{}^{(h-q)} e_g{}^q$$

$$\left[E_{bb} - (i_1 - i_2)\frac{R_L}{4} \right]^k = \sum_{p=0}^{k} (-1)^p \left[\frac{k(k-1)\cdots(k-p+1)}{p!} \right] E_{bb}{}^{(k-p)} \left[(i_1 - i_2)\frac{R_L}{4} \right]^p = \sum_{p=0}^{k} (-1)^p S_{p+1} E_{bb}{}^{(k-p)} \left[(i_1 - i_2)\frac{R_L}{4} \right]^p$$

$$\left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right]^k = \sum_0^k \left[\frac{k(k-1)\cdots(k-p+1)}{p!}\right] E_{bb}{}^{(k-p)} \left[(i_1 - i_2)\frac{R_L}{4}\right]^p$$

$$= \sum_0^k S_{p+1} E_{bb}{}^{(k-p)} \left[(i_1 - i_2)\frac{R_L}{4}\right]^p \qquad (36)$$

Substituting the functions of Eq. (36) in the representative term of Eq. (35), we obtain

$$\frac{A_{hk}}{2}(E_c + e_g)^h \left[E_{bb} - (i_1 - i_2)\frac{R_L}{4}\right]^k - \frac{A_{hk}}{2}(E_c - e_g)^h \left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right]^k$$

$$= \frac{A_{hk}}{2}\left\{\sum_0^h r_{q+1} E_c{}^{(h-q)} e_q{}^q \sum_0^k (-1)^p S_{p+1} E_{bb}{}^{(k-p)} \left[(i_{p_1} - i_{p_2})\frac{Z_L}{4}\right]^p \right.$$

$$\left. - \sum_0^h (-1)^q r_{q+1} E_c{}^{(h-q)} e_s{}^q \sum_0^k S_{p+1} E_{bb}{}^{(k-p)} \left[(i_{p_1} - i_{p_2})\frac{Z_L}{4}\right]\right\} \qquad (37)$$

It will be noted that for all even values of p and q the terms of Eq. (37) cancel, whereas for all odd values of p and q the terms add. Since q is the exponent of e_g and p of $(i_1 - i_2)\left(\frac{R_L}{4}\right)$, it will be seen that only odd-order terms are present in the output current of a push-pull amplifier, which is a verification of the results previously obtained for the more general case.

Here we do not desire odd-order terms of degree higher than the first; hence, possible combination for q and p are

$$\left.\begin{array}{ll} q = 0, & p = 0 \\ q = 1, & p = 0 \\ q = 1, & p = 1 \\ q = 2, & p = 0 \\ q = 0, & p = 2 \end{array}\right\} \qquad (38)$$

Since q takes on all integral values from 0 to h and p all integral values from 0 to k, we see that the pairs of values given in Eq. (38)

apply to h and k too, hence also to m and n, respectively. The power series for our ideal push-pull tubes must therefore be of the form

$$\left.\begin{aligned} i_1 &= A_{00} + A_{10}(E_c + e_g) + A_{01}\left[E_{bb} - (i_1 - i_2)\frac{R_L}{4}\right] \\ &\quad + A_{11}(E_c + e_g)\left[E_{bb} - (i_1 - i_2)\frac{R_L}{4}\right] + A_{20}(E_c + e_g)^2 \\ &\quad + A_{02}\left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right]^2 \\ i_2 &= A_{00} + A_{10}(E_c - e_g) + A_{01}\left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right] \\ &\quad + A_{11}(E_c - e_g)\left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right] + A_{20}(E_c - e_g)^2 \\ &\quad + A_{02}\left[E_{bb} + (i_1 - i_2)\frac{R_L}{4}\right]^2 \end{aligned}\right\} \quad (39)$$

or either tube's current may be expressed in terms of its electrode voltages as

$$i_b = A_{00} + A_{10}e_c + A_{01}e_b + A_{11}e_ce_b + A_{20}e_c^2 + A_{02}e_b^2 \quad (40)$$

Equation (40) is that of a quadric surface or conicoid in space. in general translated from the i_b, e_c, and e_b coordinate axes and rotated in the e_b-e_c coordinate plane. From physical reasoning we know that it is valid only for positive values of e_b and for such combinations of e_b and e_c as give values of i_b equal to zero or positive; *i.e.*, it is not valid below plate-current cutoff.

However, not all forms of this quadric surface are physically realizable in vacuum tubes. Thus, if $A_{11}^2 - 4A_{20}A_{02}$ is negative, then, by the principles of solid analytic geometry, Eq. (40) represents an elliptic paraboloid, which is a closed surface. It would therefore produce families of curves on the coordinate planes that are within finite limits of plate and grid voltages and hence would not correspond to those of any (at present) physically realizable tube.

If $A_{11}^2 - 4A_{20}A_{02}$ is positive, Eq. (40) represents a hyperbolic paraboloid translated from the e_b, e_c, and i_b axes and rotated with respect to the first two (Fig. 90). It would produce realizable curves upon the e_b-i_b and e_c-i_b planes (Fig. 91) and would represent the surface of a variable-μ tube. This is an interesting result in that it indicates that variable-μ tubes having parabolic character-

istics have a distortionless output when in push-pull, class A. Such tubes could be used in a volume expander without causing any distortion (perfect matching of the tubes is inferred, but is difficult to obtain in practice).

19. Constant-μ Parabolic Tubes.—Another form for Eq. (40) of practical interest is that in which

$$A_{11} = 2\sqrt{A_{20}A_{02}}$$

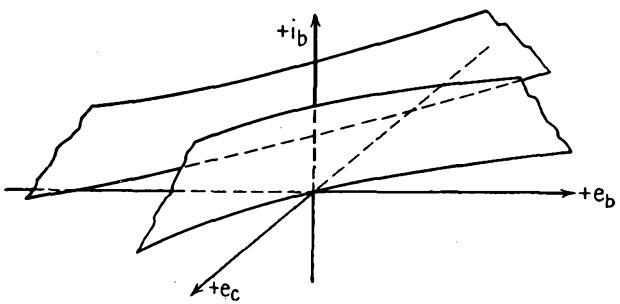


FIG. 90.—Tube surface of the form of a hyperbolic paraboloid.

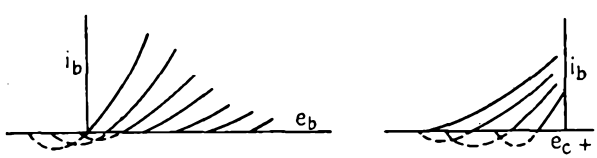


FIG. 91.—Plate and grid characteristics for a hyperbolic paraboloid.

(in which case the second-degree terms form a perfect square) and

$$\left(\frac{A_{10}}{A_{01}}\right)^2 = \frac{A_{20}}{A_{02}} = \mu$$

a constant to be identified with the amplification factor of the tube. Under these two conditions Eq. (40) becomes

$$A_{01}(\mu e_c + e_b) + A_{02}(\mu e_c + e_b)^2 = i_b \qquad (41)$$

which is a parabolic cylinder (Fig. 92) and gives rise to a family of parabolas with equidistant spacing upon the e_b-i_b and e_c-i_b coordinate planes (Fig. 93). If it is remembered that the general definition of the amplification factor is

$$\mu = \frac{\partial i_b/\partial e_c}{\partial i_b/\partial e_b} \qquad (42)$$

and that the A's of the power series are formed according to the law of a Maclaurin expansion, *viz.*,

$$A_{hk} = \frac{1}{h!k!} \frac{\partial^{h+k} i_b}{\partial e_c{}^h \, \partial e_b{}^k} \tag{43}$$

then, if μ is assumed constant, it can be shown that

$$A_{10} = \mu A_{01}$$

$$A_{11} = 2 \frac{\partial^2 i_b}{\partial e_c \, \partial e_b} = 2\mu \frac{\partial^2 i_b}{\partial e_b{}^2} = 2\mu A_{02}$$

and

$$A_{20} = \mu^2 A_{02}$$

so that

$$\frac{A_{20}}{A_{02}} = \mu^2 = \left(\frac{A_{10}}{A_{01}}\right)^2 \tag{44}$$

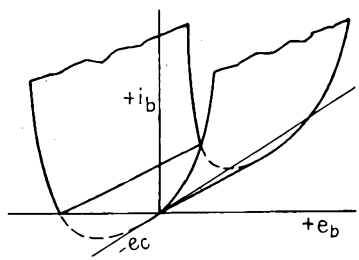


FIG. 92.—Tube surface of the form of a parabolic cylinder.

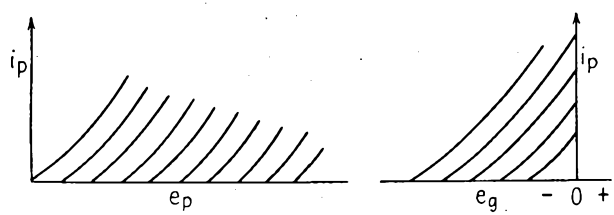


FIG. 93.—Plate and grid characteristics for a parabolic cylinder.

which are exactly the assumptions made in order that Eq. (40) reduce to Eq. (41).

It is an experimental fact that in most triodes the tube surface is tangent to the e_c-e_b coordinate plane so that

$$\frac{\partial i_b}{\partial e_c} = \frac{\partial i_b}{\partial e_b} = 0,$$

where i_b is zero, and Eq. (41) becomes

$$i_b = A_{02}(\mu e_c + e_b)^2 \tag{45}$$

which is the well-known van der Bijl's equation.

20. Properties of Square-law Tube.—The square-law type of tube has some interesting properties. Referring to Eq. (45) we may write for the balanced-amplifier circuit having a resistive load R_L

$$\begin{aligned} i_1 &= A_{02}\left[E_{bb} - (i_1 - i_2)\left(\frac{R_L}{4}\right) + \mu e_c\right]^2 \\ &= A_{02}\left[E - (i_1 - i_2)\left(\frac{R_L}{4}\right) + \mu e_g\right]^2 \end{aligned} \tag{46}$$

where $E = E_{bb} + \mu E_c$ is the equivalent diode voltage for the d.c. component and

$$i_2 = A_{02}\left[E + (i_1 - i_2)\left(\frac{R_L}{4}\right) - \mu e_g\right]^2 \tag{47}$$

From Eqs. (46) and (47) we obtain

$$(i_1 - i_2) = \frac{4A_{02}E}{1 + A_{02}ER_L}\mu e_g = 2i_L \tag{48}$$

If we substitute the value of $i_1 - i_2$ from Eq. (48) in (47) and set i_2 equal to zero, we obtain the value of grid swing for cutoff of i_2, *viz.*,

$$\mu e_g = E(1 + A_{02}ER_L) \tag{49}$$

If we substitute the latter value of μe_g in either Eq. (46) or Eq. (48), we obtain the value of i_1 where i_2 cuts off, *viz.*,

$$i_1 = 4A_{02}E^2 \tag{50}$$

At this point the plate voltage has dropped by the amount [from Eq. (50)]

$$i_1\frac{R_L}{4} = A_{02}E^2R_L \tag{51}$$

and the effective voltage has been changed from E by the amount of the corresponding grid swing given by Eq. (49), or is now equal to $E(2 + A_{02}ER_L)$. Let the remainder of the grid swing be designated by e_g'; then

$$i_1 = A_{02}\left[E(2 + A_{02}ER_L) - \frac{i_1R_L}{4} + \mu e_g'\right]^2 = 2i_L \tag{52}$$

The reader can check from this equation that i_1 equals $4A_{02}E^2$ when $\mu e_g'$ is zero.

Equation (48) gives the relation between the load current i_L and μe_g above cutoff of i_2 (class A operation), while Eq. (52) gives the relation below or beyond cutoff of i_2 (class AB operation), for the particular tube given by Eq. (45).

The slope of i_L above cutoff is evidently [differentiating Eq. (48)]

$$\frac{di_L}{d\mu e_g} = \frac{2A_{02}E}{1 + A_{02}ER_L} \tag{53}$$

The slope beyond cutoff, found by differentiating Eq. (52), is

$$\frac{di_L}{d\mu e_g'} = \frac{A_{02}[2E + \mu e_g' + A_{02}E^2R_L - (i_LR_L/2)]}{1 + (A_{02}/2)[2E + \mu e_g' + A_{02}E^2R_L - (i_LR_L/2)]} \tag{54}$$

When $\mu e_g'$ equals zero, this becomes

$$\left.\frac{di_L}{d\mu e_g'}\right]_0 = \frac{2A_{02}E}{1 + A_{02}ER_L} \tag{55}$$

which is the value given by Eq. (53). That is, the class A characteristic blends smoothly into the class AB characteristic for this type of tube, and hence the over-all characteristic may be assumed to have fewer higher harmonics than one exhibiting a sharp break at this point. This matter may be conveniently represented graphically by plotting one tube's characteristics inverted with respect to the other, but so that their operating points line up, as described in Sec. 4. Thus, in Fig. 94, $i_1 - i_2$ is tangent at cutoff to the i_1 and i_2 curves. The broken lines indicate the individual tube characteristics if they were parabolic beyond cutoff. It will be noted from Eq. (50) that i_1 at cutoff is independent of R_L: the smaller the latter is, the faster i_1 increases, but also i_2 to cutoff. Also, the blending shown in Fig. 94 is true regardless of the value of R_L or $E(=E_{bb} + \mu E_c)$: the greater the bias, or the smaller R_L is, the sooner does cutoff begin, and the sooner does the characteristic become i_1 alone and continue parabolically as shown previously or given by Eq. (52). The latter parabolic departure from linearity is not great unless the grid swing is very much beyond cutoff and is moreover upward and in a direction to offset the effects of grid current. Hence the relation between load current i_L and grid signal voltage e_g may be linear, for quite a large value of the latter.

Equation (49) shows that the cutoff grid voltage is smaller, the smaller R_L or E. The smaller E is, the greater E_c is relative to E_{bb} for a given μ. The significance is that class AB operation, *i.e.*, operation up to and beyond cutoff, is obtained either by means of overbias or by using a low value of R_L, as stated more generally in Sec. 5. At the same time, the preceding paragraph has indicated that the distortion products may nevertheless be small under these conditions. As may be recalled from Sec. 5, the value of overbiasing is that a high power-supply voltage E_{bb} may be used and yet the quiescent d.c. component I_b may be kept

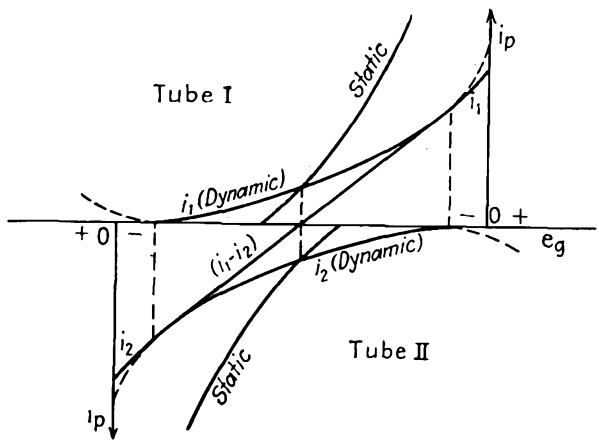


FIG. 94.—Push-pull operation beyond cutoff for parabolic tubes.

down to where the plate dissipation at no signal is within safe limits. The possible increase, thereby, of E_{bb} results in greater maximum power output.

As was also mentioned in Sec. 5, the value of being able to use low values of R_L is that the latter can be chosen for maximum power output for a given E_{bb} without the restriction (so cramping in single-side amplifiers) that the plate currents do not approach too near cutoff and produce too much distortion.

