ELECTRONIC CIRCUITS

MARTIN

PRENTICE HALL

ELECTRONIC CIRCUITS

THOMAS L. MARTIN, JR.

PRENTICE-HALL ELECTRICAL ENGINEERING SERIES

ELECTRONIC
CIRCUITS

PRENTICE-HALL ELECTRICAL ENGINEERING SERIES

W. L. EVERITT, Ph.D., *Editor*

ANNER *Elements of Television Systems*
BENEDICT *Introduction to Industrial Electronics*
DAVIS AND WEED *Industrial Electronic Engineering*
FICH *Transient Analysis in Electrical Engineering*
GOLDMAN *Information Theory*
GOLDMAN *Transformation Calculus and Electrical Transients*
JORDAN *Electromagnetic Waves and Radiating Systems*
MARTIN *Electronic Circuits*
MARTIN *Ultrahigh Frequency Engineering*
MOSKOWITZ AND RACKER *Pulse Techniques*
PUMPHREY *Electrical Engineering*, 2d ed.
PUMPHREY *Fundamentals of Electrical Engineering*
RIDEOUT *Active Networks*
RYDER *Electronic Engineering Principles*, 2d ed.
RYDER *Electronic Fundamentals and Applications*
RYDER, *Networks, Lines, and Fields*, 2d ed.
SHEDD *Fundamentals of Electromagnetic Waves*
SKRODER AND HELM *Circuit Analysis by Laboratory Methods*
STOUT *Basic Electrical Measurements*
THOMSON *Laplace Transformation*
VAIL *Circuits in Electrical Engineering*
VAN DER ZIEL *Noise*
VON TERSCH AND SWAGO *Recurrent Electrical Transients*
WARD *Introduction to Electrical Engineering*, 2d ed.

ELECTRONIC CIRCUITS

By

THOMAS L. MARTIN, Jr.

Professor of Electrical Engineering
and Head of Department
University of Arizona

Englewood Cliffs

PRENTICE-HALL, INC.

First printing . . . . April, 1955
Second printing . . . . . October, 1955

PRINTED IN THE UNITED STATES OF AMERICA
25033

"Then said a Teacher, Speak to us of Teaching.

And he said:

No man can reveal to you aught but that which already lies half asleep in the dawning of your knowledge.

The teacher who walks in the shadow of the temple, among his followers, gives not of his wisdom but rather of his faith and his lovingness.

If he is indeed wise he does not bid you enter the house of his wisdom, but rather leads you to the threshold of your own mind.

The astronomer may speak to you of his understanding of space, but he cannot give you his understanding.

The musician may sing to you of the rhythm which is in all space, but he cannot give you the ear which arrests the rhythm nor the voice that echoes it.

And he who is versed in the science of numbers can tell of the regions of weight and measure, but he cannot conduct you thither.

For the vision of one man lends not its wings to another man.

And even as each of you stands alone in God's knowledge, so must each one of you be alone in his knowledge of God and in his understanding of the earth."

PREFACE

"Electronics is the science and technology which deals primarily with the supplementing of man's senses and his brain power by devices which collect and process information, transmit it to the point needed, and there either control machines or present the processed information to human beings for their direct use."[1]

This is an extremely broad definition and encompasses an enormous body of knowledge, so enormous that it is frequently subdivided into three superficially distinct areas of inquiry, as follows:

1 *electronic components*
2 *electronic circuits*
3 *electronic systems*

It is unrealistic and probably naive to assume that *electronics* can be separated into three minor areas that can be examined independently of one another. Systems, circuits, and components are very closely tied together. System requirements often lead to the development of new components and circuits. A new component may finally result in the practical realization of a new system, and this could stimulate the development of new circuits. So you can see that the separation of *electronics* into three minor areas of interest is an academic fiction. In spite of the artificiality of the division, it seems to be the only orderly approach to arranging the huge amounts of data facing us.

Electronic components as used here will signify devices such as vacuum and gas tubes, thermistors, varistors, transistors, and magnetic and dielectric amplifiers and storage elements. The term *electronic circuit* will refer to the connection of electronic components together with the ordinary elements of resistance, capacitance, inductance, and power sources into complete circuits.

This book presents the fundamental principles and techniques associated with electronic circuits without emphasizing particular components or system applications. If vacuum tube circuits seem to receive the greatest attention, it is only because the analysis of such circuits is more highly developed at the time of writing.

[1] Reprinted with permission from "Let Us Re-Define Electronics," by W. L. Everitt, *Proc. IRE*, vol. 40, no. 8, August, 1952, p. 899.

If I were asked to summarize this book as briefly as possible, the result would read as follows:

> Electronic components can be represented by simple equivalent circuits. Electronic circuit design is thereby reduced to *ordinary* circuit design and is no longer a problem in electronics.

When the idea is flatly stated like this, I am appalled at the size of the book. However, simple ideas are usually the most difficult to execute and explain. Many illustrations under varying conditions are required before the generality and simplicity are understood.

The book is subdivided into three parts, as follows:

Part I: *Introduction*

Part II: *Class A Circuits*

Part III: *Operation in the Switching Mode*

Part I is a brief introduction to the principles of equivalent circuits and the elements of electric circuit theory based on the complex frequency and Laplace transform approach.

Parts II and III are the main sections of the book. Most of the usual and some unusual circuits using nearly all the various components are presented. All circuits covered in Part II require continuous operation of the electronic component. This is called *Class A operation* in this book. Nearly all the circuits in Part III require discontinuous operation of the electronic component. This is defined in this book as *operation in the switching mode.*

The approach is almost entirely analytical. Although many useful methods from advanced mathematics are avoided, the book is still unashamedly mathematical in nature. I feel no compulsion to apologize for this. In the light of modern developments it would be surprising to handle the subject otherwise.

Many useful and informative results, formulas, and design charts are obtained, but the emphasis is on the techniques used rather than the results themselves. No attempt has been made to write a handbook of formulas or to compile an encyclopedia of illustrative numerical problems. The book presents the methods of formulating circuits to obtain useful design formulas and performance criteria.

The book should appeal to a diversified group of readers. Practicing engineers and physicists will find it to be a usable reference in their everyday work with circuit design and development. It can also be used as a textbook at the graduate or undergraduate level covering a two- or three-semester sequence of courses.

The reader should have a background that includes elementary calculus, an introductory course in electronics, and a previous or concurrent course in a-c circuit theory will prove immensely helpful.

Thomas L. Martin, Jr.

Tucson, Arizona

ACKNOWLEDGEMENTS

The preparation of this book has left me indebted to many people. Mr. John F. Davis, formerly of the Electrical Engineering Department of the University of New Mexico, provided the enthusiastic encouragement, friendly criticism, and discussion that every text book author needs and appreciates. Dr. R. A. Hessemer, also of the University of New Mexico, contributed many helpful suggestions.

A large part of the manuscript was reviewed by Professor Clayton M. Zieman, Department of Electrical Engineering, U. S. Air Force Institute of Technology, and his assistance is deeply appreciated. Mr. J. C. Connell, of the Sandia Corporation, carefully proofread the chapters on amplifiers and multivibrators, diligently determining errors, inconsistencies, and poor phraseology. The often heated, but always friendly, arguments with Mr. S. Stanley Locus, also of Sandia Corporation, were of considerable help in clarifying obscure points.

I had the good fortune to take some courses under Professor J. M. Pettit, of the Electrical Engineering Department of Stanford University. His interesting and scholarly presentation undoubtedly sparked many of the ideas in this text.

Professor Ralph W. Tapy, Head of the Electrical Engineering Department at the University of New Mexico, gave me a tremendous boost by arranging my teaching schedule to parallel my work on the manuscript.

The many students at the University of New Mexico, graduate and undergraduate on the campus and at Los Alamos Scientific Laboratory, deserve recognition and thanks for their cooperation.

It is particularly pleasant to thank the representatives of Prentice-Hall, Karl Lorentzen, R. Frank Bitner, and Dean W. L. Everitt of the University of Illinois. They were all extremely helpful and encouraging.

Finally, I want to acknowledge the understanding and consideration given me by my wife and children. The book was written for them and it is nice to know they understand.

CONTENTS

Part I: Introduction

Part II: Class A Circuits

Part III: Operation in the Switching Mode

Part I

INTRODUCTION

Chapter I

PRINCIPLES OF EQUIVALENT CIRCUITS

The ordinary electric circuit elements of resistance, inductance, and capacitance are generally assumed to be *linear* and *bilateral.* Over some specified range of operation, the current through such components is a linear function of the applied voltage, and current flows with equal ease in either direction. As a result, it is possible to characterize such devices by single constants. Hence, it is said that a given component is a resistor of so many ohms, a capacitor of a certain number of farads, or an inductor of a specified number of henries. Actually, such a description is true only over a specified range of operation, and is therefore simply an *equivalent* representation of the device within certain operating limits.

In some cases, thermistors for example, the device might be very nonlinear over its customary operating range. It is not possible then to represent the component by a single equivalent. Instead, the characteristics are usually described graphically by showing a plot of the current through the device as a function of the applied voltage. This is called a *current-voltage* characteristic. For ferromagnetic or ferroelectric components, a plot of flux density as a function of field intensity is usually used.

Electronic devices are usually *nonlinear* and *unilateral.* In other words, elements such as vacuum and gas tubes, thermistors, transistors, and varistors have nonlinear current-voltage characteristics, and current flows through them more readily in one direction than in the other. The only really convenient method of representing the properties of such devices is graphically in the form of current-voltage characteristics.

The usefulness of these characteristics can hardly be overstated for they are widely used in the solution of practical problems. However, the fact remains that a purely graphical analysis of many electronic circuits is extremely difficult, and solutions obtained in this way lack generality. Analytical methods are required in some instances and preferred in others.

Practical necessity requires that the characteristics of electronic devices be presented in some equivalent, nongraphical form that lends itself to analytical treatment. The purpose of this chapter is to develop the concepts and techniques associated with the equivalent representation of nonlinear devices.[1] Although the discussion is based upon an analysis of current-voltage characteristics, the technique is perfectly general and can be used to derive equivalent representations of *any* nonlinear element.

The really startling development that has occurred in electronic circuitry resulted from the comparative ease with which electronic devices can be represented in simple and approximately equivalent terms.

1.1. Characteristics of Simple Circuits

In this chapter, circuits composed of conventional linear circuit elements will be developed that will have *very nearly* the same graphical characteristics as the electronic devices they are supposed to represent. Thus, it is understood at the outset that effort is directed toward *approximating* the behavior of electronic devices. It is specified that the approximation should be as accurate as possible consistent with reasonable simplicity of the equivalent circuits. Exact equivalents might be devised, but they would probably be quite complex and would have little advantage over graphical analysis.

The problem may be approached conveniently by a preliminary consideration of the graphical characteristics of a few simple circuits. Several examples are shown in figure (1.1). Consider figure (1.1a). The circuit is simply an ordinary linear resistance R. As a result, the current-voltage characteristic is a straight line. Since such an element is linear and bilateral, the slope of the curve is the same in either the positive or negative current direction. When current is plotted on the ordinate, the slope of the characteristic is $1/R$. Conversely, when voltage is the ordinate, the slope is R.

An examination of figures (1.1b) and (1.1c) shows that the characteristic curve of the resistor can be translated parallel to itself, either horizontally or vertically, by simply placing the resistor in series with a voltage source or in parallel with a current source. By combining these characteristics with a switch, many types of characteristic curves can be produced.

[1] This chapter was conceived while I was taking a course from Professor J. M. Pettit of the Electrical Engineering Department of Stanford University. It reflects Professor Pettit's treatment in some respects and where this occurs it is with his consent.