21. Optimum Value of Load Resistance.—The optimum value of R_L for class A operation is easily formulated. Thus, for values of μe_g below cutoff [given by (49)], and from (48), we have that the power output is

$$P_o = \frac{(i_1 - i_2)^2 R_L}{8} = \frac{2A_{02}^2 E^2 \mu^2 e_g^2 R_L}{(1 + A_{02} E R_L)^2} \tag{56}$$

The optimum value of R_L occurs where $\partial P_o/\partial R_L = 0$, or

$$R_L = A_{02}E = 2R_p \tag{57}$$

since the plate resistance R_p is given by

$$R_p = \frac{\partial e_b}{\partial i_b} = \frac{1}{2}A_{02}E \tag{58}$$

It is to be noted that the above value of R_p is that at the operating point.

If the grid swing is greater than the cutoff value (class AB_1 or AB_2), we have seen that we must modify the above results since the two tubes are in the picture for only part of the grid swing. A complete mathematical treatment is involved and unnecessary, since actual tubes are only approximately parabolic in their characteristics. In Sec. 10 it was shown that

$$R_L = 4R_p \tag{59}$$

where R_p is now the plate resistance at peak swing.

22. Self-bias.—Where self-bias is employed, the change in d.c. component must be small; otherwise, the bias will be increased with impress of e_g, and this in turn means a lower operating point and decreased output. It is assumed, of course, that the bias resistor is adequately by-passed; otherwise, as pointed out above, odd harmonics will be generated too. A small change in d.c. component is obtained by using a higher value of R_L than for fixed bias, since then the cutoff grid voltage is prolonged and the self-rectification in each tube is less. The grid swing generally has to be decreased too, to minimize the latter effect. The result of increasing R_L from its optimum value and reducing e_g is to decrease the power output; hence, fixed bias is preferable when available.

23. Mid-branch and Winding Resistances.—It is of interest to show, analytically, how the individual tube's load line is straightened out by a mid-branch resistance R_b. The plate voltage can be written

$$e_{b_1} = \left[E_{bb} - (i_1 + i_2)R_b - (i_1 - i_2)\frac{R_L}{4} \right] \tag{60}$$

If R_b is equal to $R_L/4$, Eq. (60) becomes

$$e_{b_1} = E_{bb} - i_1\frac{R_L}{2} \tag{61}$$

or the plate voltage of tube I is not dependent upon the plate current of tube II—i_{b_2}. This means that that particular value of $R_b(= R_L/4)$ has decoupled tube I from tube II by introducing a coupling equal and opposite to that of the two halves of the output transformer. The load line for either tube will therefore be a straight line determined by the resistance $R_L/2$ and is tangent at the operating point to the curved load line of either tube when there is no mid-branch resistance.

In the case of winding resistance in the output transformer, the corrections as detailed in the previous section give the individual tubes' currents. From these the power output can be calculated, but it must be remembered that this is the output into the transformer. To obtain the power output into the load resistance R_L, the losses in the output transformer must be subtracted.

24. Summary of Plate-circuit Relations.—In this chapter there have been formulated some general theorems on balanced-amplifier circuits. Assuming truly balanced conditions, it has been shown that

1. The output contains odd-order but no even-order modulation products.

2. The mid-branch current contains even-order but no odd-order modulation products.

3. Mid-branch impedances produce voltages that cross modulate with the signal voltage to produce odd-order modulation products in the output, in vector addition to those normally produced by the tubes themselves.

4. Mid-branch voltages produce no output directly but do cross modulate with the signal voltage to produce odd-order terms in the output.

5. A tube whose power series contains no term higher than the second will give distortionless output in a balanced amplifier operating class A. This is true even if its amplification factor is variable, subject to the above restrictions.

6. In the case where the above type of tube has a constant amplification factor, even in class AB a nearly distortionless output can be obtained. In particular, the square-law tube (van der Bijl's equation) gives a smooth dynamic characteristic for the above operation, which indicates a lack of high-order modulation products. This is true for all reasonable values of load resistance, bias, and signal voltage.

7. The above tube can give greater output in conjunction with a similar tube in balanced-amplifier operation than in single-side operation, because a higher supply voltage can be used, with overbias to keep the no-signal plate dissipation down. Also, an optimum value of load resistance can be used, even though either tube cuts off before the peak of the signal swing is reached. Thus, for this tube operating class *A*, when the load resistance (plate-to-plate) is twice the tube resistance (at the operating point), maximum power output is obtained. If operation beyond class A is desired, *i.e.*, class *AB*, maximum power output is obtained when the load resistance is four times the tube resistance at the point of peak signal swing. The above results are approximately true for tubes of parabolic characteristics.

8. Maximum grid swing is determined by the grid current. The latter depends, however, upon the plate voltage, which is at a minimum when the grid voltage is at a maximum. The value to which the plate voltage drops is determined by the load resistance as well; hence the peak grid swing is best determined graphically for the value of load resistance determined previously.

9. The full-signal plate dissipation depends upon a number of factors, so that it is best calculated after the additional d.c. component has been determined graphically. It can be reduced to within safe limits by reducing the grid swing, increasing the load resistance, or increasing the bias. The latter is often done, which results in more nearly class *B* operation. The advantage is better all-day operating economy; the disadvantage is mainly higher percentage of distortion products.

10. For self-bias, the change in d.c. component (hence change in bias) must be minimized. This can be done by operating more nearly class *A*, which means generally a higher value of load resistance and reduced signal voltage, for the same percentage of distortion products.

25. Desirability of Driving Grids Positive.—At this point in the discussion of balanced amplifiers the analysis of the grid circuit will be developed. If the grids were not to be driven positive, then the input circuit for the balanced-amplifier stage would not differ materially from that of the single-side circuit. The only additional precaution of any consequence would be that of ensuring that the two grid voltages (in the two halves of the secondary of the input transformer) were in phase opposition at

all frequencies. This is because the balanced-amplifier input transformer is essentially a three-winding transformer, and it is therefore possible for the two secondary voltages to be other than in phase opposition, particularly at the higher frequencies due to unbalances between the distributed capacities of the windings and mutual inductances between the two halves of the secondary as well as between each half and the primary.

However, the power output of a balanced amplifier under the above conditions is relatively low; and while the plate-supply voltage may be increased (and the C bias also increased in order concomitantly to keep the plate dissipation within prescribed limits, especially at no signal) and thus more power output obtained, there is an upper limit to this increase that is determined by the insulation strength of the tube and the cost of the power supply, particularly the filter condensers.

Since, as has been shown previously, balanced-amplifier operation is not limited by plate-current cutoff, it would appear equally desirable to remove the other restriction necessary for class A operation, that the grids be not driven positive. Therefore, experiments have been directed toward the expanding of the range of operation, with the result that, under proper design, grid swings of amplitude sufficient to drive the grids positive have been successfully employed. Thus, there has been a large increase in power output and also in plate efficiency, and, in addition, the obtaining of large power outputs at reasonably low B voltages. This mode of operation has been previously defined as class AB_2 (Sec. 5).

26. Driver-tube Considerations.—It is a fundamental fact in the theory of electrical circuits that, when current flows in the same direction as the voltage acting in that circuit, power is absorbed by the latter. In the case of the grid circuit, whenever the grid is driven positive, electron flow is from the cathode to the grid and then externally around the grid circuit back to the cathode, which means that the conventional flow of electric current is in the same direction as the grid signal voltage. Hence, it represents energy absorbed in the grid circuit, and this energy must come from the source of signal voltage, *i.e.*, the tube preceding the power-amplifier stage. We therefore see that the latter tube has been elevated from the role of a voltage-amplifier stage to that of a power-amplifier stage. Therefore, it is usually

called a *driver tube; i.e.*, it drives the grids of the balanced-amplifier stage alternately positive and supplies the electrical power absorbed by them when positive. The driver tube must therefore have a power rating adequate for the grid requirements and indeed, as will be shown, must have more than this rating in order that the distortion produced in the grid circuit be kept within allowable limits.

It may be well to elaborate the latter point. The load fed by the driver tube, *viz.*, the grid circuit of the balanced-amplifier stage, is very nonlinear in its characteristic. For example, if the grids are negatively biased, no current flows until they are actually driven positive by the driver, so that they appear to the latter as an infinite load resistance for the early part of the signal cycle. Then in that portion of the cycle where they are driven positive they suddenly appear as a finite resistance, which, however, is not constant, for the grid current is not even in direct proportion to the positive grid voltage. It is therefore evident that the voltage drops in the driver and associated coupling transformer will depend upon this variable grid current and hence the terminal voltage across either half of the transformer will be distorted by the internal drops in it and the driver tube even though an equivalent sinusoidal voltage be generated in the tube itself.

Our problem is therefore twofold:

1. To determine the grid-current flow during the portion of the cycle of the signal voltage when the grid is positive:

2. To determine the permissible internal impedance of the driver source for a prescribed allowable distortion of the grid signal voltage. It will be found that this consideration indicates a tube of a size greater than grid power (as averaged over the entire signal cycle) would in itself require. In short, driver internal regulation is the determining factor in the choice of the size of tube, rather than grid power considerations.

We shall discuss problem 1 first, in Sec. 27.

27. Determination of Grid Current.—The instantaneous grid current is a function both of the grid and the plate voltage. The relative effect of the plate voltage, as compared with the grid voltage, upon the grid current has been called the "reflex factor" and, while analogous to the amplification factor, is practically always of a value less than unity. However, under extreme

conditions of operation it is possible for the plate voltage to have at some instant a large effect upon the grid current, due to the relative division of the space current between the two electrodes and also to secondary emission and even possible dynatron effects occurring either at the plate or at the grid. Consequently, any attempt to determine the grid-current vs. grid-voltage characteristic must take into account the plate voltage. The latter in turn depends upon the actual grid-to-cathode voltage and load impedance (as well as any regulation in the B power supply).

Thus, it would seem that a method of simultaneous solutions must be employed. However, it is possible first to solve the

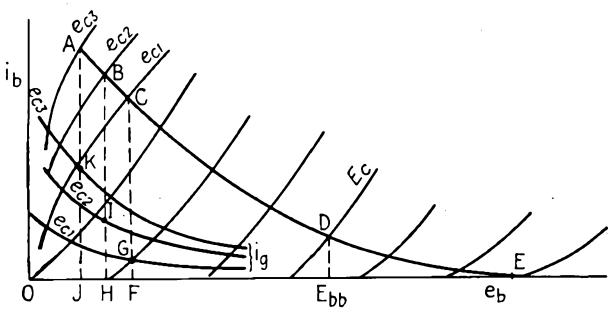


FIG. 95.—Determination of operating grid current.

plate-circuit condition (as given in the previous sections) upon the assumption of a certain form of grid voltage (usually sinusoidal) and then determine the alteration of grid voltage permissible as far as distortion of the output of the plate circuit is concerned. This is tantamount to assuming, as a first approximation, that the driver internal impedance is zero. Hence, if the load line for the balanced amplifier has been determined for this form of grid voltage, the grid current can then be found. For this purpose, it is convenient to plot the latter as a function of the plate voltage, with the grid voltage as a parameter. This gives rise to a family of curves that is superimposed on the plate family as shown in Fig. 95.

It will be noted that for a very positive grid voltage e_{c_3} the grid current at any plate voltage is higher than for a lower positive grid voltage e_{c_1}. It will also be noted that the curve for any one positive grid voltage rises as the plate voltage decreases. As mentioned previously, this is due to the fact that the grid, being

closer to the cathode, diverts space current to itself. As a result, the plate-current curve drops sharply downward where the corresponding grid-current curve rises sharply upward; hence, the load resistance R_L must be chosen low enough so that its curved push-pull load line cuts the plate family of curves above the region where they thus droop sharply, or otherwise excessive distortion will result.

If the grid voltage is more than twenty volts positive or thereabouts, secondary emission may occur at this electrode. This phenomenon is not well understood as yet, although at present there is considerable research to determine its characteristics and quantitative relations. The amount of secondary emission depends not only upon the potential of the electrode but also upon its substance and surface condition. By suitably treating the surface of a grid, the secondary emission may be considerably reduced.

When secondary emission occurs at the grid and the plate at that instant is more positive than the grid, secondaries may go from the latter to the plate. The result is that the grid current may cease to rise with increase of grid potential, or it may rise less rapidly than for lower grid potentials, or, in extreme cases, it may even decrease, sometimes even to the point of going negative. The latter case gives rise to a dynatron action; the grid-to-cathode variational resistance is negative, and at that moment the grid or input circuit may oscillate (depending upon the positive damping in this circuit).

The plate current will reflect these oscillations in its wave shape, and the result is distortion that is not even of harmonic frequency with respect to the fundamental. Such dynatron action may occur if the plate voltage is high when the grid swings sufficiently positive and will reveal itself by the shape of the right-hand portion of the grid-current curves.

However, owing to the load resistance, the plate current increases in the tube whose grid is swinging positive, and the plate voltage is concomitantly falling, as indicated, for instance, in Fig. 95. That is, when the grid e_g is at its peak positive value, then the plate voltage at that instant is at its minimum value; hence, such dynatron action can occur only if e_b is sufficiently high even when at its minimum value. This is not usually the case unless a very low value of load resistance is employed or a very high value of

B voltage. Consequently, this action may be expected more in large high-voltage tubes than in small tubes used for radio receivers, etc. Of course, the surface of the grid also has an important bearing on the matter.

In extreme cases, dynatron action may occur at the plate. The result is ordinarily a flattening of the peak of the plate-current wave, although, for high values of load resistance and large positive grid swings, the plate-current wave may show at its center a subsidiary minimum instead of a peak. This, of course, means that, as the grid swings positive to its peak value, the plate current at first increases but then decreases, owing to the grid diverting the space current more completely to itself and also possibly to its robbing the plate of secondaries, if, at that instant, the grid is actually at a higher potential than the plate, with respect to the cathode. The latter condition would indicate a very large grid swing and high load resistance, a condition not usually encountered in audio amplifiers, particularly if operation is confined to the region mentioned in the previous sections of this chapter.

After this brief discussion we are now ready to plot the grid current per tube for a given plate load resistance. In Fig. 95, $ABCDE$ is the load line presented to each tube by a load resistance R_L (plate-to-plate). It intersects the plate-current curves for successively higher positive values of the grid parameter in C, B, and A, respectively, as shown. The corresponding instantaneous values of plate current and plate voltage are CF, BH, AJ, and OF, OH, OJ, respectively. Thus, at maximum grid swing e_{c_3} the plate voltage has dropped to its minimum value OJ from its normal value at no signal of OE_{bb}. For that value of plate voltage OJ and for that grid voltage e_{c_3}, the grid current at that instant is evidently KJ. Similarly, for the grid swing e_{c_2} and plate voltage OH, the grid current is IH; for e_{c_1} and OF, it is GF. Thus, the actual grid current i_g for different grid voltages and the given value of R_L can be found. We now plot i_g vs. e_c or i_g against the signal voltage e_g (either grid to cathode) as shown in Fig. 96. On this same graph the load current $i_L = (i_1 - i_2)/2$ through R_L can also be plotted, as detailed in the previous sections. It is necessary to plot only one half cycle of e_g, as the other half cycle gives identical results from the symmetry of the circuit.