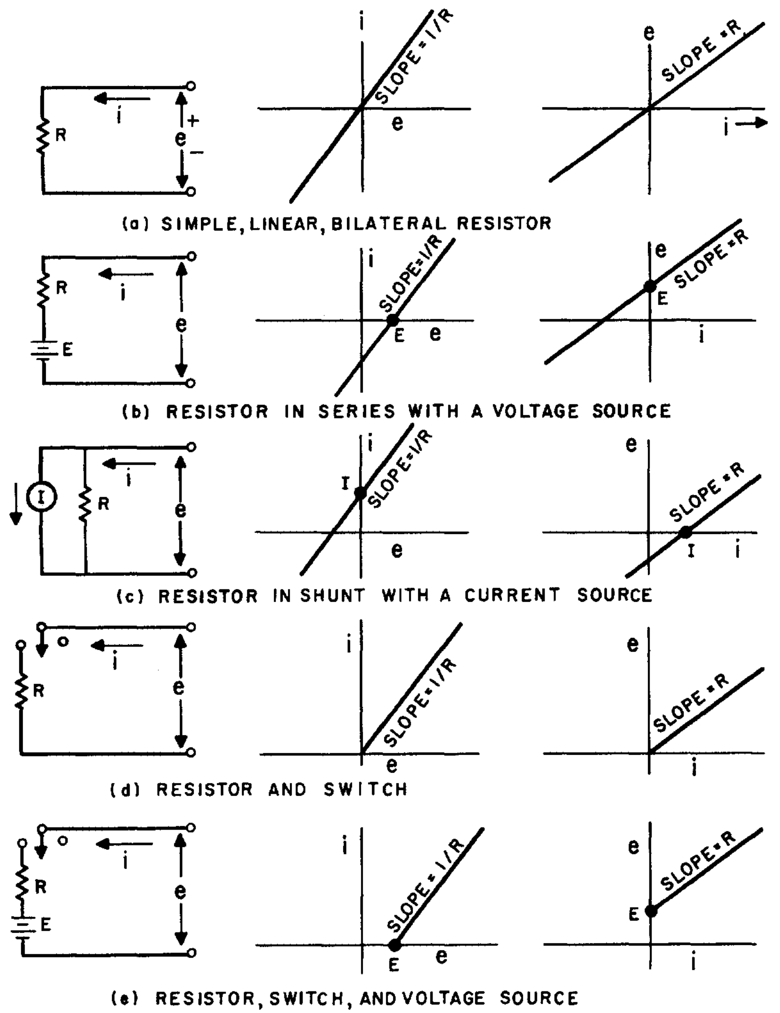


Fig. 1.1. Graphical characteristics of simple circuits.

A word of explanation concerning the switch is in order. In any given circuit it is assumed that the switch operates automatically at some specified current or voltage. In the cases shown, the switch automatically moves to the open position whenever the current through it is zero or negative. In other cases, switch operation could occur under some other specified conditions. A knowledge of these conditions is obviously an important factor in the application of these circuits to the equivalent representation of electronic devices.

1.2. Principle of Linear Approximation

It is proposed that the simple circuits shown in figure (1.1) can be used in various ways to approximate closely the characteristics of electronic devices. The method of doing this should be easily understood from the discussion that follows.

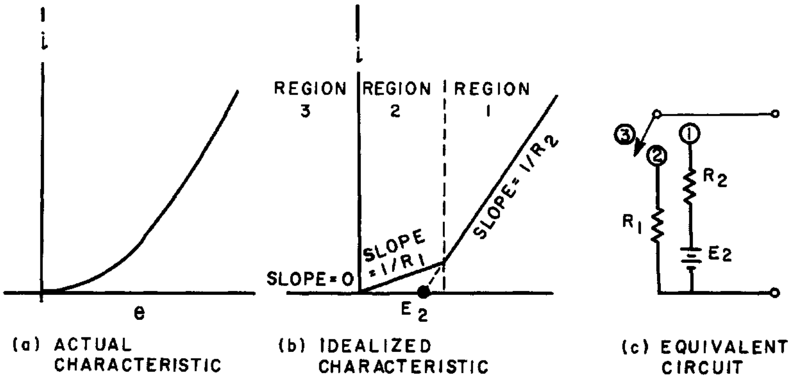


Fig. 1.2. The principle of linear approximation.

To illustrate the principle and technique to be used in developing equivalent circuits, suppose that a given nonlinear circuit element has the current-voltage characteristic shown in figure (1.2a). The reader is undoubtedly familiar with the idea of representing such a continuous curve by a discontinuous series of straight line segments, because such an approach is used at times in the introductory calculus courses. The actual curve and the segmented approximation become more nearly the same as the number of line segments used in the approximation is increased. In this text, this approximation process is called the *principle of linear approximation.*

In the case under discussion, a fairly accurate approximation is obtained by representing the actual curve by three line segments as shown in figure (1.2b). The linear approximation of the actual characteristic will be called the *idealized* characteristic.

Each line segment in this figure may be considered a *part* of a current-voltage characteristic of some simple circuit. In region 1 the characteristic would correspond to that produced by a resistance R_1 in series with a battery E_2. In region 2, the characteristic could result from a simple resistor R_2. The device is an open circuit in region 3. Therefore, three equivalent circuits are required to represent the device, one for each region. The complete equivalent circuit is the combination of these three circuits together with a selector switch to select automatically the proper circuit. The result is shown in figure (1.2c).

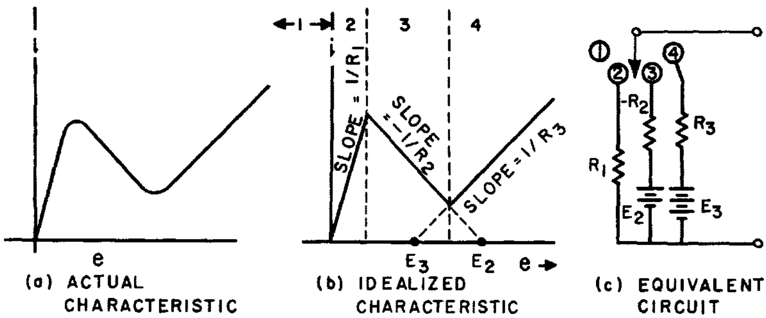


Fig. 1.3. Another example of linear approximation.

This example illustrates the basic technique to be followed in this chapter in developing equivalent circuits of electronic devices. The procedure involves three steps, as follows:

(1) Draw the actual current-voltage or volt-ampere characteristic.

(2) Approximate the characteristic by a judicious choice of straight line segments, using as many segments as required to obtain the desired accuracy.

(a) Identify the various regions of operation.

(b) Label the slopes and intercepts of the line segments.

(3) Construct an equivalent circuit for each line segment and use a switch to choose the proper one for a given operating condition.

One final example of the technique is given in figure (1.3). Note that the slope of the idealized characteristic is negative in region 3 and this leads to the negative resistance in the equivalent circuit. This type of characteristic occurs in oscillators and trigger circuits covered in later chapters.

1.3. Diode and Gas Triode Equivalent Circuits

The current-voltage characteristics of diodes, varistors, and gas triodes are fairly simple. If you understand the relatively complex examples covered in the preceding section, you should be able to construct the equivalent circuits given here without difficulty. Thus figures (1.4) through (1.7) are assumed to be self-explanatory.

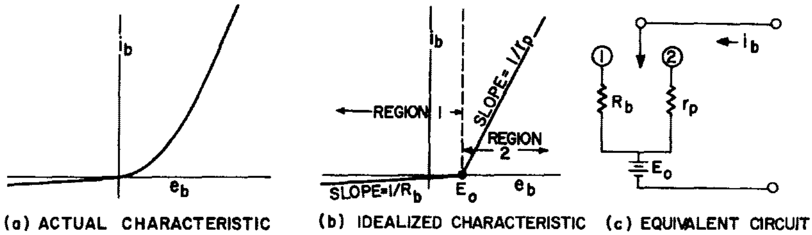


Fig. 1.4. Development of varistor equivalent circuit.

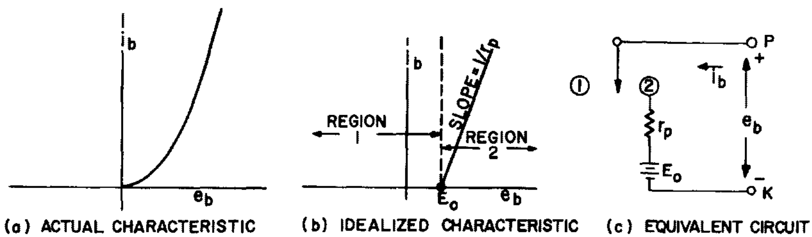


Fig. 1.5. Development of vacuum diode equivalent circuit.

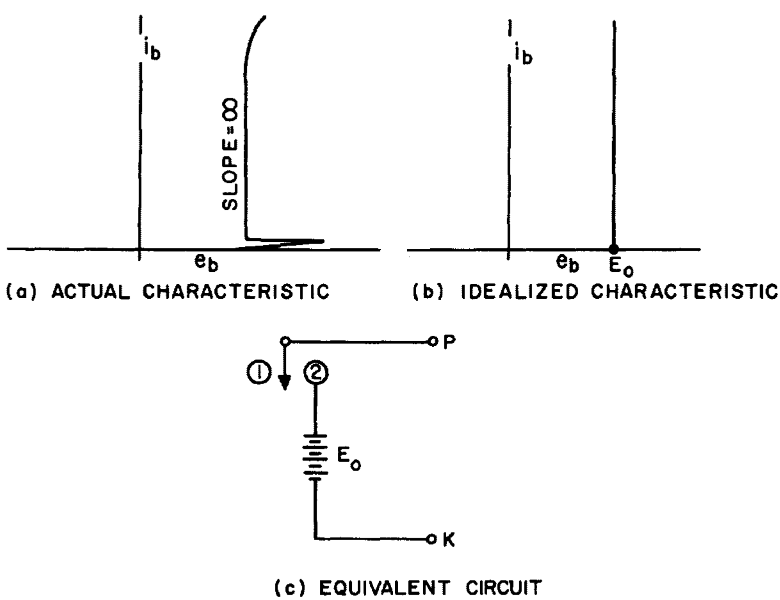


Fig. 1.6. Development of the equivalent circuit for a glow discharge tube up through the region of normal glow.

However, you should examine them carefully to familiarize yourself with the procedure and terminology.

The equivalent representation of gas triodes, such as thyratrons and ignitrons, is precisely the same as for the arc discharge diode except that the instant that the switches move from position 1 to position 2 is controlled by the grid or igniter electrode.

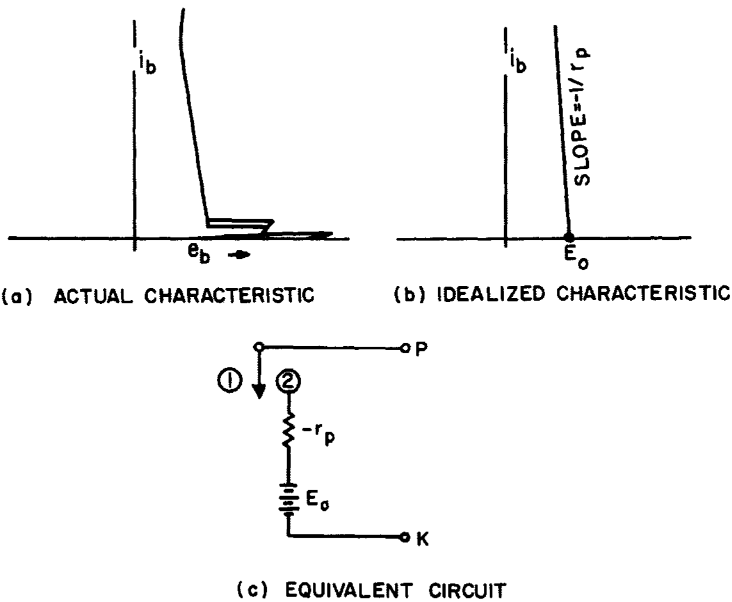


Fig. 1.7. Development of the equivalent circuit of arc discharge gas tubes.

1.4. Vacuum Triode Equivalent

The actual and idealized current-voltage characteristics of a vacuum triode are shown in figure (1.8). In contrast to the diode for which there was only a single curve, here there is a whole family of curves; the particular one generated is governed by the value of the grid voltage. A more elaborate equivalent circuit is required to approximate the characteristics of a triode vacuum tube.