28. Plate-circuit Distortion Products.—We are now ready to attack the second half of the problem stated at the end of Sec. 26. We first assume that the driver internal impedance is zero. In this case, the actual signal voltage e_g is identical with the known voltage generated within the driver despite the fact that grid current flows. Thus, the load line for i_L shown in Fig. 97 represents at the same time a plot between the load current and the generated voltage of the driver. If this load line is straight, then, as mentioned previously, the load current is an exact copy of the generated driver voltage and no distortion results. However, this graph is usually not straight for the entire range of grid

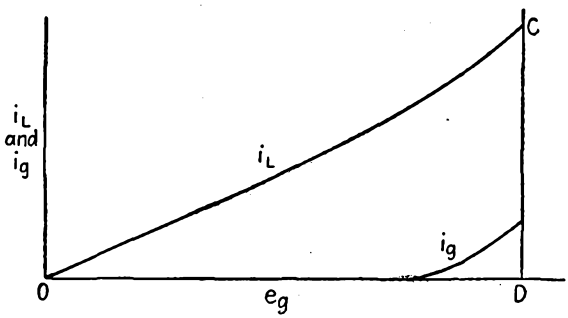


FIG. 96.—Grid- and load-current characteristics for a balanced amplifier.

swing so that we know that distortion will be present. Our problem is to determine, at least approximately, the amount of distortion corresponding to a certain degree of departure of this load line from linearity. In the case of a triode operating class *AB*, the top end of the load line rises above the linear line *OA* (Fig. 97) and is called an *overshoot*. In the case of a pentode, it is quite possible for the load line to be below *OA*, *viz.*, *OB*. Such a load line may be said to have an *undershoot*.

To determine the distortion products in either load line we assume a power-series relation between i_L and e_g, *viz.*,

$$i_L = ae_g + ce_g^3 \tag{62}$$

It will be noted that the square term is missing. It has been shown, Sec. 14, that there can be no even-order modulation products in the output of a perfectly balanced push-pull amplifier but that there can be odd-order products, which in the case of a resistive load corresponds to a power-series relationship between

i_L and e_g involving only odd powers of e_g. As a fairly satisfactory approximation, we assume that no terms above the third degree are present. This means that we assume that all the distortion is third harmonic if e_g is sinusoidal in form. If the coefficient c is positive, then the plot of this curve gives rise to load line OC (Fig. 97), whereas, if negative, it gives rise to load line OB. We therefore see that by choosing the proper sign for c we can represent, to a fair degree of accuracy, load lines having overshoots

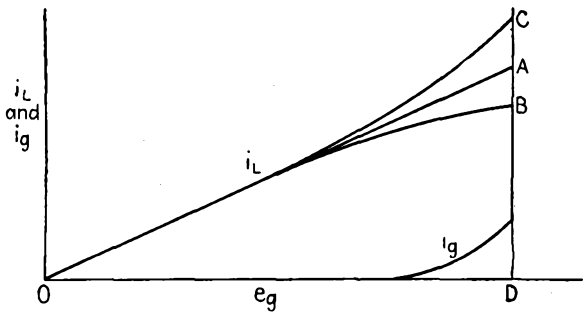


FIG. 97.—Three types of balanced-amplifier operation: *A*, linear; *B*, undershoot; *C*, overshoot.

and load lines having undershoots. In the case of a linear load line it is evident that c is zero.

Now let

$$e_g = E_g \sin \omega t \tag{63}$$

Substituting Eq. (63) in Eq. (62) we obtain

$$i_L = aE_g \sin \omega t + cE_g^3 \sin^3 \omega t \tag{64}$$

From trigonometry, we know that

$$\sin^3 \omega t = \tfrac{3}{4} \sin \omega t - \tfrac{1}{4} \sin 3\omega t \tag{65}$$

Substituting Eq. (65) in Eq. (64) we finally obtain

$$i_L = (aE_g + \tfrac{3}{4}cE_g^3) \sin \omega t - \tfrac{1}{4}cE_g^3 \sin 3\omega t \tag{66}$$

The quantity $\frac{3}{4}cE_g^3 \sin \omega t$ represents the contribution of the third-order, or cubic, term to the fundamental frequency. If c is positive, it represents an increase in the fundamental component over that given by a linear load line; if negative, it represents a decrease. The remaining term $\frac{1}{4}cE_g^3 \sin 3\omega t$ represents third-harmonic distortion produced by the third-order term.

29. Geometric Interpretation.—Referring once again to Eq. (62) we see that the quantity cE_g^3 is represented in Fig. 97 by distances CA or AB if c is positive or negative, respectively, and at the moment when e_g is at its peak value. The different wave shapes are shown in Fig. 98*A*, *B*, and *C*. Figure 98*A* shows the wave shape for an overshoot. The broken line shows the actual peak wave shape, and the solid line superimposed on it shows the amplitude of the fundamental component whose peak is $\frac{1}{4}CA$ below the peak of the actual wave. The third-harmonic distortion is shown in its proper phase about the axis. Its peak ampli-

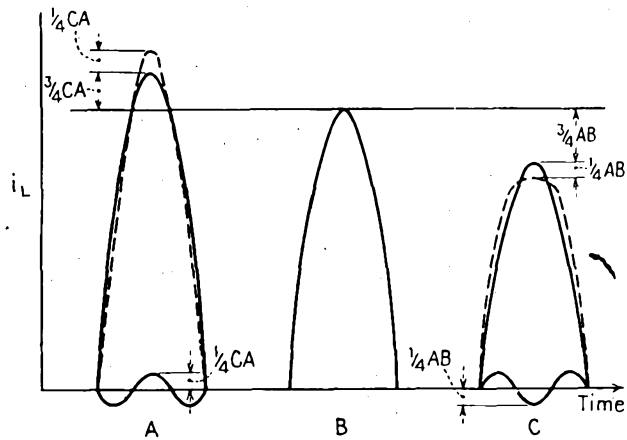


FIG. 98.—Wave shapes corresponding to *A*, overshoot; *B*, linear; *C*, undershoot in push-pull characteristics.

tude is $\frac{1}{4}CA$. In Fig. 98*B* the actual wave shape and fundamental component are identical, and the third-harmonic component is zero. In Fig. 98*C* the actual wave (broken line) has a peak amplitude $\frac{1}{4}AB$ below the fundamental peak; the latter, in turn, is $\frac{3}{4}AB$ below the fundamental peak of Fig. 98*B*. The third-harmonic distortion in the last example has a peak amplitude $\frac{1}{4}AB$, and its phase is reversed with respect to that of Fig. 98*A* because the sign of c is reversed (negative for undershoot).

It is therefore evident that the departure of the load line from linearity enables us to estimate the third-harmonic distortion produced thereby and, conversely, if the permissible amount of third-harmonic distortion is specified, the departure from linearity at the peak can be found.

The effect of internal driver resistance is to cause a flattening of the load current at its peak. It may therefore be regarded as equivalent to the case of a zero-resistance driver and a load-current undershoot. That is, either we can regard the driver resistance as zero, the terminal (actual grid voltage) and generated voltages as being identical, and the load current as having a plot such as *OB* (Fig. 97), or we may consider (as is actually the case) that the load line is *OA* or possibly even *OC* but that the terminal or grid voltage is flattened from the generated sinusoidal shape owing to the voltage drop produced by the grid current flowing through the internal resistance of the driver. Either viewpoint will give the same peak load current at the end of the first quarter of the cycle when the generated driver voltage has reached its maximum. Externally, the effect of driver resistance is to convert a tube that might otherwise exhibit an overshoot into a tube that exhibits an undershoot. We are now in a position to determine the permissible driver internal resistance.

30. Determination of Driver Resistance.

Let P_o = desired fundamental power output
I_{LF} = peak fundamental current which, flowing through
R_L = plate-to-plate load resistance, gives rise to P_o
n = permissible percentage of distortion (assumed all third harmonic). It is therefore the ratio of peak third-harmonic current to I_{LF}.

Then

$$P_o = (I_{LF})^2 \frac{R_L}{2} \tag{67}$$

or

$$I_{LF} = \sqrt{\frac{2P_o}{R_L}} \tag{68}$$

which determines I_{LF} for a given P_o and R_L.

Suppose (Fig. 99) that *OC* represents the load line for the load current. Through *O* a straight line *OA* is drawn. This should coincide as closely as possible with the lower end of *OC*. In Sec. 20 it was shown that for a parabolic tube the load line is straight up to the point of cutoff of one tube and from this point rises parabolically, since from then on the load current and the plate current of the other tube are identical. Hence, the load line is

straight up to the cutoff value of the signal voltage, and then an overshoot occurs. In actual tubes this condition is approximated so that it is usually possible to draw a line OA that coincides with the lower portion of OC. Next, a horizontal line is drawn through the ordinate I_{LF} determined by Eq. (68). Then another horizontal line is drawn through the ordinate $(1 + 3n)I_{LF}$. Third, a horizontal line is drawn through the ordinate $(1 - n)I_{LF}$.

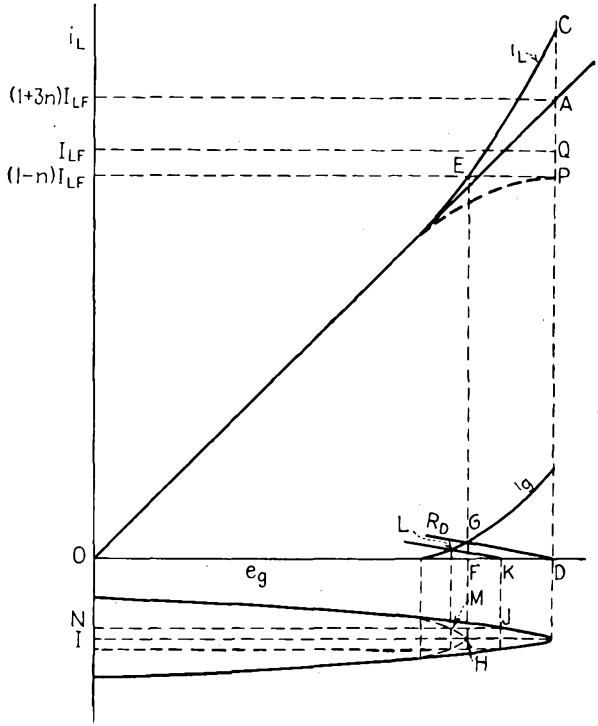


FIG. 99.—Graphical constructions for determining permissible driver resistance.

As shown in Fig. 99, the horizontal line through $(1 + 3n)I_{LF}$ intersects OA in A. OD is the peak value of the generated grid swing. The horizontal line through $(1 - n)I_{LF}$ intersects OC in E, which corresponds to a grid voltage OF. EF intersects the grid current curve in G. Then GD is the load line for the total driver resistance R_D. IH is the terminal, or actual, peak grid-to-cathode voltage of the tube which is at that moment drawing grid current. At some other moment of the generated grid-voltage cycle such as point J, the terminal voltage is found

by drawing LK parallel to GD and then projecting this down to the grid-voltage wave in point M. NM represents the terminal or grid voltage at that instant. In this way, the terminal grid-voltage wave shape can be determined as shown by the broken line. This broken line evidently coincides with the solid line for the grid signal voltage to the left of where the grid-current curve begins to rise above the e_g axis.

A study of Fig. 99 will show that the value of R_D will fulfill the requirements if the power series of Eq. (62) is sufficiently accurate. For the load line for i_L may be assumed to be OP instead of OC, as far as peak load current is concerned in the actual circuit. DP represents the peak actual current; DQ represents the peak of the fundamental component of this actual current. Load line OP undershoots the linear load line OA by a distance AP. Three-quarters of this distance, or AQ, represents reduction in the fundamental component due to the undershoot characteristic. One-quarter of AP, or QP, represents peak amplitude of third-harmonic current; and QP is evidently n per cent of I_{LF} as required, and I_{LF} or DQ is the fundamental component that will give the required fundamental output P_o.

On the other hand, if we take the actual conditions, OC is the actual relation between load current and grid signal voltage. If, at the peak swing, a peak current of FE ($= DP$) is required, the terminal grid voltage need only be OF instead of OD. Since OD is the peak generated voltage, then FD is the allowable internal voltage drop of the driver; and since for a grid voltage of OF the grid current is FG, the internal driver resistance is evidently FD divided by FG, or GD is the load line for R_D. The latter can therefore be calculated from the scales of current and voltage on the graph.

31. Driver Input Transformer.—The internal driver resistance R_D must be prorated between the winding resistances of the driver input transformer and the plate resistance of the driver tube itself. Moreover, a driver tube of adequate size must be chosen. In actual practice, the nearest standard-size tube manufactured is selected. Suppose the one chosen requires a certain B supply voltage, also a certain bias E_c, and that it has a certain μ and a certain plate resistance R_p. Assume it is to be a class A driver. Then, the equivalent voltage in the plate circuit has a maximum magnitude of μE_c.

It will be desirable, if possible, to use a grid swing less than this maximum value. This will depend upon the R_p of the tube and the winding resistances of the driver transformer. In order to reduce the number of variables present, assume that the primary winding resistance R_{pw} and the winding resistance R_{sw} of half of the secondary cause equal losses. This means that, if the turns ratio (primary to one-half secondary) is a:1, then

$$R_{pw} = a^2 R_{sw} \tag{69}$$

Now let the ratio of total winding resistance to R_p be denoted by q. A reasonable value for q is 10 per cent. Then the apparent source impedance, as viewed by the grid of either power tube, is

$$R_G = \frac{R_p + R_{pw} + a^2 R_{sw}}{a^2} = \frac{R_p(1 + q)}{a^2} \tag{70}$$

But the graphical construction of Fig. 99 requires that

$$R_G \leq R_D \tag{71}$$

If we use the equal sign, then

$$a = \sqrt{\frac{R_p(1 + q)}{R_D}} \tag{72}$$

Equation (72) thus gives the step-down ratio required in the driver transformer, after a reasonable value of q has been chosen. We are now in a position to check the suitability of the choice of driver tube.

In the midrange of frequencies for the transformer, say, around 1,000 cycles, the open-circuit reactance of the primary is sufficiently high and the leakage reactances are still sufficiently small so that, in the absence of grid current, the transformer may be regarded as practically an infinite impedance across the tube. The circuit gain is then approximately equal to the tube μ. The graphical construction gives the value of generated voltage, OD (Fig. 99), that is required. The input signal to the driver tube grid is then evidently

$$e_g = OD \frac{a}{\mu} \tag{73}$$

If e_g is less than or at most equal to the bias voltage of the driver stage, then the tube and transformer chosen for this stage are

satisfactory; if not, then a tube having a higher μ or lower R_p, that is, a higher transconductance g_m, will be necessary.

It will also be evident that the effect of the grid-current flow is to reduce the generated voltage OD to the lower terminal voltage OF, and that this latter voltage is of the proper magnitude to cause the flattening of the grid-voltage wave to the degree which permits the desired amount of power output to be obtained with the acceptable amount of distortion (assumed all third harmonic).

The driver tube is generally of a size known as a *power tube* and as such usually draws a considerable d.c. component I_b. It is rather difficult and expensive to construct a driver transformer that has a sufficiently high open-circuit reactance under these conditions, and so often a push-pull driver stage is employed, since in a push-pull stage the d.c. components of the two tubes' currents cancel each other's magnetic effects in the two primary windings. Although a push-pull family of curves could be drawn for this stage as outlined in Sec. 3 (Fig. 74), it will be sufficiently accurate to assume that the apparent internal resistance of the stage is $2R_p$ and the apparent generated voltage (on the primary side) is $2\mu e_g$. If, in addition, the turns ratio a is taken as that of the entire primary to one-half of the secondary, then Eqs. (72) and (73) will hold equally well in this case. The regulation will be half of that for a single-ended stage with the appropriate transformer.