From the idealized plate characteristics given in figure (1.8b) it appears that making the grid voltage more negative has the effect of translating the $e_c = 0$ curve to the right. The equivalent circuit of the $e_c = 0$ characteristic alone is identically the same as that of a vacuum diode. This equivalent can be modified to account for grid voltage

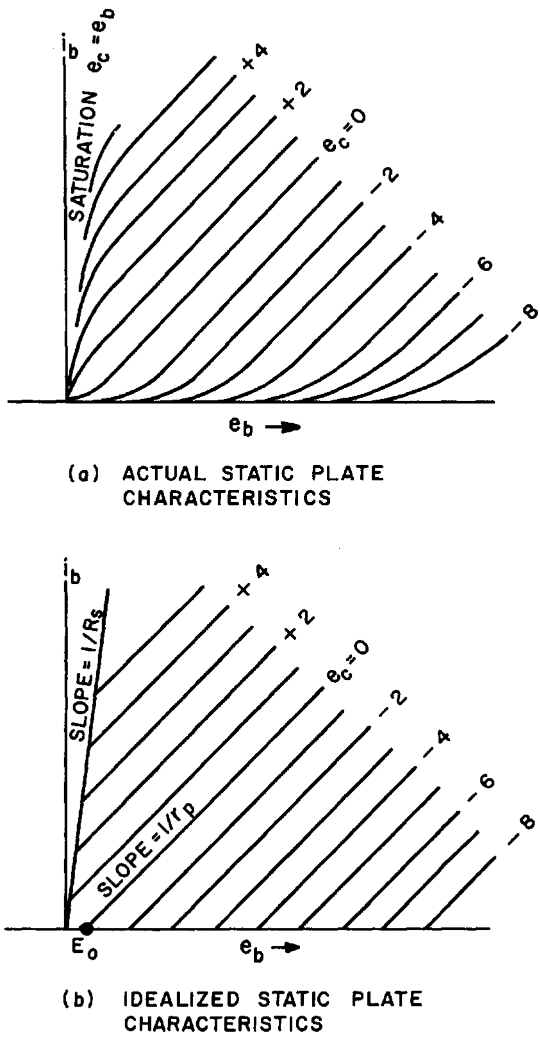
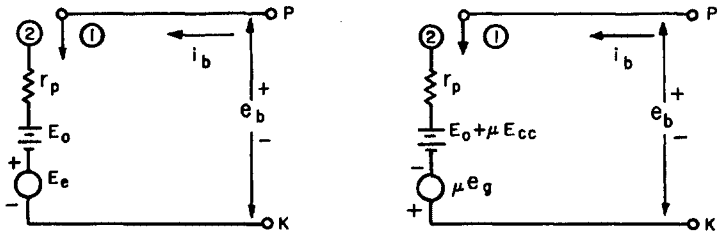


Fig. 1.8. Static plate characteristics of a triode vacuum tube.

variations by adding another voltage source E_e, in series with the intercept voltage E_o as shown in figure (1.9a). Actually, the action of the grid is rather more complex than this because the slope of the characteristic also changes as it is translated. Nevertheless, by a skillful selection of E_o and r_p, the characteristics of the equivalent circuit will closely approximate those of the triode. It remains to determine the relationship between the added equivalent generator, E_e, and the grid voltage, e_c, of the tube.



(a) FIRST EQUIVALENT CIRCUIT

(b) SECOND EQUIVALENT CIRCUIT

Fig. 1.9. Equivalent circuit of a vacuum triode for the linear region of operation, switch position 2, and the cutoff region, switch position 1.

Note that the polarity of the assumed generator, E_e, is positive in figure (1.9a). Therefore it carries its own polarity. If the actual polarity is negative, the equation for E_e will involve negative terms.

The amplification factor, μ, of the tube is defined as

$$\mu = -\frac{\partial e_b}{\partial e_c} = -\left(\frac{de_b}{de_c}\right)_{i_b = \text{const}} \tag{1.1}$$

In its exact form this equation applies only to differential voltage changes. It remains approximately true for incremental changes. Because the *idealized* characteristics are parallel, equally spaced straight lines, incremental and differential quantities have the same meaning. Hence, for the idealized characteristics,

$$\mu = -\left.\frac{\Delta e_b}{\Delta e_c}\right|_{i_b = \text{const}} \tag{1.2}$$

Or, in an alternative form,

$$\Delta e_b = -\mu\,\Delta e_c \quad \text{when} \quad i_b = \text{constant} \tag{1.3}$$

The Δe_b is the same thing as the E_e used in the equivalent circuit of figure (1.9a) because E_e *is* the Δe_b evaluated along the line of constant plate current, $i_b = 0$. Therefore

$$E_e = -\mu\,\Delta e_c \tag{1.4}$$

Here E_e is the *change* in plate voltage measured from the intercept voltage E_o of the $e_c = 0$ characteristic to the intercept of any arbitrary characteristic. Hence, the reference grid voltage is zero and $\Delta e_c = e_c$. Therefore

$$E_e = -\mu e_c \tag{1.5}$$

Ordinarily, the *total* grid voltage, e_c, involves two terms, as follows:

$$e_c = e_g - E_{cc} \tag{1.6}$$

where $-E_{cc}$ = negative d-c grid voltage = grid bias; e_g = variational

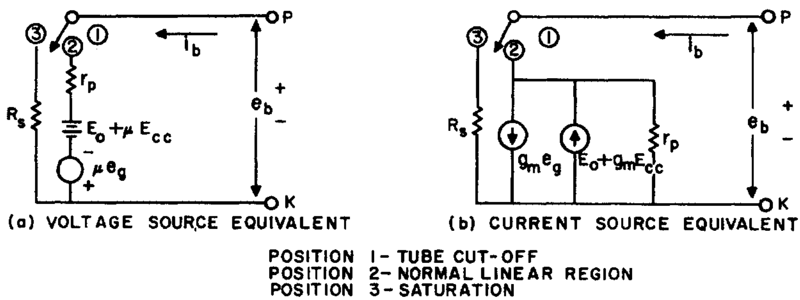


Fig. 1.10. Complete equivalent circuits of a triode vacuum tube showing three regions of operation.

or signal voltage. Therefore the equivalent voltage source simulating the action of the grid is

$$E_e = -\mu e_g + \mu E_{cc} \tag{1.7}$$

This equation leads to the equivalent circuit shown in figure (1.9b), in which all d-c terms have been combined into a single direct voltage source.

This equivalent circuit describes the operation of the tube in either of two regions:

(1) In the normal region of linear operation corresponding to switch position 2.

(2) In the *cutoff* or nonconducting region corresponding to switch position 1.

The equivalent circuit can be completed by including another switch contact and circuit to account for operation in the *saturation* region. In this mode of operation the tube acts like a simple resistor of R_s ohms, so that the complete equivalent circuit appears as shown in figure (1.10). The switch moves to position 3 and saturation results whenever the grid voltage is equal to or greater than the plate voltage.

The equivalent circuit of figure (1.10a) is called the *voltage source* equivalent circuit and sometimes the *Thevenin* equivalent circuit. An alternative form, called the *current source* or *Norton* equivalent, is shown in figure (1.10b). The method for converting one of these circuits into the other is covered in chapter 2.

1.5. Pentode and Beam Power Tube Equivalent

The actual and idealized static plate characteristics of a pentode are given in figure (1.11). Beam power tubes are sufficiently similar so that

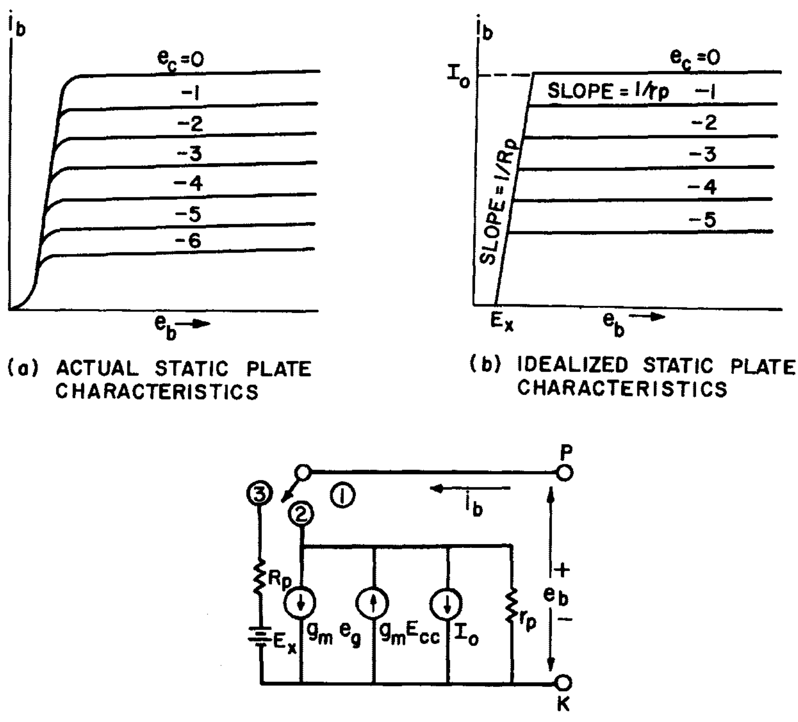


Fig. 1.11. Characteristics and equivalent circuit of a pentode or beam power tube.

they will not be treated separately. The eventual current source equivalent circuit of the pentode is given in figure (1.11c).

The region of normal linear operation corresponds to switch position 2. When the switch is in position 3 the pentode is said to be operated in the *bottoming region*. In this mode of operation the plate current is independent of the grid voltage just as in the case of a saturated triode. The switch moves to position 1 whenever the plate current through the tube is zero. This may be caused by lowered plate voltage or by the grid voltage exceeding the cutoff potential in the negative direction.

The equivalent circuits for switch positions 1 and 3 are easily constructed by the standard method previously outlined. The circuit in switch position 2 can be derived from the current source equivalent for the triode in figure (1.10b) simply by reversing the polarity of E_o or the direction of the current source $I_o = E_o/r_p$. This change is obviously necessary when you examine the characteristic and see where the E_o intercept must fall.

It is also possible to derive the current source equivalent in region 2 directly from plate characteristics, using the current intercept I_o and the definition of the mutual transconductance g_m instead of μ.

1.6. The Question of Nonlinearity

The equivalent circuits derived in the preceding sections have always involved a linear resistor as an essential part of the circuit. In the chapter introduction it was noted that tube characteristics are always nonlinear to some extent, and this nonlinearity persists even when operation is confined to a restricted region corresponding to a single position of the switch in the equivalent circuit. The equivalent circuits that have been developed do not account for the presence of any nonlinearity for any given single switch position.

A fairly thorough discussion of nonlinearity, its effects and applications, is presented in chapter 11. At that time it will be shown that the linear equivalent circuit can be retained and the nonlinearity accommodated by inserting additional generators having the nature of small correction factors. These added generators may be important or unimportant, depending upon the circuit application and operating conditions. Thus, to avoid complicating the equivalent circuits, it seems advisable to retain the present simplified and linear form and to consider the nonlinearities only when they become significant. Fortunately, this will not be necessary in a large proportion of the circuits presented in this text.

1.7. Interelectrode Capacitances and Lead Inductances

Because the electrodes of vacuum tubes and semiconductor devices are of finite size, are separated by finite distances, and are connected to their external circuits by leads of finite length, it is inevitable that the electrode leads will possess inductance and that there will be capacitances between the electrodes. The relative importance of these constants generally increases with frequency; some become significant

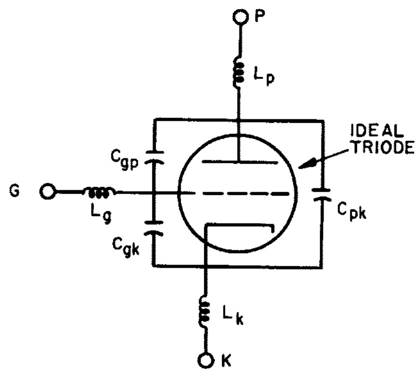


Fig. 1.12. Effect of finite dimensions on a triode.

at audio frequencies; others are not important except at ultrahigh frequencies. Any accurate representation of an electronic device must include these constants.

Strictly speaking, the lead inductances and interelectrode capacitances are distributed elements. However, even at ultrahigh frequencies

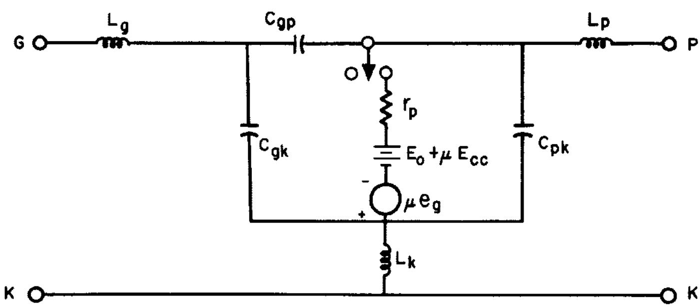


Fig. 1.13. Equivalent plate circuit of a vacuum triode including interelectrode capacitances and lead inductances.

the lead and electrode dimensions are small compared with the wavelengths involved so that these constants can be considered to be *lumped* elements. With this in mind, the representation of a triode shown in figure (1.12) is readily understood. Similar figures can be drawn for diodes, pentodes, varistors, and transistors.

The *ideal triode* shown in figure (1.12) is simply a triode that does not have interelectrode capacitance or lead inductance. This ideal triode has an equivalent plate circuit of the form shown in figure (1.10). Consolidating this equivalent circuit with figure (1.12) yields the rather awesome circuit of figure (1.13). The complication is obvious.

The effects of interelectrode capacitance are treated in considerable detail in chapter 3. Lead inductance effects are generally neglected except at high frequencies.