If it is desired to employ a driver transformer furnished by some manufacturer, then its suitability may be determined by the formulas

$$R_p + R_{pw} + a^2R_{sw} \leq a^2R_D \tag{74}$$

$$\frac{\mu e_g}{a} \leq OD \tag{75}$$

32. Further Conclusions.—The material of the preceding paragraphs may now furnish us with the following conclusions:

1. This analysis is based upon the assumption that the open-circuit reactance and the leakage reactance are very high and negligibly small, respectively. This condition is probably approached by a good driver input transformer at frequencies in the neighborhood of 1,000 cycles.

2. At low frequencies the open-circuit reactance becomes lower and comparable to the R_p of the driver tube, whereupon the

generated voltage available in the grid circuit is less, so that the power output obtained will not be so high as desired. That is, the frequency response droops off at lower frequencies. On the other hand, by Thévenin's theorem, the impedance that the grid circuit of the tube sees looking back into the source is less than its value at higher frequencies by the shunting effect of the open-circuit reactance. Hence, we might expect that the effect of the open-circuit reactance is to reduce not only the power output but the amount of distortion because the grid swing is less and the equivalent internal driver impedance is less, so that the grid-terminal voltage wave is not flattened so much.

3. At higher frequencies, say, around four and five thousand cycles, we may expect the leakage reactance to have an appreciable effect upon the output and distortion products. Since it is a series reactance, it makes the internal driver impedance appear higher and therefore causes more flattening of the grid-voltage wave, hence more of an equivalent undershoot in the plate circuit and consequently less fundamental output and more distortion. Therefore, unless it is kept down to a very low figure at those frequencies, it causes the calculations given above to be considerably in error. As a consequence, we may regard these calculations for the driver and winding resistances as giving a maximum, or upper, figure, or else we may choose for our computation a lower value of n in order to be on the safe side at the higher frequencies.

The effect of leakage reactance upon the wave shape can be determined by means of the principles established in Chap. II (Sec. 6). The driver tube can be assumed to have a linear internal resistance, and this can be combined with the winding resistances to form a value R_i. This, in conjunction with the leakage reactance, forms the finite operator curve, which is slid along the grid-current–grid-voltage characteristic. The result is a loop upon the latter characteristic, and the appropriate projections of this loop over to the characteristic give the instantaneous grid or terminal voltages. In general, it will be found that for an increasing generated voltage the above instantaneous values are lower than for corresponding moments in the next quarter cycle when the generated voltage is decreasing, so that the grid voltage wave no longer has symmetry about its values at 90 and 270 deg. of the cycle. The result is both an increase

in odd-order modulation products as well as a shifting in phase of the various components over that which would occur if no reactance were present.

The converse problem, that of determining a permissible amount of leakage reactance which will give an acceptable amount of distortion, is exceedingly difficult, if at all possible of solution— at least, from a practical viewpoint. Hence, if this factor is particularly important in some problem, a series of trials of various magnitudes of reactance may be the best method of attack. On the other hand, as noted previously, if these parameters are known, then the determination of the grid-voltage wave form (hence that of the output) is a straightforward problem.

4. The method, as has been cautioned throughout this discussion, is based on the assumption that all the distortion for a sinusoidal input is third harmonic. Actually, the load line for i_L requires a more complicated power series, which means that higher harmonics are really present. However, their amplitudes are usually very small so that if a conservative value for n is taken, say, 5 per cent or less, the effects of these higher harmonics need not be feared.

5. The driver tube should preferably be of a type that is not critical as to the value of load impedance. This means that it should be preferably operated class A, and even then that it be preferably a triode rather than a pentode. A pentode is very critical as to load resistance and will give rise to high distortion products in its plate circuit if the load impedance presented to it is very variable, such as a grid-to-cathode resistance. However, this may be obviated by the use of inverse feedback of the type that makes the apparent tube resistance lower.

6. The difficulty in obtaining a driver input transformer design whose winding resistances will fall within the required limits given by Eqs. (72) and (73) depends upon the initial load line for i_L. If this has an overshoot, then, as is evident from Fig. 99, the allowable driver resistance will be greater or, what amounts to the same thing, more internal voltage drop in the driver can be allowed before this overshoot is converted into the permissible undershoot.

7. No mention has been made of the C bias source for the balanced-amplifier stage. If the internal resistance of the latter were the same both to the a.c. and d.c. components of the grid-

current wave, then it could be subtracted from the value of R_D before proceeding to calculate a and Eq. (72). Usually, however, the impedance of this source to alternating current is negligibly small because a large by-pass condenser is placed across its terminals. Hence, its only effect is to experience a voltage drop due to the d.c. component of the grid current. This, in effect, means a change in the steady bias value. To calculate this effect and consequent effects upon the grid-current wave is quite difficult. The d.c. component of the grid-current

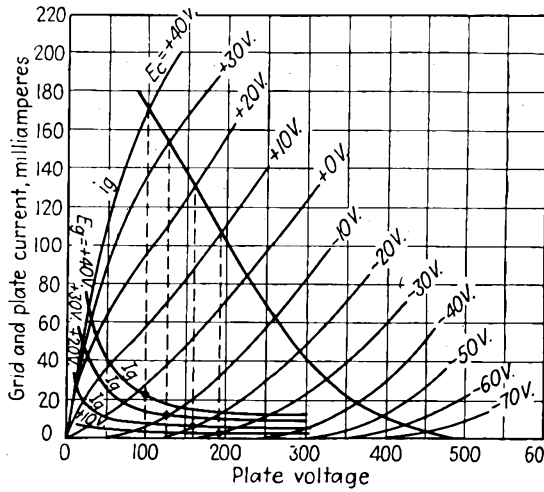


FIG. 100.—Grid-current determination for a 6F6 tube.

wave would have to be determined by a Fourier analysis, then the voltage drop in the C bias found, and then the grid-current curves shifted in a negative direction by the amount of this voltage drop. A recalculation of the voltage drop in the grid circuit would then have to be made, as well as calculation of the position of the load line for R_L, which would be shifted too. Then the correct grid-current wave would be determined, its value of d.c. component found, and another correction made to the bias. In this manner, by a method of approximations, a final state of equilibrium could be found.

8. However, the methods outlined previously will serve at least to determine feasible values for the constants of the driver stage, and at least one conclusion can be drawn, *viz.*, that the bias source should have as low a d.c. resistance as possible and

should be adequately by-passed so as to have a negligible a.c. impedance in the working range of frequencies.

33. Typical Calculations.—An illustrative example will help to make the method discussed above clearer. The 6F6 tube cited previously (page 152) will be used here. Figure 100 shows the push-pull load line as it appears to either tube. This load line intersects the +10, +20, +30, and +40 grid-voltage curves at points that are projected downward to the corresponding grid-current curves, as shown by the broken lines. The corresponding grid currents are indicated by circles. The load current i_L and the grid current i_g are then plotted against e_g (Fig. 101). A straight line is now drawn as nearly coincident with the lower part of the i_L-e_g curve as possible. It will be noted that the latter represents an overshoot compared with the straight line, although it shows signs of an ultimate undershoot for a sufficiently large grid swing.

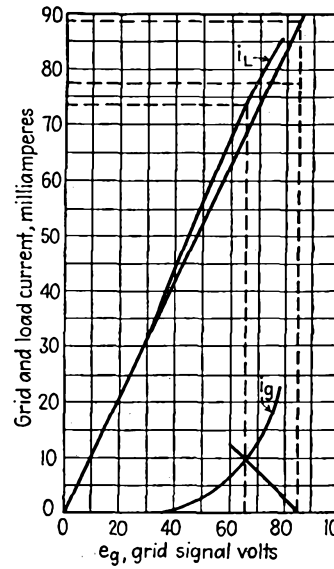


FIG. 101.—Determination of the driver resistance, 6F6 tube.

Suppose an output power P_0 of 18 watts is desired, with a third-harmonic distortion of 5 per cent (= n). The fundamental load current will therefore be

$$I_{LF} = \left(\frac{2 \times 18}{6{,}000}\right)^{1/2} \times 1{,}000 = 77.5 \text{ ma.}$$

Lines are drawn through current values of 15 per cent greater and 5 per cent less than 77.5 ma., or through 89.2 ma. and 73.6 ma., respectively. The line through 89.2 ma. intersects the linear line at a value for e_g of 84.8 volts, while the one through 73.6 ma. intersects the actual load line at a value for e_g of 65.5 volts. This corresponds to a grid current of 9.5 ma. The permissible driver resistance is therefore

$$R_D = \frac{84.8 - 65.5}{0.0095} = 2{,}030 \text{ ohms}$$

A driver tube must now be chosen. If a 6F6 triode-connected tube is chosen, it can be operated class A as follows: The B supply voltage is 250 volts; the bias, -20 volts; the plate resistance R_p is 2,600 ohms; and the amplification factor μ is 7. A value of $q = 0.1$ will be chosen. Then, by Eq. (72), the turns ratio is

$$a = \sqrt{\frac{2{,}030}{1.1 \times 2{,}600}} = 1.187:1$$

The primary winding resistance R_{pw} is $0.05 \times 2{,}600$, or 130 ohms. The secondary winding resistance R_{sw} is therefore $130 \div (1.187)^2$, or 92.3 ohms. These should be realizable in a practical transformer design.

Next the grid swing must be checked. By Eq. (73),

$$e_g = \frac{84.8 \times 1.187}{7} = 14.36 \text{ volts}$$

Since this is less than the 20 volts bias, the tube will be a satisfactory driver when used in conjunction with the above driver transformer.

It must be remembered that the leakage reactance of the driver input transformer will tend to increase the distortion products and reduce somewhat the fundamental power output. However, this should not be excessive in a carefully designed transformer. In addition, the distortion products of the driver tube will appear in the output of the balanced-amplifier stage, mainly as second harmonic. These, too, should not be excessive, since the R_p of the driver is fairly low, *viz.*, 2,600 ohms, and the grid swing of 14.36 volts is moderate.

BIBLIOGRAPHY

1. Barton, L. E.: High Audio Power from Relatively Small Tubes, *Proc. I.R.E.*, July, 1931.
2. Turner, H. M., and F. T. McNamara: *Proc. I.R.E.*, **18** (10), 1743–1747, October, 1930.
3. Peterson, E., and H. P. Evans: Modulation in Vacuum Tubes Used as Amplifiers, *Bell System Tech. Jour.*, July, 1927.
4. Kilgour, C. E.: Push-pull Amplifier Graphics, *Electronics*, March, 1933.
5. Thompson, B. J.: Graphical Determination of Performance of Push-pull Audio Amplifiers, *Proc. I.R.E.*, **21**, 591, April, 1933.

CHAPTER VI

DETECTION

1. Diodes—Rectification Curves.—The diode detector is in almost universal use today as a demodulator, but its circuit design is far removed from that of the early Fleming valve. Its present-day popularity is due to its ability to handle reasonably large signals without overlooking, its characteristic of linear detection of large signals, and ability—if required—to furnish the variable bias necessary for automatic volume control with simplicity of associated circuits.

A linear detector is one whose output is a faithful copy of the envelope of the modulated h.f. wave impressed upon it; *i.e.*, the former is directly proportional to the latter. Such detection is facilitated by the use of an ideal diode. An ideal diode is one whose impedance to current flow in one direction is zero and whose admittance to current flow in the opposite direction is zero. Actual diodes practically meet the latter condition, but not the former, since they present a finite, usually nonlinear resistance in the conductive direction.

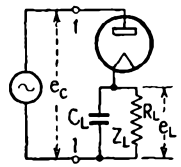


FIG. 102.—Elementary diode-detector circuit.

An elementary diode circuit is shown in Fig. 102. The h.f. source impedance is assumed zero. The condenser C_L is assumed to have negligible reactance at the high carrier frequency f_c and to have negligible susceptance at the low modulation frequency f_m. We shall first, however, assume that e_c is unmodulated, *i.e.*, of constant amplitude. The rectified output voltage e_L will therefore be d.c. Its magnitude, as well as that of the current flow through the diode, could be found by the methods described in Chap. IV (Sec. 10); but these become very involved when e_c is assumed to be modulated, and various impedances present in the actual circuit are taken into account. Accordingly, a simpler, although more approximate method, is to be sought.

For steady amplitude of e_c, e_L can be found experimentally as well as graphically. Thus, first, by the *compensation theorem*, the impedance Z_L can be replaced by a zero-impedance d.c. source e, so poled as to oppose current flow through the diode. If a constant value of e_c is impressed as well and its peak amplitude exceeds e, rectified current will flow through the diode and its d.c. component i can be measured. If e is varied from the peak value of e_c to zero, i will vary from zero to some maximum upper limit and give rise to an e-i curve similar to an ordinary i_b-e_b curve for a triode. If e_c is now adjusted to another fixed value, another curve will be obtained, etc., until a whole family is developed. This is shown in Fig. 103. The curves are approximately equidistant in spacing and are known as *rectification curves* for the diode. It will be noted that they proceed from right to left as e_c is increased and also that these curves can be obtained at a conveniently low frequency for e_c.

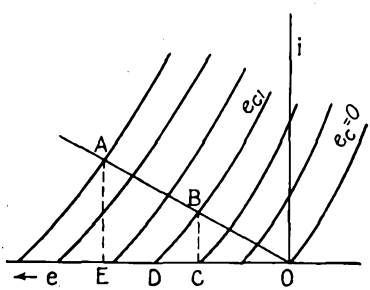


FIG. 103.—Rectification diagram and load line for a diode detector.

If e is now replaced by Z_L, i and e_L (in place of e) must satisfy the relationship that $iR_L = e_L$. For any value of e_c, i can be found from the graphical solution for R_L and the diode in series. Thus, the load line for R_L is laid off as OA, and its intersection with the diode family at B gives the output current BC for a peak alternating (carrier) voltage e_{c1}. The diode curves can represent by their slope the nonlinear internal resistance of the diode r_d. Then DO can represent an equivalent direct voltage generated in the circuit, of value equal to the maximum value of e_{c1}. The portion DC is lost as voltage drop across the diode, and the remainder CO is available across Z_L. If $Z_L \gg r_d$, then it can be seen that CO is practically equal to the maximum value of e_{c1}, and DC approaches zero. This is usually the case in actual practice. The diode thus approaches the ideal diode in performance.

Now suppose e_c represents a modulated wave, of the form

$$e_c = E_c(1 + m \sin \omega_m t) \sin \omega_c t \qquad (1)$$

where $\omega_m = 2\pi f_m$, the low modulation angular frequency, and $\omega_c = 2\pi f_c$, the higher carrier angular frequency. It is evident

from Eq. (1) that the amplitude of e_c may be regarded as made up of a constant value E_c and a variable quantity, of peak magnitude mE_c and varying at a frequency f_m. Suppose in Fig. 103 that $e_{c_1} = E_c$. Then BC and CO represent the constant components of the output current and voltage, respectively. The variation in amplitude due to mE_c produces a path of operation along AO. Assume, specifically, that $m = 1$ (100 per cent modulation) and that this carries the operation from B to A and from B to O. The output current then varies from BC to AE and down to zero, or it has an a.c. component as well as the d.c. component BC, which alone is produced when there is no modulation ($m = 0$). Similarly, the output voltage has a variable, or a.c., component of peak-to-peak value EO, as well as the d.c. component CO. It will be noted that the output voltage and current can follow the input carrier up to 100 per cent modulation. If the diode family consists of straight lines of equidistant spacing, then from the geometry of the figure it is evident that the output voltage will be a constant fraction of the envelope of the input and there will be no distortion, *i.e.*, linear detection. In particular, if $r_d = 0$, the diode curves will be vertical and e_L will be equal to the amplitude of the envelope; *i.e.*, the above-mentioned fraction will be unity. This ratio, or fraction, is known as the *detection efficiency*. If the diode family is curved and R_L is low relative to r_d, it is evident that the detection efficiency will vary over the path of operation, and hence over the modulation cycle, and also will be less than unity.

2. Input Impedance.—The output load Z_L or, more specifically, R_L (assuming C_L has negligible susceptance at the modulating frequency) presents a certain equivalent resistance to the source of e_c. The value of this resistance, call it R_c, involves r_d as well and in practice is influenced by the other impedances in the circuit, such as that of C_L (if appreciable) and of the source. However, a fair approximation to R_c may be had as follows:

The actual current wave shape through the diode is a series of narrow pulses, which represent the charging current of C_L as it charges up to the peaks of e_c through r_d. As e_c passes from its positive peak value down to its peak negative value, C_L discharges more slowly through the higher resistance R_L. The action of the diode is thus determined by two time constants, the charging time constant $C_L r_d$, and the discharge time constant $C_L R_L$. For a con-

stant amplitude carrier of peak value E_c, the output voltage e_L, the input voltage e_c of peak amplitude E_c, and the diode current i_d are as shown in Fig. 104. The wave form for e_L is exactly that for an ordinary condenser input filter. If $R_L \gg r_d$ and $1/\omega_c C_L$ is negligibly small, the ripples in e_L are negligible and e_L is practically equal to E_c. In this case, i_d consists of exceedingly narrow impulses. It is an interesting fact that such a wave form for i_d, when analyzed into a Fourier series, yields a fundamental a.c. component i_c (of frequency f_c) which is practically twice the d.c. component i_0. This is practically true regardless of the wave form of the pulses, so that we may set

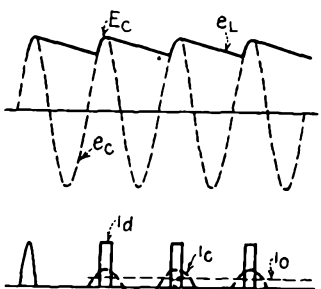


FIG. 104.—Waves shapes of current and voltage in diode detection for fixed-carrier amplitude.

$$i_c \cong 2i_0 \tag{2}$$

As the pulses become narrower and narrower, Eq. (2) approaches an equality more and more closely. This is the case as R_L exceeds r_d more and more, and such are the conditions in normal practice. If R_L does not greatly exceed r_d, then the pulses become broader, their shape is of more consequence, and i_c decreases from its value of $2i_0$. Over quite a large range of practical value of R_L and r_d, such as, for example, specific values of $\frac{1}{4}$ megohm and 5,000 ohms, respectively, Eq. (2) is practically an equation. From the above it is apparent that R_L appears to the source of e_c as one-half its actual value; *i.e.*,

$$R_c \cong \tfrac{1}{2}R_L \tag{3}$$

A further refinement is to take r_d into account as well as R_L, so that, somewhat more accurately,

$$R_c \cong \tfrac{1}{2}(R_L + r_d) \tag{4}$$

However, as stated above, r_d is usually negligible in broadcast practice as compared with R_L.

If e_c is a modulated wave, further complications arise. If the envelope drops too rapidly, C_L in conjunction with R_L may fail to follow this drop, and the inward peaks of modulation will be "clipped" and the output distorted. The rectifier is now rectifying the envelope as well as the carrier wave itself. If $f_c \gg f_m$, this clipping can be avoided by a suitably low choice of the time

constant $C_L R_L$. Figure 104 must now be replaced by Fig. 105. Actually, the pulses of i_d during inward modulation will be less than those during outward modulation because of the fact that C_L cannot discharge so rapidly through R_L as it can charge through r_d, but this factor will be ignored and the pulses represented as varying symmetrically with the envelope. This corresponds to the pulses and components of i_d of Fig. 104 being modulated by the factor $m \sin \omega_m t$. As a consequence, instead of i_0, we have an output current composed of the d.c. component i_0, acting as a (d.c.) carrier for an a.c. wave i_m of frequency f_m.

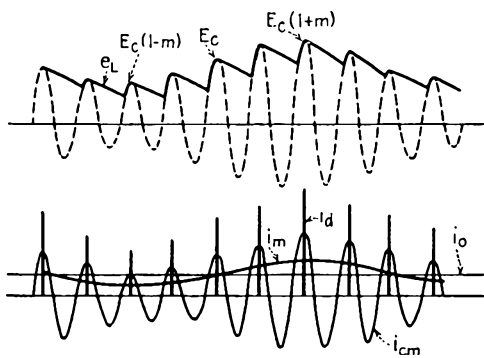


FIG. 105.—Wave shapes of current and voltage in diode detection for modulated carrier.

Similarly, the h.f. component is i_{cm}, which in turn can be broken up into a term of constant amplitude, i_c, of frequency f_c, and two side bands, i_{cm_1} of frequency $f_c + f_m$ and i_{cm_2} of frequency $f_c - f_m$.

It is evident that, like Eq. (2), the following approximate equations hold:

$$\left.\begin{aligned} i_c &\cong 2i_0 \\ i_{cm_1} + i_{cm_2} &\cong 2i_m \end{aligned}\right\} \tag{5}$$

If, tentatively, i_{cm_1} is assumed equal to i_{cm_2}, then

$$i_{cm_1} = i_{cm_2} = i_m \tag{6}$$

We now have, in addition to Eq. (4), the following:

$$R_{cm_1} = R_{cm_2} \cong R_L + r_d \tag{7}$$

where R_{cm_1} and R_{cm_2} are the apparent impedances that $R_L + r_d$ presents to side-band voltages of frequencies $f_c + f_m$ and $f_c - f_m$, respectively.

These two apparent impedances can be combined into a single equivalent impedance of $\frac{1}{2}(R_L + r_d)$ in the following manner:

In Fig. 106 is shown a vector representation of a modulated wave. The carrier voltage vector, of constant amplitude, is shown as E_c. Around it rotate in opposite directions at a frequency f_m the two side-band voltages, each $mE_c/2$ in peak amplitude. Their resultant at any instant, me_c, is clearly colinear with E_c and varying at the frequency f_m from a positive peak amplitude of mE_c to a negative peak amplitude of $-mE_c$. The equation of this voltage is therefore

$$e = me_c = (m \sin \omega_m t)E_c \sin \omega_c t \tag{8}$$

We may regard the quantity $m \sin \omega_m t$ as a modulating operator of the carrier voltage $E_c \sin \omega_c t$, and its product by E_c also represents the equation of the envelope. At any rate, to $e = me_c$, the apparent impedance of the diode load is

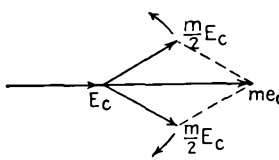


FIG. 106.—Vector relation of side bands and carrier for amplitude modulation.

$$R_{cm}' = \tfrac{1}{2}R_{cm_1} = \tfrac{1}{2}R_{cm_2} = \tfrac{1}{2}(R_L + r_d) \tag{9}$$

Now suppose that C_L does not have negligible susceptance to e_L, *i.e.*, that C_L by-passes R_L appreciably at the modulation frequency f_m. Then the a.c. component of i_L must be broken up into two quadrature components, one, i_{LC} through C_L, that leads the a.c. component of e_L by 90 deg.; the other, i_{LR} through R_L, that is in phase with e_L. The total must have been produced by a modulated carrier current capable of producing this through the diode.

Consider a modulated carrier current whose equation is

$$i = I[1 + m \sin (\omega_m t + \theta)] \sin \omega_c t \tag{10}$$

This can be expanded into the form

$$i = I \sin \omega_c t - \frac{mI}{2} \cos (\omega_c t + \omega_m t + \theta) + \frac{mI}{2} \cos (\omega_c t - \omega_m t - \theta) \tag{11}$$

The process of detecting or demodulating this current consists (among other things) in beating the side bands with the carrier. The two difference beat frequencies will be

$$\frac{i_d}{2}[\sin (\omega_m t + \theta) + \sin (\omega_m t + \theta)] = i_d \sin (\omega_m t + \theta) \tag{12}$$

The vector diagram for the current defined by Eqs. (10) and (11) is shown in Fig. 107 at the instant $t = 0$. The reference axis is along the vertical. If θ were not present, the lower and upper side bands would be, respectively, vertically up and down at this instant of time. The presence of θ causes both vectors to take up the positions shown. Thus the lower side band (l.s.b.) lags from the vertical by θ, and the upper side band (u.s.b.) leads from the vertical by θ. Their resultant nevertheless remains colinear with the carrier, so that no envelope distortion occurs. The envelope, however, regarded as a wave shape, leads its position for $\theta = 0$ by the quantity θ. This wave, upon detection, gives rise to an alternating current of frequency f_m that leads the voltage e_L by the angle θ. But it is evident from the preceding discussion that e_L is in phase with the envelope of the modulated voltage wave me_c. Hence, a modulated current i whose envelope leads that of the modulated voltage me_c by the angle θ and whose upper and lower side bands therefore lead and lag, respectively, the corresponding side bands of the voltage wave gives rise to a demodulated current i_d, which leads the demodulated voltage e_L by the same angle θ.

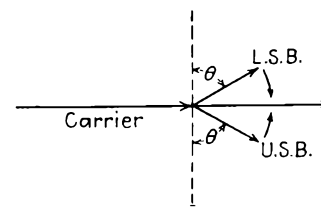


FIG. 107.—Vector relations at some instant of time for the case where the sideband currents are out of phase with the side-band voltages.

We may at least *expect* the converse to be true, *viz.*, that, if i_d leads e_L by θ, then the modulated h.f. current has upper and lower side bands that lead and lag those of the modulated h.f. voltage. At the same time, since the actual current is in the form of pulses, the sum of the side-band currents is twice i_d.

We shall now attempt to find a form of impedance which, when placed directly in the h.f. circuit, has the same effect upon the current components as Z_L acting through the diode. This effect can be expressed in the following manner: The diode translates the side bands down in the spectrum from the carrier frequency f_c to a zero frequency carrier. The lower side band would therefore have a negative frequency (be below zero in frequency). A negative frequency is physically equivalent to a positive frequency; *i.e.*, the angle of the lower side band is reversed in sign, and thus the two side bands become directly additive. In the

case of the u.s.b. and l.s.b. currents, respectively, leading and lagging the corresponding voltage side bands, the translation through detection produces the following results: The l.s.b. voltage becomes one of negative frequency and hence becomes directly additive to the upper side band to form the total voltage e_L (a.c. component) of frequency f_m. The l.s.b. current becomes one of negative frequency and lagging angle θ. Upon reversing to obtain a positive-frequency wave, θ changes sign too, to become a leading phase angle. The l.s.b. current thus becomes in phase with the u.s.b. current after detection and is therefore directly additive to the latter, to form a total current i_L (a.c. component) that leads e_L by the angle θ. This is a very striking and interesting feature of detection.

The h.f. impedance equivalent to Z_L is a parallel resonant circuit shunted by a resistance (Fig. 108*A*). Suppose it is antiresonant at f_c. Then

$$LC = \frac{1}{\omega_c^2} \tag{13}$$

The impedance at this frequency is evidently simply R, and this corresponds to $(R_L + r_d)/2$ by Eq. (7). Now consider the admittance A of the LC portion of the circuit at the frequencies $f_c \pm f_m$, that is, at the u.s.b. and l.s.b. frequencies, respectively.

$$A = \frac{1}{j(\omega_c \pm \omega_m)L} + j(\omega_c \pm \omega_m)C \tag{14}$$

Equation (14), in conjunction with Eq. (13), gives

$$A = \frac{1 - 1 \mp (2\omega_m/\omega_c) - (\omega_m^2/\omega_c^2)}{j(\omega_c \pm \omega_m)L} \approx \pm 2j\omega_m C \tag{15}$$

since if $\omega_c \gg \omega_m$, the term ω_m^2/ω_c^2 is negligibly small. The corresponding impedance is therefore

$$Z \approx \mp \frac{j}{2\omega_m C} \tag{16}$$

This is an inductive reactance to the l.s.b. voltage and a capacitive reactance of equal amount to the u.s.b. voltage. It will thus give rise to leading u.s.b. and lagging l.s.b. currents, which, upon detection in a pure resistive load, will give rise to a

current i_L that leads the voltage e_L. Hence, the impedance Z is potentially equivalent to Z_L. The equivalence is complete if[2]

$$\left.\begin{aligned} C &= C_L \\ L &= \frac{1}{\omega_c^2 C} \\ R &= \frac{R_L + r_d}{2} \end{aligned}\right\} \tag{17}$$

because the h.f. currents total to double i_L. More accurately, r_d, the diode resistance, should appear in series with Z for complete equivalence but in practical cases can usually be ignored, as well as in the third equation (for R).

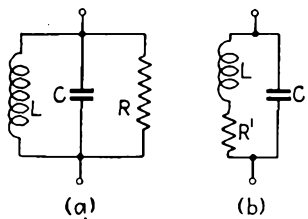


FIG. 108.—Equivalent carrier-frequency circuits: *A*, circuit giving desired phase shift to side-band currents; *B*, circuit equivalent to that in (*A*) for a limited frequency range.

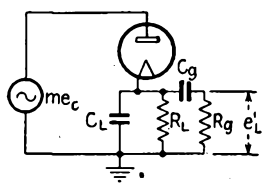


FIG. 109.—Actual diode circuit.

For $\omega_m \ll \omega_c$, Fig. 108*B* is practically equivalent to Fig. 108*A* if

$$R = \frac{\omega^2 L^2}{R'} \tag{18}$$

and so either circuit may be used to represent the apparent impedance that the diode and load present to the h.f. source.

One further type of diode load of practical interest is shown in Fig. 109. The circuit $C_g R_g$ permits the grid of the following tube to be coupled to the diode load without assuming the direct bias voltage developed across R_L. Its equivalent h.f. impedance is represented in Fig. 110. The condenser C_g appears as L' and C' in parallel, and these are in series with R', which represents R_g. At sufficiently high modulation frequencies, C_g has practically no effect, nor have, therefore, L' and C' any effect, so that R' becomes simply parallel to R, as do R_g and R_L.

3. Diode Performance—Resistive Circuit.—It is of interest to study first the diode performance when the tuned circuit is assumed to present a source impedance R to the carrier and side-band frequencies, and the load a resistance R_L to the carrier and a resistance R_L' to the modulation frequencies, where R_L' is equal to R_L and R_g in parallel (Fig. 109). It is therefore assumed that C_g has negligible reactance and C_L has negligible admittance at the modulation frequency f_m, while the tuned circuit has negligible admittance to the side-band frequencies $f_c \pm f_m$. Throughout this discussion, C_L is assumed to have negligible reactance to all h.f. components flowing through it.

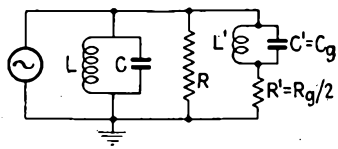


FIG. 110.—Equivalent h.f. circuit for actual diode circuit.