1.8. Modes of Operation

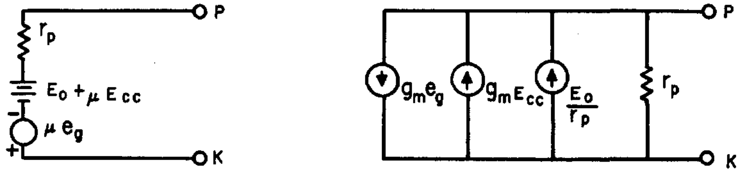
The treatment of electronic circuits presented in this book has two major subdivisions:

(1) Class A operation, covered in part II.

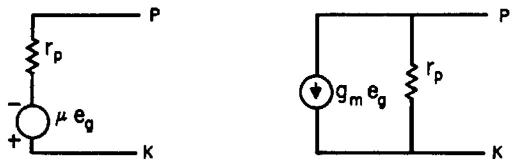
(2) Operation in the switching mode, covered in part III.

This subdivision of material is based upon the behavior of the electronic device in the circuit as represented by its equivalent circuit. Every electronic component has a switch in its equivalent circuit. This switch may do either of two significantly different things, depending upon the nature of the signal voltages and currents. That is, the switch in the equivalent circuit: (1) may remain in one position at all times, or (2) may move back and forth between the various contacts. If the switch is always fixed in the position corresponding to normal linear operation, the device is said to operate in *Class A*. If the switch moves back and forth between contacts, *operation in the switching mode* is obtained.

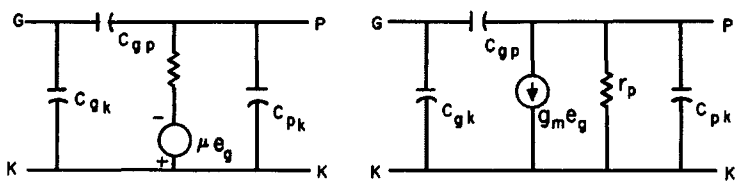
When a vacuum tube operates in the switching mode, the complete equivalent circuits previously derived must be used. However, a considerable simplification is possible when operation is Class A. In the case of vacuum tubes the switch is then always in position 2 in the circuits of figure (1.10). As a result, the equivalent circuit reduces to that shown in figure (1.14). Moreover, in class A operation it is often necessary to consider *only* the variational or signal components of current and voltage. The principle of superposition allows us to replace all d-c terms by their internal impedances. The equivalent circuits simplify to those shown in figure (1.14b). More complete equivalent circuits including the interelectrode capacitances are given in figure (1.14c). These equivalent circuits will be used to represent vacuum tubes throughout part II of the book. The *complete* equivalent circuits including the switches and d-c terms are used in part III.



REVERSE POLARITY OF E_0 TERM FOR A PENTODE
(a) FIRST EQUIVALENT CIRCUITS INCLUDING DC TERMS



(b) SIMPLIFIED EQUIVALENT CIRCUITS FOR SIGNAL FREQUENCY COMPONENTS



(c) INTERELECTRODE CAPACITANCES INCLUDED

Fig. 1.14. Equivalent circuits for Class A operation of vacuum tubes, except diodes.

1.9. Transistor Equivalent Circuits

Sample static characteristics of a point contact transistor are given in figure (1.15). Similar characteristics are obtained for junction transistors; they differ mainly in the slope resistance values.

The idealization of the transistor characteristics is *not* an immediately obvious procedure; careful thought is required. This is especially true because the characteristics given in figure (1.15) are more nonlinear than the usual case. The choice was deliberately made so that we could have the experience of working out a relatively difficult equivalent circuit.

The essential key to the choice of regions required for the linear approximation is found in the *forward* characteristics. You can see that there are four fairly distinct regions of operation for any given value of collector current. These regions are defined as follows.

(1) Region 1. The emitter current I_e is negative and the collector voltage V_c is virtually independent of the emitter current.

(2) Region 2. Transition between regions 1 and 3.

(3) Region 3. The emitter is operating in the forward, positive current direction and the collector voltage is nearly a linear function of

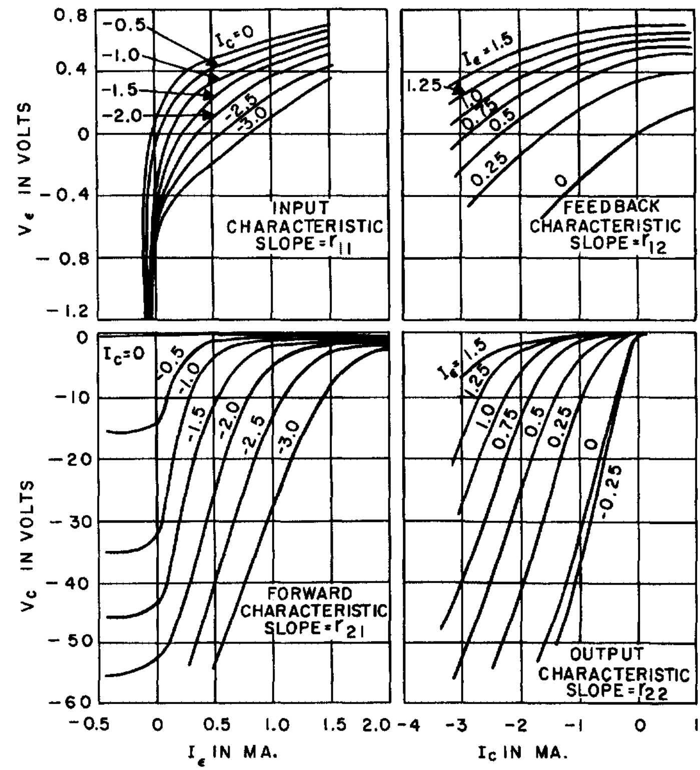


Fig. 1.15. Sample static characteristics of a point contact transistor. (From R. M. Ryder and R. J. Kircher, "Some Circuit Aspects of the Transistor," *Bell System Tech. J.*, vol. 28, pp. 367–401, July, 1949.)

the emitter current. This is the usual linear region of operation of a transistor.

(4) Region 4. The collector voltage is practically independent of the emitter current, but it does depend somewhat upon the collector current.

The regions of operation have been defined, so the characteristics can now be idealized. Draw the dividing lines between the four regions on the forward characteristics as shown in figure (1.16). Now take data from these lines so that similar lines can be plotted on each of the other three sets of characteristic curves. The final result will appear as shown in figure (1.16).

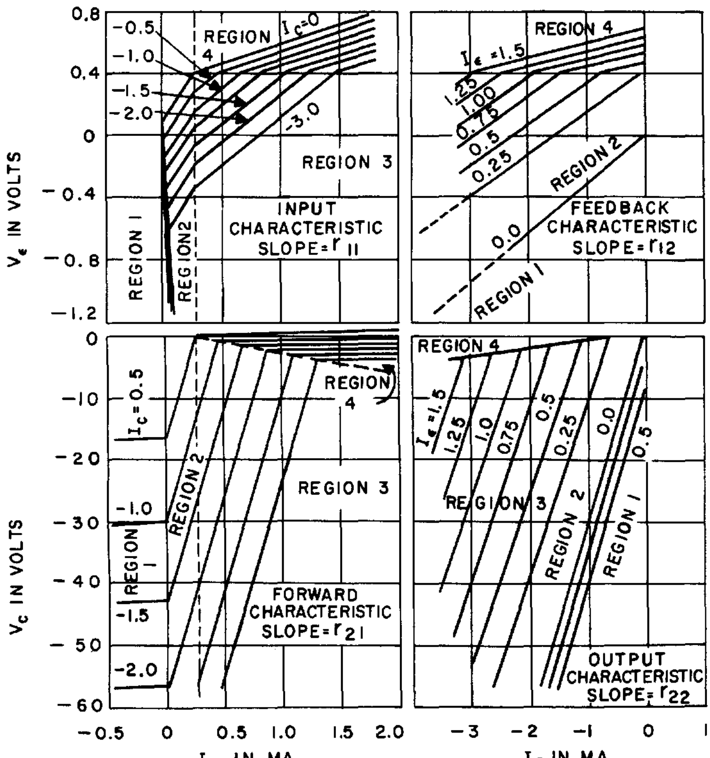


Fig. 1.16. One way of idealizing the static characteristics of transistors.

In most cases region 2 is so small that it need not be considered. In other words, region 3 will usually be combined with region 2 because the transition region is usually smaller and more abrupt than it appears here.

With the idealized characteristics completed, as shown in figure (1.16), it is a simple matter to construct the equivalent circuit. To avoid confusion in symbolism, the following terminology is used for the slope resistances and intercept voltages in the various regions.

(1) Region 1, superscript 0
(2) Region 2, superscript ′
(3) Region 3, no superscript
(4) Region 4, superscript ″.

Subscripts are used with the intercept voltages as follows:

(1) Input characteristics, subscript 1
(2) Output characteristics, subscript 2.

Consider the idealized input characteristics. There are four regions of operation, so that a four-position selector switch is required. Each equivalent circuit corresponding to each switch position will consist of a series combination of slope resistance, intercept voltage, and a variable voltage source of a magnitude determined by the collector current. Designate this variable component of voltage in any switch position E_x and make it positive in the conventional sense. Therefore, for any region of operation you can see that

$$E_x = \left(\frac{\Delta V_\varepsilon}{\Delta I_c}\right)_{V_c \, \mathrm{const}} \Delta I_c \tag{1.8}$$

However, because the idealized characteristics are parallel straight lines in any one region, incremental and differential quantities have the same meaning, so that

$$E_x = \left(\frac{\partial V_\varepsilon}{\partial I_c}\right) \Delta I_c = r_{12} \, \Delta I_c \tag{1.9}$$

The slope resistance can be superscripted to specify the region of operation.

Now ΔI_c is measured from the $I_c = 0$ characteristic as the reference. Therefore

$$\Delta I_c = I_c \tag{1.10}$$

and

$$E_x = I_c r_{12} \tag{1.11}$$

Appropriate superscripts are now used to make this specification apply to all regions of operation.

The results of this derivation are shown by the equivalent circuit given in figure (1.17). This circuit is a simplification because common terms were combined wherever possible. You can work it out and see that it comes out this way.

An exactly similar analysis can be made for the output characteristics and the result will appear as shown in figure (1.17). This is the complete equivalent circuit for the transistor. As noted previously, region 2 can frequently be omitted entirely.

Generally speaking, the emitter and collector currents will consist of two terms:

(1) A d-c or bias term that is time invariant.

(2) A signal or variational component.

Therefore

$$I_\varepsilon = i_\varepsilon + I_{\varepsilon\varepsilon} \tag{1.12}$$

$$I_c = i_c - I_{cc} \tag{1.13}$$

You can see that the generators in the equivalent circuit of figure

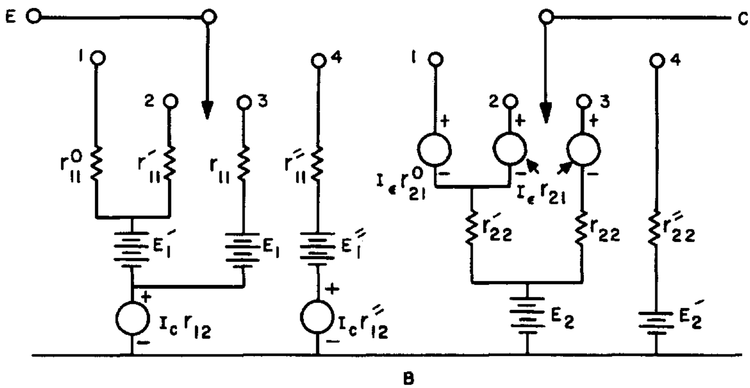


Fig. 1.17. Complete transistor equivalent circuit. E_1' and E_2 are virtually zero in most cases; generally $I_\varepsilon = i_\varepsilon + I_{\varepsilon\varepsilon}$ and $I_c = i_c - I_{cc}$.

(1.17) will consist of two components:

$$I_c r_{12} = i_c r_{12} - I_{cc} r_{12} \tag{1.14}$$

$$I_\varepsilon r_{21} = i_\varepsilon r_{21} + I_{\varepsilon\varepsilon} r_{21} \tag{1.15}$$

Unless a very sophisticated analysis is desired, it is usually possible to neglect the intercept voltages, E_1, E_1'', E_2, and E_2''.

When a transistor is operated in class A, both switches are normally in position 3. Operation in the switching mode may cause the switch to move over all switch positions or just from region 1 through 3 and back again. This depends upon the function being fulfilled by the transistor.

1.10. Transistor Equivalent for Class A Operation

A transistor, like a vacuum tube, has two main modes of operation. The class A equivalent circuit is easily deduced from the general equivalent circuit of figure (1.17) by applying the same simplification process as that followed for vacuum tubes in section (1.8). The result is shown in figure (1.18a). This circuit is useful, but it has been found[2] more convenient to convert to the equivalent tee section shown in figure (1.18b). The resistances in this circuit are defined as follows:

r_ε = emitter resistance
r_c = collector resistance
r_b = base resistance
r_m = mutual resistance

Before the equivalent tee circuit can be used, equations expressing the tee resistances in terms of the slope resistances r_{11}, r_{22}, r_{12}, and r_{21} are required.

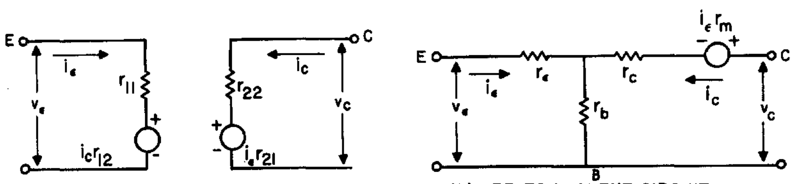


Fig. 1.18. Transistor equivalent circuits for Class A operation.

The required equivalence formulas for the two circuits of figure (1.18) are easily established by writing the network equations. Thus for figure (1.18a),

$$V_\varepsilon = I_\varepsilon r_{11} + I_c r_{12} \tag{1.16}$$

$$V_c = I_c r_{22} + I_\varepsilon r_{21} \tag{1.17}$$

Similarly for figure (1.18b),

$$V_\varepsilon = I_\varepsilon(r_\varepsilon + r_b) + I_c r_b \tag{1.18}$$

$$V_c = I_c(r_c + r_b) + I_\varepsilon(r_b + r_m) \tag{1.19}$$

From a term-by-term comparison of these two pairs of equations it is clear that

$$r_{11} = r_\varepsilon + r_b \tag{1.20}$$

$$r_{22} = r_c + r_b \tag{1.21}$$

$$r_{12} = r_b \tag{1.22}$$

$$r_{21} = r_b + r_m \tag{1.23}$$

[2] R. M. Ryder and R. J. Kircher, "Some Circuit Aspects of the Transistor," *Bell System Tech. J.*, vol. 28, pp. 367–401, July, 1949; and R. L. Wallace, Jr. and W. J. Pietenpol, "Some Circuit Applications of *n-p-n* Transistors," *Bell System Tech. J.*, vol. 30, pp. 530–563, July, 1951.

These four equations can be rewritten in the reverse form as

$$r_b = r_{12} \tag{1.24}$$

$$r_\varepsilon = r_{11} - r_{12} \tag{1.25}$$

$$r_c = r_{22} - r_{12} \tag{1.26}$$

$$r_m = r_{21} - r_{12} \tag{1.27}$$

Another useful relationship can be derived in terms of these resistances. The current amplification factor α of a transistor is defined as

$$\alpha = \left.\frac{\partial I_c}{\partial I_\varepsilon}\right|_{V_\varepsilon = \text{const}} \tag{1.28}$$

This can be rewritten as

$$\alpha = \frac{(\partial I_c/\partial V_c)}{(\partial I_\varepsilon/\partial V_c)} = \frac{1/r_{22}}{1/r_{21}} = \frac{r_{21}}{r_{22}} \tag{1.29}$$

$$\alpha = \frac{r_b + r_m}{r_b + r_c} \tag{1.30}$$

1.11. Summary

By using the principle of linear approximation, electronic devices can be represented by approximately equivalent circuits composed of simple linear resistances, voltage and current sources, and switches. When the occasion demands, interelectrode capacitances and lead inductances can be added to the simple equivalent circuits.

There are only two fundamentally different modes of operating electronic devices and these were defined as class A operation and the switching mode.

As a result of the simple equivalent representation of electronic devices, the problem of electronic circuit theory is largely reduced to ordinary electric circuit theory.

Although the principle of linear approximation has been applied only to current-voltage characteristics, it can be applied to any physically realizable, nonlinear element such as ferromagnetic inductors and ferroelectric capacitors, d-c motors, photocells of all types, strain gages, and so on.

PROBLEMS

1.1. Derive the equivalent circuit for the linear region of operation of a pentode using current intercepts and the definition of the mutual transconductance g_m.

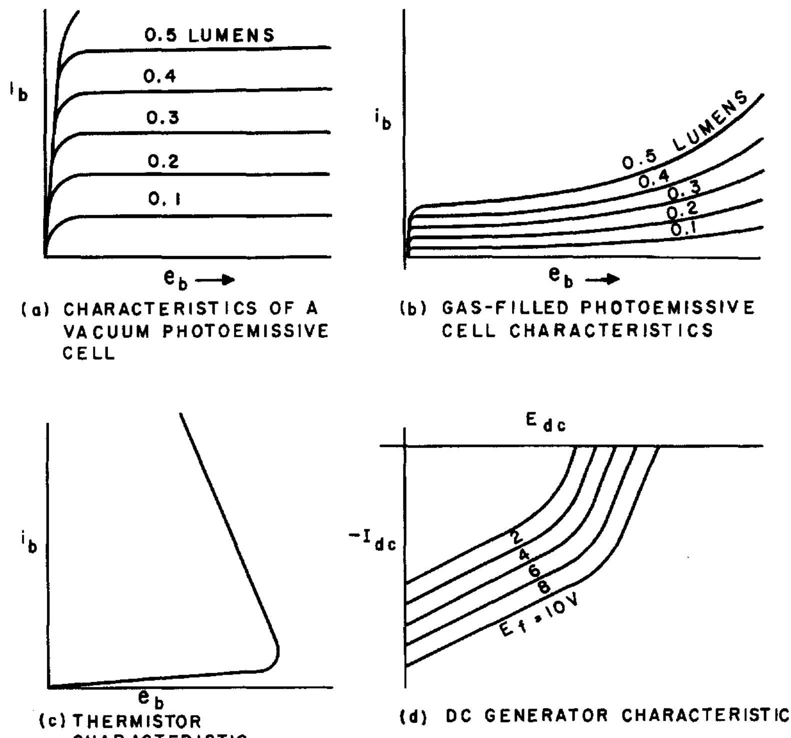


Fig. 1.19.

1.2. The current-voltage characteristics of a vacuum tube photoemissive cell are given in figure (1.19a). Derive an equivalent circuit from these characteristics. Use the current source equivalent in the linear region and define a parameter as follows:

$$S_c = \frac{\partial i_b}{\partial L} \quad \text{where} \quad \begin{array}{l} i_b = \text{plate current} \\ L = \text{lumens} \end{array}$$

1.3. Repeat problem (1.2) for the gas-filled photoemissive cell having the static characteristics shown in figure (1.19b).

1.4. Construct the equivalent circuit of the thermistor having the current-voltage characteristic shown in figure (1.19c).

1.5. Derive equivalent circuits for the three transistor characteristics given in figure (17.31).

1.6. The external characteristics of a particular separately excited d-c generator are shown in figure (1.19d). The current is shown as negative because it flows out of the generator rather than into it, which is the positive direction assumed in the examples in this chapter. The generated and field voltages are related by

$$K = \frac{\partial E_{dc}}{\partial E_f}$$

Derive an equivalent circuit for the generator.

Chapter 2

PRINCIPLES OF CIRCUIT THEORY

The first chapter of this book had one main objective, the development of approximate equivalent circuits composed of linear circuit elements to represent the various electronic devices such as gas and vacuum tubes and transistors. Once this has been achieved, electronic circuits can be treated with the powerful tools of linear circuit analysis and some useful methods and results can be illustrated.

This chapter presents a brief summary of some of the more general aspects of linear circuit analysis. Regardless of your background, you should be familiar with this material, for it is used repeatedly in the remainder of the book.

2.1. Responses

One of the basic techniques characteristic of the scientific method is the study of the *responses* of physical systems to known applied stimuli. Whether the physical system is a human, a radio, white mice, or an electronic computer does not alter this basic approach. A known stimulus is applied and the system response is studied. The technique can be used experimentally or in theoretical analysis. Such studies are useful in many ways. For example:

(1) If the physical system operates in an unknown manner or according to unknown principles, the ratio of the stimulus to the response may be used as an aid in deducing the nature of the system.

(2) The relative merits of various systems designed to provide the same service can be compared by applying a standard stimulus and evaluating the responses relative to a proposed ideal.

(3) By using standardized stimuli, past experience with various systems can be easily tabulated and filed away for future reference, study, and correlation.

In circuit theory this basic problem is indicated symbolically as shown in figure (2.1). The stimulus is usually called the *excitation function*, and the response is appropriately called the *response function.* The ratio of the response function to the excitation function is called the system *transfer function.* That is

$$\begin{pmatrix}\text{transfer}\\ \text{function}\end{pmatrix} = \left(\frac{\text{response function}}{\text{excitation function}}\right) \tag{2.1}$$

This can be rewritten in an alternative form as

$$\begin{pmatrix}\text{response}\\ \text{function}\end{pmatrix} = \begin{pmatrix}\text{excitation}\\ \text{function}\end{pmatrix}\begin{pmatrix}\text{transfer}\\ \text{function}\end{pmatrix} \tag{2.2}$$

Finally, the reciprocal of the transfer function is called the *characteristic function,* so that

$$\begin{pmatrix}\text{response}\\ \text{function}\end{pmatrix} = \left(\frac{\text{excitation function}}{\text{characteristic function}}\right) \tag{2.3}$$

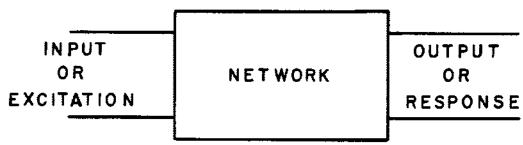


Fig. 2.1. General network terminology.

It will be shown later that the characteristic and transfer functions are determined solely by the properties of the system.

The electrical engineering curriculum in circuit theory is based upon equation (2.3). The particular course presented depends upon the nature of the excitation function, as follows:

(1) When the excitation function is time invariant, the course is called direct current (d-c) circuit theory.

(2) When the excitation function is a single frequency, constant amplitude sinusoid, the course is called alternating current (a-c) circuit theory.

(3) If the excitation function is neither of the preceding two types, the course is usually called transient analysis.

All three of these subdivisions of circuit theory must be used in the analysis of electronic circuits and we are naturally prompted to ask if there is not some general method of formulating circuits, irrespective of the excitation, so that the excitation will be relegated to a secondary role. Fortunately, such a method does exist and it is no more difficult than the usual a-c and d-c methods. It makes use of the *Laplace transformation* and it is treated briefly in this chapter after a short review of some important aspects of circuit theory.

2.2. Elements of Electric Circuits

Electric circuits and networks are connected systems of *active elements*, such as voltage and current sources, and the *passive elements* of resistance, inductance, and capacitance. The active elements supply energy to the circuit. The passive elements of capacitance and inductance act as reservoirs of electrostatic and magnetic energy, respectively, while resistance causes an irreversible transformation of electric energy into heat.

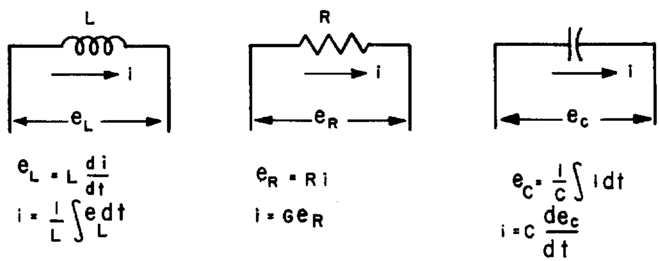


Fig. 2.2. Passive electric circuit elements.

The circuit symbols of the three passive elements are given in figure (2.2) together with the fundamental equations that express their behavior characteristics.

When a number of circuit elements are combined into a circuit or network, the behavior of the system can be evaluated by the basic laws of electric circuit theory. These are called *Kirchhoff's laws* and are as follows:

(1) *Kirchhoff loop law*:
Around any closed path, called a *loop*, in a network, the sum of the instantaneous voltage drops in a specified direction is zero. That is, around a closed path,

$$\sum e_K(t) = 0$$

(2) *Kirchhoff node law*:
At a junction, called a *node*, of two or more circuit elements, the sum of the currents into (or away from) the node is zero. That is, at a node,

$$\sum i_K(t) = 0$$

The particular law used depends upon the nature of the problem at hand. The following examples should make this clear.

Consider the series *RLC* circuit shown in figure (2.3a). There are three *pairs* of independent nodes in this circuit, but only a single loop.