In the preceding section it was shown that R_L had an apparent resistance to the h.f. source of $R_L/2$. In the same manner, R has an apparent resistance, when viewed from the diode l.f. output terminals of value $2R$. The envelope appears from these same terminals as a voltage e_L, composed of a d.c. component

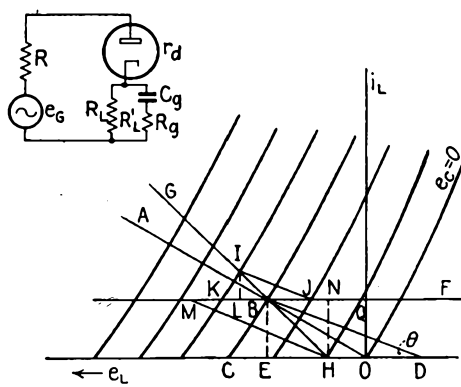


FIG. 111.—Graphical construction for actual diode circuit, showing clipping for high degrees of modulation.

equal to the carrier peak amplitude E_c, and an a.c. component equal to the vector sum of the side-band amplitudes, or

$$mE_c \sin \omega_m t.$$

The above is sufficient information for proceeding with the graphical construction.

The "successive steps" are shown in Fig. 111. First the quiescent point for the unmodulated carrier E_c is determined. Through

O, the load line OA for R_L is drawn. Then a rule is slid at an angle θ such that

$$\theta = \cot^{-1} (R_L + 2R) \tag{19}$$

Where it intersects OA in a diode rectification curve, at B, such that CD equals E_c, is the quiescent point. The load line for $R_L + 2R$ is DB. The reader can check that OD is the voltage drop in the apparent source impedance (as viewed from the diode output terminals), CE is the voltage drop in the diode itself, and EO is the voltage developed across R_L.

Now suppose that the carrier is modulated and that the instantaneous peak carrier voltage me_c is at the moment under consideration greater than E_c (envelope rising above its average carrier value), by an amount ΔE_c. To find the path of operation for the modulated carrier, proceed as follows:

Shift the voltage axis up to B, so that its new position is BF. Through B draw the steeper load line for R_L', *viz.*, GH. Now slide a rule IJ at an angle

$$\theta' = \cot^{-1} (R_L' + 2R) \tag{20}$$

to a position where its intersection with GH in I is also that of a characteristic curve IK and such that KJ equals ΔE_c. Then BJ represents the voltage drop in $2R$; KL, that in the diode; and LB, that developed across the a.c. diode load resistance R_L'. This process is then repeated until all values of the envelope have been used.

Since the path of operation is clearly along GH, it is in practice necessary to find the points corresponding only to the peaks of the envelope. It is also evident from the figure that a peak envelope which carries the path of operation beyond H will result in output distortion. We shall now discuss this in greater detail.

The resistance R_L' to the a.c. component of the output current i_L is less than the resistance R_L to the d.c. component. This would mean that, if $m = 1$, the peak modulation voltage is equal to the average, or carrier, voltage E_c; hence, the a.c. component of i_L would exceed the d.c. component, or the total would have to be negative during peak inward modulation. But this is impossible, since the diode cannot conduct reverse current. Hence, the path of operation for inward modulation is first along BH (Fig. 111) and then along the e_L axis, or the negative peaks

of i_L are "clipped," with resulting distortion. This is similar to excessive grid swing in a three- (or more) element tube, where the operation is carried beyond plate-current cutoff. There also results some self-rectification, with consequent shifting of the operating point away from B.

To avoid such clipping, the maximum modulation of the generated input signal must be less than m. To find this maximum permissible value m', for m, draw MH parallel to IJ, that is, at the slope for $2R + R_L'$. It is then evident from the figure that the generated carrier voltage is

$$E_c = CE + EO + OD = i_{dc}(r_d + R_L + 2R) \tag{21}$$

whereas the maximum permissible decrease in the generated carrier envelope is

$$m'E_c = QN + NB + BM = i_L(r_d + R_L' + 2R) \tag{22}$$

From Eqs. (21) and (22), $m' = \dfrac{i_L}{i_{dc}} \dfrac{(r_d + R_L' + 2R)}{(r_d + R_L + 2R)}$. But for maximum permissible modulation, i_L just equals i_{dc}, so that

$$m' = \frac{r_d + R_L' + 2R}{r_d + R_L + 2R} \tag{23}$$

It is to be noted from Eq. (23) that, the smaller R_L' is compared with R_L, the smaller m' will be, whereas, the greater R is, the more nearly does m' approach unity. In short, R tends to counteract the difference between R_L' and R_L in reducing the value of m'; but this, of course, does not indicate that a high value of R is desirable, since the actual ouput voltage e_L would be reduced. The main point in presenting the action of R is to show that the percentage modulation at which clipping just occurs is dependent upon all parameters, including r_d, although it is primarily caused by the lower value of R_L' as compared with R_L. If r_d is comparable with R_L', then the reflection of R into the output circuit will be less than $2R$ and quite difficult to determine. However, for quite a range of $r_d < R_L$ the above value of $2R$ holds fairly accurately; indeed, in usual broadcast practice where $r_d \ll R_L$ (diode approaches the ideal), Eq. (23) reduces to

$$m' = \frac{R_L' + 2R}{R_L + 2R} \tag{24}$$

The difference in total circuit resistance for the envelope $2R + R_L'$ and the carrier $2R + R_L$ results in a reduction in the envelope at the combined diode and load terminals as compared with the generated voltage. From Fig. 111 it is evident that the generated envelope voltage is MQ and the terminal voltage is BQ (including the drop in the diode). It is also evident that the carrier generated voltage is CD and the carrier terminal voltage is CO. The percentage modulation m_G of the generated signal is MQ/CD, while the percentage modulation m_T of the terminal voltage is BQ/CO. From the figure it is evident that

$$\left.\begin{aligned} m_G &= \frac{MQ}{CD} = \frac{i_L(2R + R_L' + r_d)}{i_{dc}(2R + R_L + r_d)} \\ m_T &= \frac{BQ}{CO} = \frac{i_L(R_L' + r_d)}{i_{dc}(R_L + r_d)} \end{aligned}\right\} \tag{25}$$

Then the percentage reduction in modulation is

$$\frac{m_T}{m_G} = \left(\frac{R_L' + r_d}{R_L + r_d}\right)\left(\frac{2R + R_L + r_d}{2R + R_L' + r_d}\right) \tag{26}$$

If a greater percentage of modulation is employed than that given by Eq. (24), it will be found that around peak inward modulation, where cutoff of i_L occurs, the envelope terminal voltage will become equal to the generated envelope voltage, whereas for other portions of the modulation cycle Eq. (26) indicates that the terminal voltage will be less. Hence, the envelope terminal voltage will be distorted in that it will have its negative half cycles peaked, whereas the output voltage e_L will have its negative peaks flattened.

From the above, the following points are evident:

1. For the usual diode circuit, where the a.c. output resistance is less than the d.c. output resistance, a value of modulation less than unity must be employed if distortion is to be avoided.

2. The source resistance tends to counteract this effect, but only in degree, and at the expense of the output signal.

3. The above two considerations also result in a lesser depth of modulation of the carrier voltage developed across the diode and load resistance as compared with that generated in the source.

The construction shown in Fig. 111 is based on the assumption that $r_d \ll R_L'$, for otherwise R would not appear as $2R$ in the

diode output circuit. However, it does take r_d into account in determining the maximum permissible percentage modulation m' and the percentage reduction in modulation. If r_d is negligible, Eqs. (24) and (26) will be found to coincide with those given by Wheeler[2] and also those given by Court.[3]

It will also be evident that, if a very large inductance whose winding resistance equals R_L is substituted for R_L, or (what is equivalent) a large inductance of negligible winding resistance be placed in series with R_L, and if, in addition, R_g is made equal to R_L, then the d.c. and a.c. resistances will be equal. In this case, 100 per cent modulation can be accommodated without "clipping." However, the cost and bulk of the inductance, as well as its susceptibility to hum pickup, mitigate against its use in practice.

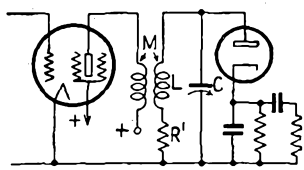


FIG. 112.—Complete diode circuit including h.f. source.

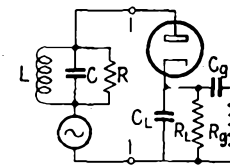


FIG. 113.—Equivalent circuit for that in the preceding figure.

It is also possible to obtain distortionless operation for 100 per cent modulation with the ordinary type of resistive load by inserting a positive direct voltage in series with the diode. Its value should be equal to HO (Fig. 111), in which case it would move the load line for R_L' over to the origin O instead of H. However, its value is dependent upon the magnitude of the carrier voltage E_c, and distortion can occur for values of E_c greater or smaller than the value that just brings the load line for R_L' over to the origin O. For a complete discussion, the reader can consult Wheeler[2] or Court.[3]

4. Diode Performance—Tuned Source Impedance.—The analysis given above is sufficient for most purposes. However, there may occur in practice a circuit such as that shown in Fig. 112. The pentode here may be regarded as a constant current generator whose current is $G_m e_{g1}$. This current induces in the tuned secondary coil, having inductance L and resistance R, a voltage $j\omega M G_m e_1$. At this point, Fig. 112 may be replaced by Fig. 113

by the use of Thévenin's theorem and also in view of the discussion in Sec. 2 and Figs. 108 and 108*A*.

As viewed from terminals 1-1, the apparent source impedance is as shown, where $R = \omega^2 L^2/R'$. The apparent generated (open-circuit) voltage is that developed across C before the diode is connected. For the carrier it is

$$e_c = je_{g1}G_m\omega_c M \frac{1/j\omega_c C}{R' + j\omega_c L + (1/j\omega_c C)} = e_{g1}\frac{G_m M}{R'C} \qquad (27)$$

if L and C are adjusted to resonance at the carrier frequency f_c. Similarly, for the upper and lower side bands, the apparent generated voltages are, respectively,

$$\left.\begin{aligned} e_c^+ &= e_{g1}^+ \frac{G_m M}{R'C} \\ e_c^- &= e_{g1}^- \frac{G_m M}{R'C} \end{aligned}\right\} \qquad (28)$$

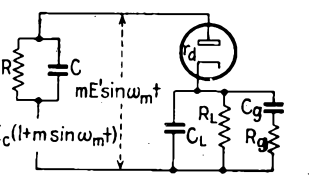


FIG. 114.—Complete diode circuit as viewed from the l.f. side.

It is evident that they are in phase with the input voltages and in the same ratio, $G_m M/R'C$. Hence, the modulated wave is undistorted by the tuned circuit, at least if $\omega_m \ll \omega_c$. These voltages and the apparent source impedance can be reflected to the output side of the diode circuit, and the circuit will appear as in Fig. 114. At the higher modulation frequencies where C is of some importance, C_g is usually unimportant, so that R_L and R_g in parallel may be replaced by R_L' as in Sec. 3.

While the quiescent point for the carrier voltage is determined as in the preceding section, the a.c. path requires the method of Sec. 10 of Chap. IV for its determination. Although fairly accurate linear-circuit calculations are possible and generally more desirable, an example will be worked out to show the graphical application.

In this example, $2R$ will be taken as 0.2 megohm, C as 200 $\mu\mu$f, C_L as 500 $\mu\mu$f, R_L as 0.125 megohm, and R_g as 0.5 megohm. The diode rectification family is that shown in Fig. 116 and is that for a 6H6 tube.

The problem as given presents formidable difficulties. The rigorous method of attack is to take C_g into account as well as the other diode load parameters. Not only does this increase the complexity of the circuit and hence the labor; but also, owing

to the large R_gC_g time constant as compared with the period of the modulation frequency to be used (5,000 c.p.s.), many cycles of operation will have to be traversed before steady-state conditions are attained.

This characteristic of the finite operator method is due to its generality and completeness of attack. A similar situation arises if it be employed to solve a choke-feed power-output stage. In the latter case, a simplification and usually a satisfactory approximation are made in assuming that there are two resistive load lines, a d.c. load line corresponding to the ohmic resistance of the choke, and an a.c. load line corresponding to the load resistance as it appears in parallel with the choke. As a first approximation, these two are assumed to intersect at the quiescent point, and this intersection is then shifted to correct for self-rectification. The finite operator method, on the other hand, would give a very narrow load spiral or scroll, which would finally, after many cycles, close into a steady-state narrow loop practically coinciding with the final position of the load line as determined by the more approximate method.

However, the phenomenon to be exhibited here is that of clipping due to the shunting effect of C_L upon R_L. Hence, we shall assume that R_g is a negligible shunt upon R_L, that is, that the a.c. and d.c. resistances are identical and equal to R_L. We shall assign a value of 0.1 megohm to this parameter. This represents the above 0.5- and 0.125-megohm resistors in parallel.

The graphical construction will proceed from the origin of the family of curves rather than from the quiescent point, and 100 per cent modulation will be assumed for e_c as determined by Eq. (27) in conjunction with Eq. (28). The equivalent voltage, as it appears across the output side of the diode circuit, will be taken as

$$e_c = 20 - 20 \cos (2\pi \times 5{,}000t) \tag{29}$$

While this is a rather low voltage, it will exhibit the clipping phenomenon described above, particularly for a modulation frequency of 5,000 c.p.s. The circuit parameters given above are entirely practical, although C_L has been chosen rather large to emphasize the effect.

Let the voltage drop across R and C be e_1 and that across the diode output load Z_L (R_L and C_L in parallel) be e_2. Also, let

i_{CL} be the instantaneous current through C_L, i_C that through C, i_R that through R, i_{RL} that through R_L, and i_T the total current. Then

$$i_T = i_{CL} + i_{RL} = i_C + i_R \tag{30}$$

$$e_1 = \frac{\Delta t}{C_L}\left(i_{CL} + \sum i_{CL}\right) = i_{RL}R_L \tag{31}$$

$$e_2 = \frac{\Delta t}{C}\left(i_C + \sum i_C\right) = i_R R \tag{32}$$

$$e_c = e_1 + e_2 + i_T r_d \tag{33}$$

where r_d is the nonlinear diode resistance as given by the family of curves.

From Eqs. (30) to (32) can be obtained

$$\left.\begin{aligned} e_1 &= i_T Z + Z\Sigma i_C \\ e_2 &= i_T Z_L + Z_L \Sigma i_{CL} \end{aligned}\right\} \tag{34}$$

where $Z = 1/(R + \Delta t/C)$ and $Z_L = 1/(R_L + \Delta t/C_L)$

Then, from Eqs. (33) and (34) is obtained

$$e_c - Z_L\Sigma i_{CL} - Z\Sigma i_C = i_T Z_L + i_T Z + i_T r_d \tag{35}$$

Equations (34) and (35) form the basis of the graphical construction. Thus, at any stage of the process, Σi_C and Σi_{CL} are known by summing up the currents from the first application of e_c, when we can assume, if we wish, no initial charges on C_L and C; that is, Σi_C and Σi_{CL} are zero. Therefore, the net voltage

$$e_n = e_c - Z_L\Sigma i_{CL} - Z\Sigma i_C$$

is known and is applied in a manner similar to that described in the preceding two sections of this chapter.

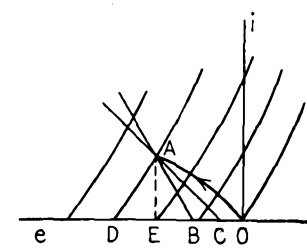


FIG. 115.—Graphical construction for a diode circuit having appreciable shunting capacitance at modulation frequencies.