Thus three equations would be required to formulate the circuit on the node basis, while only one equation is needed on the loop basis. Clearly, loop formulation would be the best approach. The loop equation for this circuit is

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt - V(t) = 0$$

or

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = V(t) \tag{2.4}$$

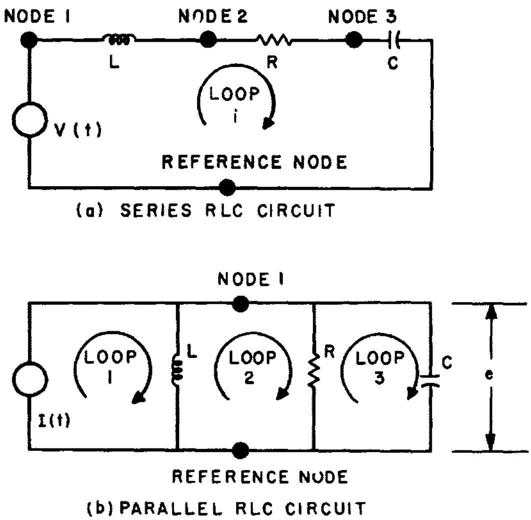


Fig. 2.3. Typical electric circuits.

The parallel *RLC* circuit of figure (2.3*b*) has only a single pair of independent nodes, but there are three geometric loops. Therefore the circuit is best adapted to node formulation and the circuit equation is

$$C\frac{de}{dt} + Ge + \frac{1}{L}\int e\,dt - I(t) = 0$$

or

$$C\frac{de}{dt} + Ge + \frac{1}{L}\int e\,dt = I(t) \tag{2.5}$$

It should be noted that there is a one-to-one correspondence between the mathematical forms of equations (2.4) and (2.5). The circuits are said to be *duals* when their mathematical formulations are so related.

Most circuits cannot be formulated this easily and it is often necessary to use a combination of loop and node analysis. In other cases, network theorems can be used to simplify the network into purely series or parallel forms.

2.3. Some Useful Network Theorems

Several network theorems are stated here without proof because it is assumed that the reader is familiar with their validation through other books. Proofs may be found elsewhere.[1]

One of the great advantages resulting from the restriction of interest to linear circuits is that the *principle of superposition* can be used. As applied to electric circuits, this principle states that if a network contains several active elements, the complete network solution can be obtained with equal validity in either of two ways:

(1) The complete solution can be obtained by writing the Kirchhoff equations with all active sources considered in the formulation.

(2) A separate solution can be obtained for each active element considered alone, with all other active elements replaced by their internal impedances. The sum of all such solutions is the same as the complete solution obtained in (1).

This principle is immensely helpful for it is quite common to find a network with d-c sources and a-c sources of different frequency. By superposition, a separate solution independent of other solutions can be obtained for each generator. It was the use of this principle that permitted the development of the simplified class A equivalent circuits of vacuum tubes and transistors in which all active d-c elements were replaced by their internal impedances and only variational quantities are considered.

Thevenin's theorem is frequently used in this text. According to this theorem, any two-terminal network can be replaced by an ideal voltage source (where ideal means zero internal impedance) in series with an impedance. The magnitude of the voltage generated by this source is equal to the open-circuit voltage measured across the two terminals of the network, and the series impedance is the impedance looking into the network with all active elements replaced by their internal impedances. This theorem is illustrated in figure (2.4a).

A corollary to this theorem is *Norton's theorem*, which is illustrated in figure (2.4b). In this case the two-terminal network is replaced by an

[1] See for example, Charles R. Vail, *Circuits in Electrical Engineering*, Prentice-Hall, Inc., New York, 1950.

equivalent circuit composed of an ideal current source (where ideal means infinite internal impedance) of generated current equal to the short-circuit current through the terminals. It is placed in shunt with an impedance Z_{in}, equal to the impedance defined in the statement of Thevenin's theorem.

The *compensation theorem* can be used to advantage in many cases. According to this theorem, any impedance in a network can be replaced by an ideal generator with a terminal voltage equal to the voltage drop across the impedance at every instant.

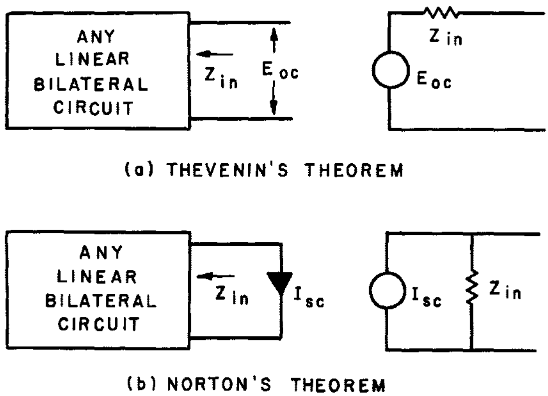


Fig. 2.4. Theorems on two terminal networks.

2.4. A Generalized Concept of Frequency

The fundamental equations expressing the behavior of ordinary linear circuit elements were given in figure (2.2) and are reproduced in equations (2.6).

$$\left.\begin{aligned} e_R &= Ri && (a) \\ e_L &= L\frac{di}{dt} && (b) \\ e_C &= \frac{1}{C}\int i\,dt && (c) \end{aligned}\right\} \qquad (2.6)$$

If the current flowing through each of these circuit elements is sinusoidally oscillating with constant amplitude at an angular velocity ω, it is an easy matter to state this fact mathematically with the tools of a-c circuit theory. That is,

$$i = I\varepsilon^{j\omega t} \qquad (2.7)$$

where $I=$ amplitude of the oscillating current; $\omega=2\pi f=$ angular velocity of the oscillation; $f=$ frequency of the oscillation. Consequently, the voltage drops across each of the three types of circuit elements can be obtained by substituting this equation for the current into the voltage equations given in (2.6).

$$\left.\begin{aligned} e_R &= RI\varepsilon^{j\omega t} && (a) \\ e_L &= j\omega LI\varepsilon^{j\omega t} && (b) \\ e_C &= \frac{1}{j\omega C}I\varepsilon^{j\omega t} && (c) \end{aligned}\right\} \quad (2.8)$$

Note that the exponential function $\varepsilon^{j\omega t}$, which symbolized that the current was a sinusoidal function of time, retains its form through differentiation and integration, and every voltage equation contains such a factor. Because this $\varepsilon^{j\omega t}$ term is just an angle function of the form

$$\varepsilon^{j\omega t} = \cos\omega t + j\sin\omega t = 1\,\underline{/\omega t} \quad (2.9)$$

then it can be considered as a unit vector or phasor rotating at an angular velocity ω. The phasor diagrams of a-c theory were derived from this viewpoint because all three voltages and the current contain this factor. Thus a phasor diagram is a stroboscopic picture, so to speak, of the relative positions existing between the various phasors as they all rotate with constant speed ω. In the case of d-c circuits, ω is zero and there is no rotation.

Now suppose that the current is *not* the simple sinusoid specified in equation (2.7), but is multiplied by an additional factor as shown in equation (2.10).

$$i = I\varepsilon^{j\omega t}\varepsilon^{\sigma t} \quad (2.10)$$

In this case the original constant amplitude sinusoidal current $I\varepsilon^{j\omega t}$ has been multiplied by an exponential factor in which σ is a real number. Because this additional factor affects only the magnitude of the current, the equation could be rewritten

$$i = (I\varepsilon^{\sigma t})\varepsilon^{j\omega t}$$

If σ is positive the exponential term $\varepsilon^{\sigma t}$ continuously increases with time and the amplitude of the current also increases. On the other hand, if σ is negative, the amplitude of the current continuously decreases with increasing time. Thus, depending upon the sign attached to σ, for a given ω, equation (2.10) can represent either a growing or a decaying sinusoidal oscillation. If ω is zero, we then have a rising or decaying exponential current. If σ is also zero, the current is time invariant or pure direct current.

Equation (2.10) can be rearranged slightly as

$$i = I\varepsilon^{(\sigma + j\omega)t} \tag{2.11}$$

A new term can now be defined as follows:

$$s = \sigma + j\omega = \text{complex frequency} \tag{2.12}$$

Hence, the equation for the current reduces to

$$i = I\varepsilon^{st} \tag{2.13}$$

and the corresponding voltage equations are

$$\left.\begin{aligned} e_R &= RI\varepsilon^{st} && (a) \\ e_L &= sLI\varepsilon^{st} && (b) \\ e_C &= \frac{1}{sC} I\varepsilon^{st} && (c) \end{aligned}\right\} \tag{2.14}$$

As before, it is observed that the exponential factors are common to all currents and voltages.

This concept of a complex frequency, $s = \sigma + j\omega$, involving both real and imaginary parts is not necessarily a mathematical fiction. As has been shown, a term involving a complex frequency exponential term can signify direct current, growing or decaying direct current, steady state alternating current, or growing or decaying sinusoids.

In the ordinary a-c case where $s = 0 + j\omega$, the impedance elements were found to be

$$R = \text{resistance}$$

$$j\omega L = \text{inductive impedance}$$

$$\frac{1}{j\omega C} = \text{capacitive impedance}$$

From equation (2.14) you can see that the impedance elements in the complex frequency case are

$$R = \text{resistance}$$

$$sL = \text{inductive impedance}$$

$$\frac{1}{sC} = \text{capacitive impedance}$$

There is no difference in form in the two cases. The only difference is that the $j\omega$ has been replaced by the complex frequency s. Thus, to evaluate impedances in terms of this generalized frequency, proceed according to normal a-c theory methods, merely replacing $j\omega$ by s. For example, the input impedance of the circuit of figure (2.5) as a function of complex frequency is

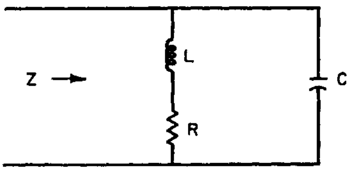


Fig. 2.5. Typical circuit.

$$Z(s) = \frac{(R + sL)(1/sC)}{R + sL + 1/sC} \tag{2.15}$$

Alternatively, in terms of conventional a-c frequency,

$$Z(j\omega) = \frac{(R + j\omega L)(1/j\omega C)}{R + j\omega L + 1/j\omega C} \tag{2.16}$$

It will be shown later that this concept of complex frequency is closely related to the variable in the Laplace transformation.

2.5. Simple Transformations

The Laplace transformation,[2] which is introduced in the next section, is a mathematical technique that transforms certain kinds of functions of a real variable into other functions of a complex variable. As a result of this functional transformation, a number of useful mathematical processes essential to circuit theory are simplified.

The use of transformations is not an unusual experience for electrical engineers. Following an analogy by Gardner and Barnes,[2] the process of taking logarithms involves the transformation of a number into some other number. As a result of the transformation, certain mathematical operations are simplified.

Another simple transformation is used in a-c circuit theory and was reviewed in the preceding section. In this case a sinusoidally oscillating current or voltage of fixed amplitude and frequency is represented by a rotating vector or phasor. This transformation permits an enormous simplification in the purely mechanical processes of computing circuit behavior.

In all these cases an initial uneasiness is often felt by the user, and the feeling persists until familiarity with the process is gained. The Laplace transformation is not a difficult tool to use. On the contrary, it is simple, direct, and immensely more powerful than more elementary methods.

[2] Murray F. Gardner and John L. Barnes, *Transients in Linear Systems*, John Wiley & Sons, Inc., New York, vol. 1, 1942.

2.6. The Laplace Transformation

It was previously stated that the Laplace transformation is a *functional transformation*, transforming certain classes of functions of a real variable into other functions of a complex variable. Specifically, in the case of many physical problems, a function of the real variable *time* t is transformed into a function of the complex frequency $s = \sigma + j\omega$. The transformation process is indicated symbolically:

$$\mathcal{L}[f(t)] = F(s)$$

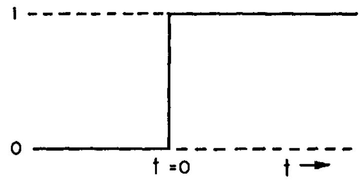


Fig. 2.6. Unit step function.

This states that the Laplace transformation of $f(t)$ is equal to $F(s)$, where $\mathcal{L}$ = Laplace transformation of; $f(t)$ = function of the real variable t; $F(s)$ = function of the complex variable s.

The Laplace transformation is defined by the following equation, which is given without proof:

$$\mathcal{L}[f(t)] = F(s) = \int_0^\infty \varepsilon^{-st} f(t)\, dt \tag{2.17}$$

While the integral may appear a little frightening at first, it is not too difficult to evaluate in most cases.

For example, suppose that the function of the real variable is given by the following expression:

$$f(t) = 1 \quad \text{for} \quad t > 0$$
$$f(t) = 0 \quad \text{for} \quad t \leq 0$$

The form of this function is shown in figure (2.6). It is called a *unit step function*. The Laplace *transform* of this particular function is obtained by substituting this expression for $f(t)$ into equation (2.17) for the transform. In this case

$$F(s) = \int_0^\infty \varepsilon^{-st}(1)\, dt = -\frac{\varepsilon^{-st}}{s}\Big|_0^\infty = \frac{1}{s}$$

Therefore,

$$\mathcal{L}[1] = \frac{1}{s} \tag{2.18}$$

As another example, assume that

$$f(t) = \sin \omega t \quad \text{for} \quad t > 0$$
$$f(t) = 0 \quad \text{for} \quad t \leq 0$$

TABLE 1

A TABLE OF FUNCTION-TRANSFORM PAIRS

Pair	$F(s)$	$f(t)$
1	$1/s$	1
2	$\dfrac{\omega}{s^2 + \omega^2}$	$\sin \omega t$
3	$\dfrac{s}{s^2 + \omega^2}$	$\cos \omega t$
4	$\dfrac{\gamma}{s^2 - \gamma^2}$	$\sinh \gamma t$
5	$\dfrac{s}{s^2 - \gamma^2}$	$\cosh \gamma t$
6	$1/s^2$	t
7	$1/s^n$	$\dfrac{t^{n-1}}{(n-1)!}$
8	$\dfrac{1}{s \mp \gamma}$	$\varepsilon^{\pm \gamma t}$
9	$\dfrac{\beta}{(s + \alpha)^2 + \beta^2}$	$\varepsilon^{-\alpha t} \sin \beta t$
10	$\dfrac{s + \alpha}{(s + \alpha)^2 + \beta^2}$	$\varepsilon^{-\alpha t} \cos \beta t$
11	$\dfrac{1}{(s + \alpha)^2}$	$t\varepsilon^{-\alpha t}$
12	$\dfrac{1}{(s + \alpha)^n}$	$\dfrac{t^{n-1}}{(n-1)!}\,\varepsilon^{-\alpha t}$
13	$\dfrac{1}{(s - \alpha)(s - \gamma)}$	$\dfrac{1}{\alpha - \gamma}(\varepsilon^{-\alpha t} - \varepsilon^{-\gamma t})$
14	$\dfrac{1}{s(s^2 + \alpha^2)}$	$\dfrac{1}{\alpha^2}(1 - \cos \alpha t)$

Then the Laplace transform of this function is

$$F(s) = \int_0^\infty \varepsilon^{-st} \sin \omega t \, dt$$

$$= -\varepsilon^{-st} \left(\frac{s \sin \omega t + \omega \cos \omega t}{s^2 + \omega^2} \right)_0^\infty$$

Substitute limits and the result is

$$\mathcal{L}[\sin \omega t] = \frac{\omega}{s^2 + \omega^2} \tag{2.19}$$

The original function $f(t)$, together with its transform $F(s)$, form a *function-transform pair*. It is clear that after a few transformations have been worked out, using the procedure illustrated, a table of function-transform pairs can be constructed. Table 1 is representative of this type of data summation. Extensive tables of function-transform pairs can be found elsewhere.[3] Once such tables are available, only unusual functions need to be transformed; most conventional types will appear in the table.

2.7. Operation-Transform Pairs

The great advantage of the Laplace transformation is not just the ability to transform functions, but to simplify mathematical operations. It converts problems in linear integral, differential, and difference equations into ordinary algebraic problems. The principle involved should be clear from the following derivations.

Assume an arbitrary function of a real variable, call it $f(t)$. It is assumed that it can be transformed so that

$$F(s) = \mathcal{L}[f(t)]$$

or

$$F(s) = \int_0^\infty \varepsilon^{-st} f(t) \, dt \tag{2.20}$$

This integral is the product of two functions of time, and the evaluation must be performed by integrating by parts, using the general form

$$\int u \, dv = uv - \int v \, du \tag{2.21}$$

So from equation (2.20), let

$$u = \varepsilon^{-st} \quad \text{so that} \quad du = -s\varepsilon^{-st} \, dt$$

$$dv = f(t) \, dt \quad \text{so that} \quad v = \int f(t) \, dt$$

[3] See for example, *ibid.*, pp. 338–356; also *Transformation Calculus and Electrical Transients*, S. L. Goldman, Prentice-Hall, Inc., New York, 1946, pp. 416–423.

Substitution of these factors into equation (2.21) yields

$$F(s) = \left[\varepsilon^{-st}\int f(t)\,dt\right]_0^\infty + s\int_0^\infty \varepsilon^{-st}\left[\int f(t)\,dt\right]dt \tag{2.22}$$

Examination of the last term in this equation shows that it can be written

$$s\int_0^\infty \varepsilon^{-st}[\int f(t)\,dt]\,dt = s\mathcal{L}[\int f(t)\,dt]$$

Moreover, substitution of limits into the first term in equation (2.22) reduces it to the following form:

$$\left[\varepsilon^{-st}\int f(t)\,dt\right]_0^\infty = (0)\int^{t=0} f(t)\,dt - \int^{t=0} f(t)\,dt$$

where $\int^{t=0} f(t)\,dt =$ value of the integral at $t = 0$. This is called an *initial condition.* The first term in the foregoing equation must vanish, otherwise $f(t)$ is nontransformable. In other words, ε^{-st} must converge to zero more rapidly than $\int f(t)\,dt$ diverges as t approaches infinity. As a result of this restriction, the product term is zero.

The complete form of equation (2.22) therefore reduces to

$$F(s) = -\int^{t=0} f(t)\,dt + s\mathcal{L}[\int f(t)\,dt]$$

or

$$\mathcal{L}[\int f(t)\,dt] = \frac{F(s)}{s} + \frac{\int^{t=0} f(t)\,dt}{s} \tag{2.23}$$

This equation states that the Laplace transformation of the integral of the function $f(t)$ is equal to the Laplace transform $F(s)$ of the function $f(t)$, divided by s, plus a term accounting for the initial condition. Thus the process of integrating with respect to t has been simplified to dividing by s, plus inserting an appropriate initial condition factor.

A simplified method of differentiating can be derived in a similar manner. As before,

$$F(s) = \int_0^\infty \varepsilon^{-st} f(t)\,dt$$

This can be integrated by parts, using the following identification of terms:

$$u = f(t) \qquad \text{so that} \quad du = \frac{df(t)}{dt}\,dt$$

$$dv = \varepsilon^{-st}\,dt \quad \text{so that} \quad v = -\frac{\varepsilon^{-st}}{s}$$

Proceeding as in the previous derivation yields

$$F(s) = \frac{f(0)}{s} + \frac{\mathcal{L}\left[\frac{df(t)}{dt}\right]}{s}$$

or

$$\mathcal{L}\left[\frac{df(t)}{dt}\right] = sF(s) - f(0) \tag{2.24}$$

In other words, the Laplace transformation of the derivative of $f(t)$ is equal to s times the transform of $f(t)$ minus the initial value of $f(t)$. Thus differentiation with respect to t has been replaced by multiplication by s, with the inclusion of an appropriate initial condition term.

TABLE 2

OPERATION-TRANSFORM PAIRS

Function of real variable	Laplace transform
$f(t)$	$F(s)$
$\int f(t)\,dt$	$\frac{F(s)}{s} + \frac{\int^{t=0} f(t)\,dt}{s}$
$\frac{df(t)}{dt}$	$sF(s) - f(0)$
$\frac{d^2f(t)}{dt^2}$	$s^2F(s) - sf(0) - \frac{df(0)}{dt}$

Where the initial condition terms are specified as follows:

$$\int^{t=0} f(t)\,dt = \text{value of } \int f(t)\,dt \text{ at } t = 0$$

$$f(0) = \text{value of } f(t) \text{ at } t = 0$$

$$\frac{df(0)}{dt} = \text{value of } \frac{df(t)}{dt} \text{ at } t = 0$$

The transform of the second derivative can be evaluated in the same way and the results, together with those just derived are summarized in table 2, which is called a table of *operation-transform pairs*.

2.8. The Inverse Laplace Transformation

The process of taking the inverse Laplace transformation is simply the reversal of the process just described for finding the transform. That is, the operation involves changing the transform $F(s)$ back into the original function $f(t)$. This process is indicated symbolically:

$$\mathcal{L}^{-1}[F(s)] = f(t)$$

where the $\mathcal{L}^{-1}$ denotes the inverse Laplace transformation.

Direct analytical methods exist for evaluating the inverse transform,[4] but the operation depends upon a more complete knowledge of complex variable theory than can be conveniently assumed here. Actually, a method has already been established for finding the inverse transform of a function $F(s)$. Reference to the table of function-transform pairs shows that if a given transform $F(s)$ appears in the table, then the corresponding inverse transform $f(t)$ is also in the table. For example, according to pair number 6 in this table,

$$\mathcal{L}^{-1}\left[\frac{1}{s^2}\right] = t$$

According to pair number 5,

$$\mathcal{L}^{-1}\left[\frac{s}{s^2 - \omega^2}\right] = \cosh \omega t$$

Many other examples can be worked out in the same way.

It is clear that it would be impractical to attempt to construct a table of function-transform pairs that would include all possible transforms and their corresponding inverse transforms. However, if the more complicated forms of $F(s)$ can be factored or otherwise separated into simpler functions that are in the table, the inverse transformation can be accomplished. In the succeeding chapters of this book, great interest centers about the determination of inverse transforms. Thus, methods of simplifying complicated functions are of importance. The next article outlines one useful method for making the necessary simplification.

[4] See for example, Goldman, *op. cit.*, pp. 214–218; also Valley and Wallman, *Vacuum Tube Amplifiers*, McGraw-Hill Book Co., Inc., New York, vol. 18, Radiation Laboratory Series, 1948, pp. 31–39.

2.9. Partial Fraction Expansion

The transform *response* functions usually met in practice are of the general form of a rational fraction of two polynomials. A typical case with the polynomials factored can be written

$$F(s) = \frac{(s + a_1)(s + a_2)}{(s + b_1)(s + b_2)(s + b_3)} \tag{2.25}$$

Note that the roots of both polynomials are all different, or *distinct*. It has been shown in a number of algebra books[5] that a fraction of this type can be expanded into a series of *partial fractions*. Equation (2.25) can be written in partial fraction form as

$$F(s) = \frac{A_1}{s + b_1} + \frac{A_2}{s + b_2} + \frac{A_3}{s + b_3} \tag{2.26}$$

The values of the coefficients, A_1, A_2, and A_3, are unknown. All we have to do now is solve for the coefficients of the partial fractions.

The solutions for A_1, A_2, and A_3 are easily obtained. For example, to find A_1, multiply equations (2.25) and (2.26) by $(s + b_1)$ to obtain

$$(s + b_1)F(s) = (s + b_1)\frac{(s + a_1)(s + a_2)}{(s + b_1)(s + b_2)(s + b_3)}$$

$$= A_1\frac{s + b_1}{s + b_1} + A_2\frac{s + b_1}{s + b_2} + A_3\frac{s + b_1}{s + b_3}$$

Cancel the common $(s + b_1)$ factor where it occurs in both of these equations, so that

$$(s + b_1)F(s) = \frac{(s + a_1)(s + a_2)}{(s + b_2)(s + b_3)}$$

$$= A_1 + A_2\frac{(s + b_1)}{(s + b_2)} + A_3\frac{(s + b_1)}{(s + b_3)}$$

Now let $s = -b_1$. This makes all but two of the terms zero, so that

$$A_1 = \frac{(-b_1 + a_1)(-b_1 + a_2)}{(-b_1 + b_2)(-b_1 + b_3)}$$

Hence, A_1 has been determined in terms of the roots of the two polynomials.

Then A_2 and A_3 are calculated in the same way. To find A_2, multiply equations (2.25) and (2.26) through by $(s + b_2)$ and then let $s = -b_2$. Do the same thing for A_3 with $s = -b_3$.

[5] See for example, R. S. Burington and C. C. Torrance, *Higher Mathematics*, McGraw-Hill Book Co., Inc., New York, 1939, p. 184.

The process can be simply stated in mathematical terms. Assume the factors of the function in the denominator are all distinct. If this is so, then

$$A_n = (s + s_n)F(s)\Big|_{s=-s_n} \tag{2.27}$$

where A_n = coefficient of the partial fraction involving the $(s + s_n)$ factor of the denominator of $F(s)$.

A simple example should illustrate the use of equation (2.27). Assume that

$$F(s) = \frac{(s+1)}{(s+2)(s+3)}$$

Expand this into the partial fraction form as

$$F(s) = \frac{A_1}{s+2} + \frac{A_2}{s+3}$$

Now use equation (2.27) to find A_1.

$$A_1 = (s+2)F(s)\Big|_{s=-2} = \frac{(s+2)(s+1)}{(s+2)(s+3)}\Big|_{s=-2}$$

So $$A_1 = \frac{(s+1)}{(s+3)}\Big|_{s=-2} = \frac{(-2+1)}{(-2+3)} = -1$$

Then, in the same way,

$$A_2 = (s+3)F(s)\Big|_{s=-3} = \frac{(s+3)(s+1)}{(s+2)(s+3)}\Big|_{s=-3}$$

$$= \frac{(s+1)}{(s+2)}\Big|_{s=-3} = \frac{(-3+1)}{(-3+2)} = 2$$

Both coefficients are now known.