In Fig. 115 is shown the construction. The net voltage at the moment under consideration is CD. The voltage $Z_L\Sigma i_{CL}$ is laid off from the origin as OB. Through B is drawn BA to represent the finite operator Z_L. Then AC, which represents the finite operator $Z + Z_L$, is slid parallel to itself until its inter-

section with AB in A is also that of a diode rectification curve AD, such that DC equals the net voltage

$$e_n = e_c - Z_L \Sigma i_{CL} - Z \Sigma i_C.$$

Then AE equals i_T, DE represents the diode voltage drop $i_T r_d$, and EO equals $i_T Z_L + Z_L \Sigma i_{CL}$ or e_2, the voltage across the diode load. In addition, it is evident that Eq. (35) is satisfied. The important point to note is that r_d is here a family of curves, and hence the proper member of the family must be selected. This has been done by drawing Z_L or AB at a distance $BO(=Z_L \Sigma i_{CL})$ from the origin O, so that point E will be distant from O by the amount e_2, the voltage across the diode load.

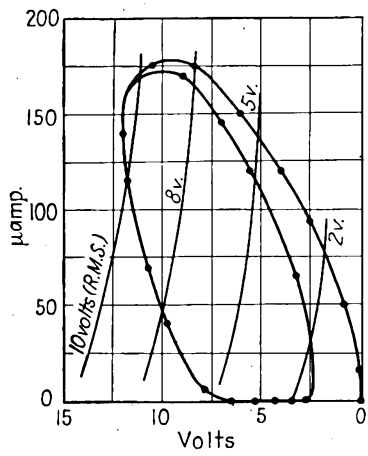


FIG. 116.—Path of operation for diode circuit for 100 per cent modulation, showing clipping on inward modulation.

From AE, the new values of i_{CL} and i_C can be found. To determine i_{CL}, for instance, divide $e_2(=EO)$ by R_L to obtain i_{RL}. Subtract this from i_T, and the difference is i_{CL}. This may turn out positive or negative, depending upon whether C_L is charging from the source or discharging through R_L during the particular time interval Δt. Similarly, i_c can be found from e_1 and R. Note that e_1 is equal to BC plus $Z \Sigma i_C$, or alternatively to $i_T Z + Z \Sigma i_C$.

The successive values of i_T, Σi_{CL}, Σi_C, i_{CL}, i_C, e_1, e_2, e_n, etc., can be arranged in the form of a tabular schedule, in order to systematize the work. Where possible, slide-rule calculations should be made to shorten the work, as in the calculations for i_{CL} and e_n. Note that these could also be obtained graphically, but, as pointed out in the dynatron example of Chap. II, the graphical process is best reserved for such work as must be done on the nonlinear element itself, here r_d.

In the example cited, 20-deg. intervals in the cycle were chosen, so that $\Delta t = \frac{1}{18} \times 1/5{,}000 = 1/90{,}000$ sec. Then $Z_L = 18{,}330$ units, and $Z = 43{,}400$ units. From Eq. (29) it is

evident that, at the start ($t = 0$), $e_c = 0$ volts. In addition, zero initial charges are assumed for the two condensers. During the first time interval, e_c is assumed to have the constant value of 1.2 volts, which it actually attains at the end of the time interval. Since $\Sigma i_{CL} = \Sigma i_C = 0$, e_n is therefore 1.2 volts too. The first position of Z_L is also evidently through the origin (Fig. 116), and the intersection with Z and a diode curve is such that $i_T = 15\mu a$, $e_2 = 0.3$ volt, and $e_1 = 0.7$ volt. Then $i_{RL} = 3\mu a$, $i_{CL} = 15 - 3 = 12\mu a$, and $Z_L \Sigma i_{CL} = 0.22$ volt. Similarly, $i_R = 3.5\mu a$, $i_C = 11.5\ \mu a$, and $Z\Sigma i_C = 0.5$ volt. Therefore, $Z_L \Sigma i_{CL} + Z\Sigma i_{CL} = 0.72$ volt.

During the next time interval Δt, e_c rises to 4.7 volts. Therefore, $e_n = 4.7 - 0.72 = 4.0$ volts. The operator Z_L is now drawn through 0.22 volt. Its intersection with Z and the proper diode curve is at a value of $i_T = 50\mu a$, $e_2 = 1.0$ volt, and

$$e_1 = 0.5 + 2.3 = 2.8 \text{ volts.}$$

The several computations are then made as in the first step, and the process repeated. For the initial conditions chosen, the load loop does not close until about $1\frac{1}{2}$ cycles have been computed. It will be observed that the loop is flattened at the bottom, since i_T cannot become negative. Over this flat part, it will be found that e_n is negative, so that Z and Z_L intersect on the voltage axis, and i_T is therefore zero. The diode curve that intersects with the above two operators is the appropriate one that meets the axis e_n units to the right of the intersection and then, of course, proceeds along the axis to the left, as in the case of class AB balanced-amplifier constructions. The flat portion of the loop represents a period of time during which the condensers discharge through their respective resistances, thus diminishing Σi_{CL} and Σi_C (hence $Z_L \Sigma i_{CL}$ and $Z\Sigma i_C$) until their sum is less than e_C, whereupon e_n becomes positive once more, i_T becomes greater than zero, and the loop rises once again. The steady-state condition is given by the narrower loop in the figure.

The above circuit, however, is amenable to fairly simple and accurate analytical treatment. Thus, the a.c. component of i_L, or i_m, is given by

$$i_m = \frac{e_m}{Z_L + 2Z} \tag{36}$$

where e_m is the voltage corresponding to the modulation envelope, $Z_L = R_T'/(1 + j\omega C_L R_L)$, and $Z = R/(1 + j\omega CR)$, where R_L' is R_L and R_g in parallel.

The d.c. component of i_L, or i_C, is due to a voltage corresponding to the carrier amplitude, e_c, and is given by

$$i_c = \frac{e_c}{R_L + 2R} \tag{37}$$

The percentage modulation of the carrier and also of the d.c. output current is

$$m_C = \frac{i_m}{i_C} = \frac{e_m}{e_c}\frac{R_L + 2R}{Z_L + 2Z} = m_G \frac{R_L + 2R}{Z_L + 2Z} \tag{38}$$

where m_G is the percentage modulation of the carrier voltage e_c. But m_C cannot exceed unity, since i_m cannot exceed, at its negative peak value, i_C, since then the diode current would be reversed, which is impossible. Hence, the maximum value of m_G that corresponds to $m_C = 1$ is, by Eq. (38),

$$m_G = \frac{Z_L + 2Z}{R_L + 2R} \tag{39}$$

The terminal voltages (envelope and carrier) across the diode and Z_L are related to the generated voltages e_m and e_c, respectively, by

$$\left.\begin{aligned} e_{Tm} &= e_m \frac{Z_L}{Z_L + 2Z} \\ e_{Tc} &= e_c \frac{R_L}{R_L + 2R} \end{aligned}\right\} \tag{40}$$

The percentage modulation of the terminal carrier voltage is

$$m_T = \frac{e_{Tm}}{e_{Tc}} = \left(\frac{e_m}{e_c}\right)\left(\frac{Z_L}{R_L}\right)\left(\frac{R_L + 2R}{Z_L + 2Z}\right) = m_G\left(\frac{Z_L}{R_L}\right)\left(\frac{R_L + 2R}{Z_L + 2Z}\right) \tag{41}$$

The maximum value of m_T is also unity, since otherwise the instantaneous voltage would reverse across the diode, whereupon the latter would short out the signal voltage and thus prevent the latter from having any appreciable reverse value. Hence, the maximum value of m_G for $m_T = 1$ is

$$m_G' = \left(\frac{R_L}{Z_L}\right)\left(\frac{Z_L + 2Z}{R_L + 2R}\right) = \frac{R_L}{Z_L} m_G \tag{42}$$

In short, the permissible value of m_G here is greater than that given by Eq. (40), so that, for the circuit postulated, clipping due to current cutoff is the limiting factor. The above derivation assumes r_d is negligibly small, whereas the graphical construction does not.

For the example given, m_G comes out to be 59.3 per cent. If R_L is taken as 0.125 megohm, m_G is 54.8 per cent, so that no great error was made in assuming that the a.c. and d.c. load lines are identical. At low modulation frequencies the construction of Fig. 111 is permissible, since C_L and C have negligible susceptance. The maximum permissible modulation is then given by Eq. (24) and for this example comes out to be 92.3 per cent.

In Fig. 117 has been plotted the load loop for a 20-volt generated carrier wave modulated 59.3 per cent by a 5,000-cycle frequency. The equation of the envelope is therefore

$$e_c = 20 - 11.9 \cos (2\pi \times 5{,}000t) \tag{43}$$

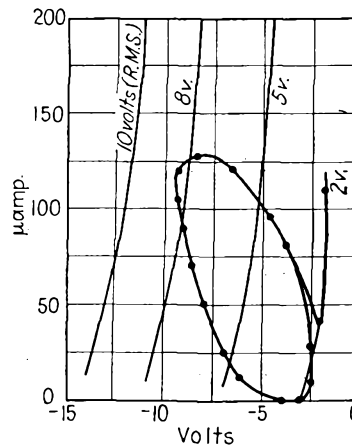


FIG. 117.—Path of operation for a diode circuit with maximum permissible percentage modulation at which no clipping occurs.

It will be observed that the loop is just perceptibly flattened; it is considered not only that this is in good agreement with Eq. (40), but also that Eq. (40) is definitely a maximum value for the permissible modulation, since clipping just occurs for this value. In passing, it is well to note that rarely if ever is such high modulation encountered in practice at such a high modulation frequency. The question may still arise as to what happens when a strong l.f. and weak h.f. complex modulation wave is impressed. This can be answered by the graphical method, but, unfortunately, each wave shape requires individual treatment. This question in general has never been satisfactorily answered for a large number of nonlinear circuit problems.

5. Transrectification Diagrams.—Another form of detector, used more in the past than at present, is the plate-circuit detector. Usually, a tuned circuit feeds the grid, which is biased close to

or at plate-current cutoff of a triode or pentode tube. Detection therefore takes place in the plate circuit; and if a large input grid signal is employed, the detection is approximately linear—sometimes called "power detection" (although this may also be had with the old grid leak detector).

The plate load is a resistor, and the plate is by-passed to ground (or the cathode) by a condenser sufficiently large to short-circuit the h.f. currents developed in the plate circuit, but not (too) appreciably the higher modulation frequencies. The plate may be coupled to the following grid by means of a grid coupling condenser and resistor.

The plate load therefore corresponds to that of the diode, *viz.*, Z_L; indeed, the tube functions like a diode fed from a resistive source, the R_p of the tube. It is therefore possible to develop experimentally a family of transrectification curves similar to that for a diode. Such a family is shown in Fig. 118. The load resistor may be replaced by a zero-impedance d.c. source of active voltage. The parameter is carrier voltage as in the case of a diode, but here the carrier voltage is applied to the grid, rather than directly in the plate circuit. The family depends upon the value of grid bias chosen. The curve at the extreme right, for zero carrier voltage, is the ordinary static characteristic of the tube for the bias voltage. Progressively higher carrier voltages produce successive curves to the left. Ordinarily, carrier voltages exceeding in peak value the bias voltage are not used or shown, since the grid would draw current and produce distortion.

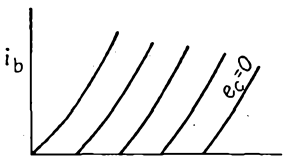


FIG. 118.—Transrectification curves.

The graphical constructions are exactly the same as for the diode. Thus, if the plate load resistance is lower to the envelope voltage than to the carrier, clipping occurs for modulation percentages approaching 100 per cent. Modifications in the graphical constructions for application to the triode must be made as indicated in Chap. II (Sec. 13).

From the foregoing it is evident that the plate-circuit detector is no better than the diode as regards clipping and consequent distortion for high degrees of inward modulation. The two advantages of plate-circuit detection are

1. That it does not load the tuned circuit, and thus reduce the voltage developed at that point, as well as flatten the resonance curve.

2. That the tube serves as an amplifier of the input signal, as well as a detector.

However, adequate amplification is easily obtained in the preceding tuned stages in the case of the diode, and the plate-circuit detector may overload owing to grid current if the signal input is unduly large, as on outward modulation exceeding 100 per cent. This distortion is in addition to that produced in the plate circuit on inward modulation in the form of clipping, which also occurs in diode detection. Possibly the main reason for the diode superseding the plate-circuit detector is the fact that the load resistor is at a positive potential to ground and hence not in an effective position to furnish automatic volume control to the r.f. and i.f. stages.

BIBLIOGRAPHY

1. LAPORT, E. A.: Characteristics of Amplitude Modulated Waves, *RCA Rev.*, April, 1937.
2. WHEELER, H. A.: Design Formulas for Diode Detectors, *Proc. I.R.E.*, June, 1938.
3. COURT, W. P. N.: Diode Operating Conditions, *Wireless Engineer*, November, 1939.
4. KILGOUR, C. E., and J. M. GLESSNER: Diode Detection Analysis, *Proc. I.R.E.*, **21**, 930, July, 1933.
5. TERMAN, F. E., and N. R. MORGAN: Some Properties of Grid Leak Power Detection, *Proc. I.R.E.*, December, 1930.

CHAPTER VII

MISCELLANEOUS GRAPHICAL CONSTRUCTIONS

1. Feedback Constructions—Voltage Feedback.—An interesting graphical construction has been developed for a tube to which feedback has been applied.[1,2] A typical circuit is shown in Fig. 119. Here, the actual voltage e_c applied between grid and cathode is composed of the signal voltage e_g and the feedback voltage e_f, which is in phase opposition with e_g. (The direct bias voltage E_c is not considered.) If the condenser C be sufficiently large, then its reactance compared with R_1 and R_2 over the range of frequencies under consideration is negligibly small, and

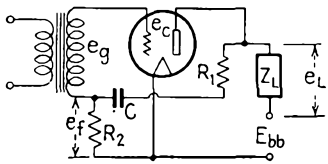


FIG. 119.—Typical voltage feedback circuit.

$$e_f = \frac{R_2}{R_1 + R_2} e_L = ne_L \tag{1}$$

where e_L is the output voltage due to e_g impressed and

$$n = \frac{R_2}{(R_1 + R_2)}$$

is the percentage feedback. Since the latter is a real number (ratio of resistances), it represents a voltage in phase with e_L. From the figure it is evident that

$$\begin{aligned} e_c &= e_g - e_f = e_g - ne_L \\ &= e_g - \frac{\mu n e_g Z_L}{R_p + Z_L} \end{aligned} \tag{2}$$

Solving Eq. (2) for e_c, we obtain

$$e_c = e_g \frac{R_p + Z_L}{R_p + Z_L(1 + \mu n)} \tag{3}$$

Then e_L is found to be, from Eq. (3) by reference to Eq. (2),

$$\begin{aligned} e_L &= e_g \frac{\mu Z_L}{R_p + Z_L(1 + \mu n)} \\ &= e_g \left[\frac{\mu}{1 + \mu n}\right] \left[\frac{Z_L}{R_p/(1 + \mu n) + Z_L}\right] \end{aligned} \tag{4}$$

Equation (4) indicates that the tube may be replaced (as far as the external terminals are concerned) by an equivalent tube whose amplification factor is $\mu/(1 + \mu n)$ and whose internal resistance is

$$\frac{R_p}{(1 + \mu n)}.$$

If condenser C were omitted, then feedback would occur for the d.c. as well as for the a.c. components; but, with C present, only the a.c. components are affected by the feedback. Nevertheless, the following graphical construction, based on the absence

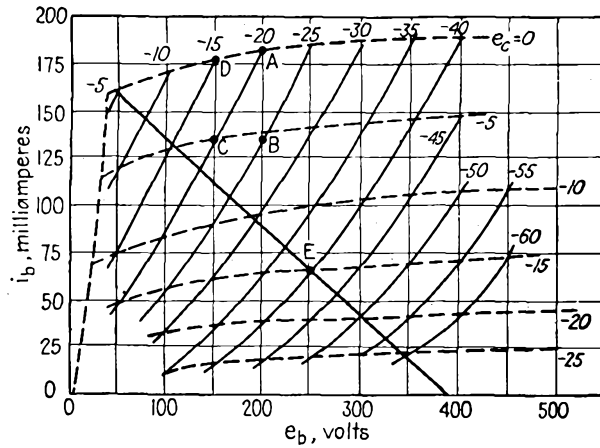


FIG. 120.—Modification in plate characteristics of a tube, due to voltage feedback.

of C, can be used with simple modifications to determine the behavior of the stage when C is present.

Equation (4) indicates that the feedback may be regarded as modifying the i_b-e_b family of curves, rather than Z_L. Hence, the graphical construction can be used to modify the original tube family as follows:

In Fig. 120 is shown a beam-power-tube family of characteristics. Suppose the percentage feedback n equals 10 per cent. Then, for any plate voltage e_b, e_f equals $0.1e_b$, and this must be added to any chosen value of e_c to give the applied input voltage e_g. Thus, suppose we take the 200-volt ordinate. This represents a feedback voltage of -20 volts. The intersection at A with the zero-volt grid curve therefore represents a value of e_g equal to -20 volts; its intersection with the -5-volt grid

curve at B represents a value of e_g equal to -25 volts; etc. Next, consider the 150-volt ordinate. Its intersection with the -5-volt curve at C represents a value of e_g equal to -20 volts; its intersection with the zero-volt curve at D represents $e_g = -15$ volts, etc. Points of the same value of e_g may now be joined by a curve, as shown by each of the heavy lines.

The curves represent the characteristics of the equivalent tube. They are steeper and more closely spaced, hence (as explained in Chap. I) represent a tube of lower R_p and μ, respectively. This checks with Eq. (4). It must be remembered that this family varies with the value of n, just as in Chap. III it was shown that the push-pull family of curves varied with the quiescent point.

This family may be used to determine the performance of a push-pull or single-ended stage and represents an interesting graphical combination of methods. Thus, suppose the tube is used single-ended in conjunction with a 2,220-ohm load resistor. Assume further that condenser C (Fig. 119) is present and that the bias is -15 volts and $E_{bb} = 250$ volts. The latter values, being d.c., are laid off on the original tube family, as shown at point E.

For a.c. grid swings, the feedback i_b-e_b characteristics are to be used. The maximum positive grid swing is evidently

$$40 - 5 = 35 \text{ volts},$$

whereas without feedback it would be 15 volts. The ratio is $\frac{35}{15} = 2.33$. This ratio may be formulated analytically as follows:

The gain without feedback is evidently

$$a = \frac{\mu Z_L}{R_p + Z_L} \tag{5}$$

From Eq. (4), the gain with feedback is found to be

$$\frac{e_L}{e_g} = a_f = \frac{\mu Z_L}{R_p + Z_L(1 + \mu n)} \tag{6}$$

From Eqs. (5) and (6), we obtain

$$\frac{a_f}{a} = \frac{1}{1 + na} \tag{7}$$

or the gain is reduced by the factor $1/(1 + na)$. In this example, this quantity should come out to be 1/2.33, or $1 + na$ should be 2.33.

Referring to Fig. 120, we note that

$$a = \frac{250 - 50}{15} = 13.3$$

The positive half cycle has been taken, and distortion in the output ignored. Then $1 + na = 1 + (0.1 \times 13.3) = 2.33$, which checks the ratio of relative grid swings required for the same output voltage e_L.

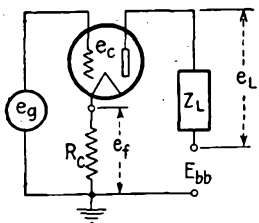


FIG. 121.—Typical current feed-back circuit.

2. Feedback Constructions—Current Feedback.—Another type of feedback circuit is shown in Fig. 121. Here feedback occurs in the un-by-passed cathode resistor R_c and depends upon the plate current. The feedback voltage E_f is therefore proportional to the current through the load impedance Z_L, rather than the voltage e_L developed across it.

The analytical formulation for this case is

$$e_c = e_g - i_p R_c = e_g - e_c \frac{\mu R_c}{R_p + R_c + Z_L} \tag{8}$$

from which

$$e_c = e_g \frac{R_p + R_c + Z_L}{R_p + R_c(1 + \mu) + Z_L} \tag{9}$$

Since

$$e_L = \mu e_c \frac{Z_L}{R_p + R_c + Z_L} \tag{10}$$

we obtain finally

$$e_L = \mu \frac{Z_L}{R_p + R_c(1 + \mu) + Z_L} \tag{11}$$

Equation (11) indicates that the tube is equivalent to a tube having the same μ, but an internal plate resistance equal to that of the former, plus an amount $R_c(1 + \mu)$. Hence a graphical

construction will be sought which converts the i_b-e_b characteristics of the actual tube into that of the equivalent one.

In Fig. 122 are shown the characteristics of a triode tube (solid lines). Suppose this has an un-by-passed self-bias resistor R_c of 2,100 ohms in series with the cathode. When the plate current is 4 ma., the drop across R_c is 8.4 volts. If the input voltage e_g (Fig. 121) is 6 volts, then the actual voltage between grid and cathode, or e_c, will be -2.4 volts. For this bias, a plate voltage (to cathode) of 95 volts is required to maintain the plate current. The voltage between the plate and ground, or the total, is therefore 95 + 8.4, or 103.4, volts. The values for the

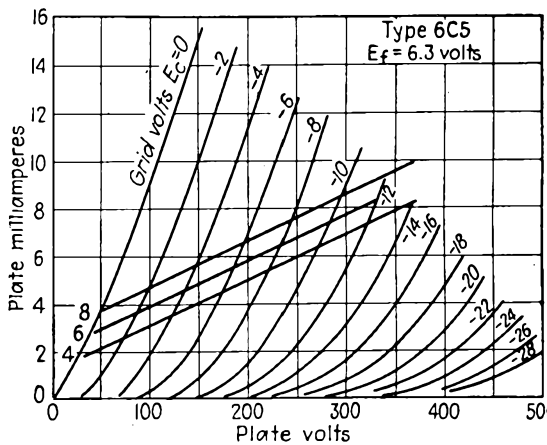


FIG. 122.—Modification in plate characteristics of a tube, due to current feedback.

equivalent tube are $e_c' = 6$ volts, $e_b' = 103.4$ volts, and $i_b = 4$ ma. The point (4, 103.4) can be laid off on the actual tube's characteristics, and this, in conjunction with other such points corresponding to the value of $e_c' = 6$ volts, will form one curve of the equivalent tube's characteristics, *viz.*, the one for which the grid voltage is 6 volts. For convenience, these are exhibited in the accompanying table. The values from the first and last columns are plotted on Fig. 122 and give rise to the +6-volt equivalent-tube characteristic (heavy line).

Next, a value of +4.0 volts can be taken for e_c' and a similar table prepared and plotted. For example, at $i_b = 4$ ma., $e_f = 8.4$ volts, $e_c = -4.4$ volts, $e_b = 140$ volts, and $e_b' = 148.4$ volts.

In Fig. 122 three such characteristics have been plotted, *viz.*, the 4-, 6-, and 8-volt characteristics for the equivalent tube. It

CALCULATIONS FOR EQUIVALENT TUBE

i_b, ma.	e_f	e_c'	e_c	e_b	e_b'
3	6.3	+6.0	− 0.3	44	50.3
4	8.4	+6.0	− 2.4	95	103.4
5	10.5	+6.0	− 4.5	140	150.5
6	12.6	+6.0	− 6.6	195	207.6
7	14.7	+6.0	− 8.7	250	264.7
8	16.8	+6.0	−10.8	310	316.8
10	21.0	+6.0	−15.0	400	421

will be evident that the equivalent R_p is much higher, as predicted by Eq. (11). It will also be noted that the curves are more nearly linear and hence that the distortion, for a constant impedance load, will be less. On the other hand, a variable impedance load, such as a loud-speaker, will not be damped so effectively as by the same tube using voltage feedback and may give rise to an objectionable transient response.

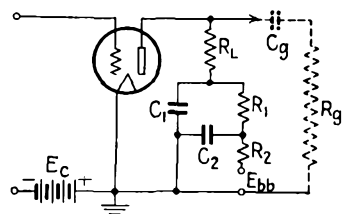


FIG. 123.—Two-stage plate-isolation circuit.

The graphical constructions for resistance coupling, for instance, are the same as in the preceding case. In this example, the feedback is present for the d.c. as well as the a.c. components, but the load line is drawn on the characteristics just the same as in the examples in Chap. I.

3. Plate Isolation.—In audio and video voltage amplifier stages, it is usually necessary to employ plate isolation or decoupling circuits to obviate undesirable feedback through the internal impedance of the common B power supply. Such a circuit is shown in Fig. 123, in which R_1 and R_2, C_1 and C_2 represent the resistors and condensers, respectively, of a two-section plate isolation filter. Except at the very low end of the frequency range, C_1 and C_2 can be regarded as having negligible reactance compared with R_L, R_1, and R_2. Hence, the impedance to the a.c. component i_p of the plate current i_b is practically R_L, while the impedance to the d.c. component I_b is $R_L + R_1 + R_2$. It is desired to analyze this circuit graphically, in order to determine optimum values for the circuit parameters.

The problem can be formulated in several ways, but possibly the following is the most significant: Suppose a peak-to-peak output voltage of value $2e_L$ is desired and that the B supply voltage is E_{bb}. (Clearly, E_{bb} must exceed $2e_L$; and, owing to the drop in $R_1 + R_2$, it must exceed it by a considerable amount. In order that the stage have good gain, R_L should be at least $3R_p$. The procedure is then as follows:

In Fig. 124 is shown a set of tube characteristics. The above assumed value of R_L gives the load line to the a.c. component i_p. This load line is slid along the curves in such manner that its left-hand tip at all times touches the $e_c = 0$ curve. Where its projection on the e_b axis attains the value of $2e_L$ gives the proper location for it on the tube characteristics. A position somewhat

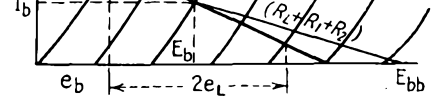


FIG. 124.—Graphical construction for plate-isolation circuit.

higher may be desired in order to reduce the distortion components. If the tube is a triode, these may be evaluated as second-harmonic distortion by Eq. (29), Sec. 17 of Chap. III. In this way, the peak negative value of the grid voltage may be ascertained, e_{c_1}.

The bias is evidently half this voltage, as shown. Point A is then evidently the value of the d.c. component I_b. From A a line is drawn to E_{bb}. This is the load line for the d.c. component and represents a value of resistance equal to $R_L + R_1 + R_2$. Since R_L has been assumed, $R_1 + R_2$ can then be found. Then C_1 and C_2 can be selected so as to give the proper time constants in conjunction with R_1 and R_2 for adequate plate isolation. (This is a matter of ordinary linear circuit analysis.)

If the results are not entirely satisfactory, another value of R_L can be assumed and the process repeated. The object generally is to obtain a maximum value for $R_1 + R_2$ consistent with a minimum of distortion content in $2e_L$. Usually, in small-voltage amplifier tubes, the value of I_b and E_b and hence the plate dissipation are of no particular concern.

If a certain value of $R_1 + R_2$ is desired, then, after choosing R_L, the load line for $R_L + R_1 + R_2$ can be laid off from point A and its intersection with the e_b axis, or E_{bb}, thus found. Thus the solution may be for the requisite value of B supply voltage for a given amount of $R_1 + R_2$ in conjunction with an assumed R_L.

In low-level stages the tube is usually oversize; *i.e.*, the grid swing is but a small fraction of the maximum possible. This, in turn, of course means that $2e_L$ is also very small. In such a case, point A can be close to the e_b axis (I_b can be small) and also close to the i_b axis (E_b can be small). Usually, the distortion products need not be considered, since for small grid swings the amplification is practically linear.

From the above it is evident that the load line for R_L will be low on the tube characteristics and also that either $R_1 + R_2$ can be much higher for a given E_{bb} or E_{bb} can be much lower for a given $R_1 + R_2$ plus assumed value of R_L. This is important because generally the low-level stages require the most filtering, and the higher the values of R_1 and R_2 the smaller C_1 and C_2 for the required amount of filtering.

It should be obvious that the above applies equally well to a one-section plate isolation filter. As far as the graphical constructions are concerned, the result obtained is the total amount of isolation resistance possible, regardless of into how many sections it is divided.

4. Effect of Low Grid Coupling Resistance.—The preceding construction may be used for another problem that may arise in practice. Suppose that the conventional method of coupling to the next stage is employed, *viz.*, C_g and R_g, as shown in dashed lines in Fig. 123. If the tube shown is a pentode, then R_L is usually of high value, -0.1 megohm or more. If the following tube is a power tube, then R_g may have to be of the same order of magnitude as R_L owing to grid-current considerations. Then, if we assume that the reactance of C_g is negligible in the range of frequencies under consideration, the impedance to the a.c. component is R_g and R_L in parallel, a value materially less than R_L, while the resistance to the d.c. component is at least R_L (if R_1 and R_2 are not employed). Thus the a.c. output is less than would be obtained if the load line for the two components were substantially the same ($R_g \gg R_L$).

The construction is the same as in Fig. 124, except that, in the case of a triode, R_g is known and R_L is to be determined so that from point A it passes through E_{bb}. It is not evident whether this is always possible; in any event, it can apparently be determined only by a series of trials. However, a solution can always be found by assuming a value for R_L, then drawing the load line for R_g and R_L in parallel, and then drawing from point A (Fig. 124) the load line for R_L alone, thus determining E_{bb}.

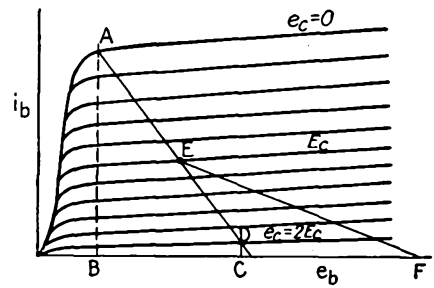


FIG. 125.—Graphical constructions for the case of lower a.c. than d.c. plate-load resistance—pentode tube.

This permits an even more unusual solution in the case of a pentode. In Fig. 125 are shown the i_b-e_b characteristics for such a tube. Operation to the left of the knee (to the left of AB) is to be avoided. Distance BC represents $2e_L$. Therefore AD represents the load line for R_g and R_L in parallel such that the distortion products are not excessive and $e_c = 2E_c$ is the maximum negative grid swing. Half this value is the value of bias E_c to be used (point E). Since R_g is predetermined and known, R_L can be found from load line AD. The load line for R_L is drawn through E, and where it intersects the e_b axis in F gives the value of B supply voltage required. This construction can obviously be further modified to take into account the presence of plate isolation resistors.

BIBLIOGRAPHY

1. SCHADE, O.: Beam Power Tubes, *Proc. I.R.E.*, March, 1938.
2. MARTIN, LOUIS: Characteristics of Inverse Feedback Circuits, *Radio Eng.*, May, 1937.

INDEX