The procedure just outlined works as well for complex factors as it does for real ones. For example, let

$$F(s) = \frac{s}{s^2+\omega^2} = \frac{s}{(s+j\omega)(s-j\omega)}$$

Expand the function into its partial fraction form.

$$F(s) = \frac{A_1}{(s+j\omega)} + \frac{A_2}{(s-j\omega)}$$

Now use equation (2.27) to find A_1.

$$A_1 = (s+j\omega)F(s)\Big|_{s=-j\omega} = \frac{s}{s-j\omega}\Big|_{s=-j\omega} = \frac{-j\omega}{-2j\omega} = \frac{1}{2}$$

In the same way you can show that $A_2 = 1/2$.

Although we have not proved it, it is always true that if there are complex conjugates in the factors of the polynomial in the denominator of $F(s)$, the coefficients of the partial fractions containing these factors will also be complex conjugates.

A second case exists when a repeated factor is present in the denominator of $F(s)$. That is, if the function has a factor of order 2 or more, a different method must be used to find the coefficients in the partial fraction expansion. For example, assume that

$$F(s) = \frac{s+1}{s(s+2)^2}$$

The $(s+2)$ factor is repeated twice. This is expanded into partial fractions as

$$F(s) = \frac{A_1}{(s+2)} + \frac{A_2}{(s+2)^2} + \frac{B}{s}$$

Then B, the coefficient of the *nonrepeated* factor, is evaluated by using equation (2.27) as in the preceding examples. However, if this technique is applied to the repeated factor, indeterminate forms such as 0/0 will result.

To evaluate the coefficients of the terms involving the repeated factor, first multiply the function $F(s)$ through by the repeated factor. In the example, this would require that $F(s)$ be multiplied by $(s+2)^2$. This results in a new function that we shall call $P(s)$. That is,

$$P(s) = (s+2)^2 F(s)$$

in this example. The coefficients of terms involving the repeated factor are then computed from the following formula,[6] which is stated without proof:

$$A_{n-r} = \frac{1}{r!} \cdot \frac{d^r}{ds^r} P(s) \bigg|_{s=-a} \qquad (2.28)$$

where $-a =$ value of the repeated root (-2 in example); $r = 0, 1, 2, 3, \ldots (n-1)$.

The application of this formula to the assumed problem will illustrate the technique of use. A second-order factor is present, so that $n = 2$ and $r = 0$, $r = 1$. Hence, to compute A_2, let $r = 0$, $n = 2$, and $a = 2$ in formula (2.28). Thus,

$$A_{2-0} = A_2 = \frac{1}{0!} \cdot \frac{d^0}{ds^0} \left(\frac{s+1}{s}\right)_{s=-2} = \left(\frac{s+1}{s}\right)_{s=-2} = \frac{1}{2}$$

[6] Ruel V. Churchill, *Modern Operational Calculus in Engineering*, McGraw-Hill Book Co., Inc., New York, 1944, pp. 48–49.

In the same way for A_1, let $n = 2$, $r = 1$, and $a = 2$, so that

$$A_{2-1} = A_1 = \frac{1}{1!}\cdot\frac{d}{ds}\left(\frac{s+1}{s}\right)_{s=-2} = \left[\frac{s-(s+1)}{s^2}\right]_{s=-2} = \frac{1}{4}$$

It is clear that the procedures outlined in this article permit relatively complex transform functions to be expanded into a series of simpler functions that should appear in the table of function-transform pairs.

2.10. Use of Partial Fractions in Inverse Transformation

The partial fraction expansion is a method of easily determining inverse Laplace transforms. An example will serve to clear up most of the loose ends remaining. For example, assume that we want to find the inverse transform $f(t)$ of

$$F(s) = \frac{2}{s^4 + 10s^3 + 26s^2 + 48s + 45} \tag{2.29}$$

Fortunately, this function can be factored as follows:

$$F(s) = \frac{2}{[s+3]^2[(s+1)^2+4]} \tag{2.30}$$

This transform does not appear in the table of function-transform pairs; it will have to be simplified before taking the inverse transformation.

The quadratic factor can be divided into two subfactors, so that we can write the transform

$$F(s) = \frac{2}{(s+3)^2(s+1+2j)(s+1-2j)}$$

We can now expand this into a series of partial fractions as

$$F(s) = \frac{A_1}{(s+3)} + \frac{A_2}{(s+3)^2} + \frac{B}{(s+1+2j)} + \frac{B'}{(s+1-2j)}$$

where $B' =$ complex conjugate of B.

Equation (2.28) is used to find A_2 and A_1. After making the appropriate substitutions this gives

$$A_2 = (s+3)^2F(s)\Big|_{s=-3} = \frac{2}{(s+1)^2+4}\Big|_{s=-3}$$

so that $\quad A_2 = \dfrac{1}{4}; \qquad A_1 = \dfrac{d}{ds}\left[\dfrac{2}{(s+1)^2+4}\right]_{s=-3} = \dfrac{1}{8}$

Because B and B' are complex conjugates, only one of them need be computed. Thus, using formula (2.27) for distinct factors,

$$B = (s + 1 + 2j)F(s)\,\Big|_{s=-1-2j}$$

$$= \frac{2}{(s+3)^2(s+1-2j)}\Bigg|_{s=-1-2j} = -\frac{1}{16}$$

Therefore $B' = -1/16$.

The complete expression for the expanded transform can now be written

$$F(s) = \frac{1}{4}\left(\frac{1}{s+3}\right)^2 + \frac{1}{8}\left(\frac{1}{s+3}\right) - \frac{1}{16}\left(\frac{1}{s+1+2j}\right) - \frac{1}{16}\left(\frac{1}{s+1-2j}\right)$$

Now the inverse transforms are taken term-by-term using pairs 8 and 11 from the table of function-transform pairs. Therefore, the complete inverse transform is

$$f(t) = \frac{t\varepsilon^{-3t}}{4} + \frac{\varepsilon^{-3t}}{8} - \frac{1}{16}(\varepsilon^{-t-2jt} + \varepsilon^{-t+2jt}) \tag{2.31}$$

$$f(t) = \frac{t\varepsilon^{-3t}}{4} + \frac{\varepsilon^{-3t}}{8} - \frac{1}{16}\,\varepsilon^{-t}(\varepsilon^{-2jt} + \varepsilon^{+2jt}) \tag{2.32}$$

The last term in parentheses is the exponential form of the cos $2t$. Therefore

$$f(t) = \frac{t\varepsilon^{-3t}}{4} + \frac{\varepsilon^{-3t}}{8} - \frac{1}{8}\,\varepsilon^{-t}\cos 2t \tag{2.33}$$

This is the complete inverse transform of $F(s)$.

2.11. A Sample Problem on Electric Circuits

All the material presented so far can be conveniently summarized by a typical example taken from electric circuit theory. Such a move also points the way to some additional useful ideas.

Consider the simple *RLC* circuit of figure (2.7). It is assumed that both e and i are unspecified functions of time, so that they should be written as $e(t)$ and $i(t)$. The switch is assumed to open at $t = 0$. The circuit loop equation is

$$e(t) = L\frac{di(t)}{dt} + Ri(t) + \frac{1}{C}\int i(t)\,dt \tag{2.34}$$

Now define the following terms: $E(s) =$ Laplace transform of $e(t)$; $I(s) =$ Laplace transform of $i(t)$; $\alpha = i(0) =$ initial current through L;

$q = \int i(t = 0)\, dt$ = initial charge on the capacitor; $\gamma = q/C$ = initial voltage on the capacitor.

The transforms of the individual terms in equation (2.34) can be written from the table of operation-transform pairs. The result is

$$E(s) = L(sI - \alpha) + RI + \frac{I}{sC} + \frac{\gamma}{s} \tag{2.35}$$

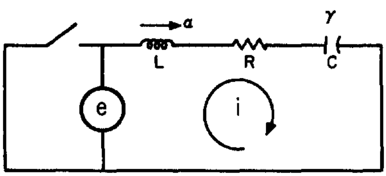


Fig. 2.7. Series *RLC* circuit.

where the argument s has been omitted from $I(s)$ for simplicity. With terms rearranged, this becomes

$$E(s) = I\left(sL + R + \frac{1}{sC}\right) - L\alpha + \frac{\gamma}{s} \tag{2.36}$$

Now solve for the transform current I.

$$I(s) = I = \frac{E(s) + (L\alpha - \gamma/s)}{sL + R + 1/sC} \tag{2.37}$$

The form of equation (2.37) is general and should strike a responsive chord. It has the form

$$\begin{pmatrix}\text{response}\\ \text{function}\end{pmatrix} = \frac{\text{excitation function}}{\text{characteristic function}} \tag{2.38}$$

It is clear from equation (2.37) that the characteristic function is dependent only upon the characteristics and constants of the circuit and is not affected by the nature of the excitation. Moreover, the characteristic function is simply the total impedance of the circuit evaluated in terms of the complex frequency s.

The excitation function involves two terms that are defined as: $E(s)$ = driving function; $(L\alpha - \gamma/s)$ = initial excitation function. The *driving function* is that excitation supplied by active circuit elements, while the *initial excitation* is that caused by energy storage in the passive circuit elements. A great many problems can be solved by neglecting the initial excitation function. All the circuits in Part II of this book are treated in this way. This is a permissible procedure because the superposition principle applies to the linear circuits covered.

The omission of the initial excitation terms simplifies the treatment of many problems. In particular, the use of the Laplace transformation becomes no more involved than ordinary a-c circuit theory. That is,

$$\mathcal{L}\left[L\frac{di}{dt}\right] = (sL)I; \qquad \mathcal{L}\,[Ri] = (R)I; \qquad \mathcal{L}\left[\frac{1}{C}\int i\,dt\right] = \left(\frac{1}{sC}\right)I$$

where sL and $1/sC$ are the impedances associated with complex frequency. Thus, as long as initial conditions can be neglected, circuit equations are written as in a-c circuit theory and as shown in section (2.4), simply replacing $j\omega$ by s. These equations will always end up in the form of

$$\text{response transform} = \frac{\text{driving transform}}{\text{characteristic transform}}$$

The response as a function of time, called the *transient* response, is computed by taking the inverse transform of the response transform by the methods that have been outlined.

It should be understood that the characteristic transform function may be any one of several types, such as impedance, admittance, or a dimensionless ratio.

2.12. Complex S Plane; Poles and Zeros

Complex numbers, their use, and their representation in a complex plane are all familiar ideas to electrical engineers. For example, the impedances and admittances of a-c circuit theory are complex quantities of the general form

$$Z = R + jX \quad \text{or} \quad Y = G + jB$$

These quantities have *real* parts R and G, and *imaginary* parts X and B. By constructing a complex Z or Y plane, such numbers can be represented as shown in figure (2.8). Resistance or conductance is plotted along the real axis, while reactance or susceptance is plotted along the imaginary axis.

It was shown in section (2.4) that the generalized frequency s is a complex number of the form $s = \sigma + j\omega$. By analogy to the technique used for a-c impedances and admittances, it follows that s can be represented graphically in a complex *s plane* with σ along the axis of reals and ω on the imaginary axis. This is illustrated in figure (2.9). Any combination of values for σ and ω uniquely determine a specific point on the complex s plane.

It has been shown that the response transforms of physical systems always have the general form

$$\text{response transform} = \frac{\text{excitation transform}}{\text{characteristic transform}}$$

Regardless of the nature of the problem, the transforms of the excitation and characteristic functions are always polynomials in s. In *practical* cases the degree of the polynomial in the numerator of the response transform is always less than that in the denominator. Thus a general response transform has the form

$$\text{response transform} = K\frac{a_1s^{n-1} + a_2s^{n-2} + a_3s^{n-3} + \ldots a_n}{s^n + b_1s^{n-1} + b_2s^{n-2} + b_3s^{n-3} \ldots b_n}$$

where $K =$ scale factor $=$ constant, independent of s.

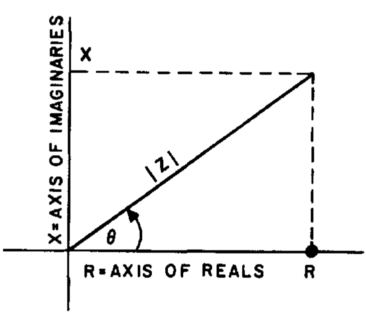


Fig. 2.8. Complex Z plane for representing a-c impedances.

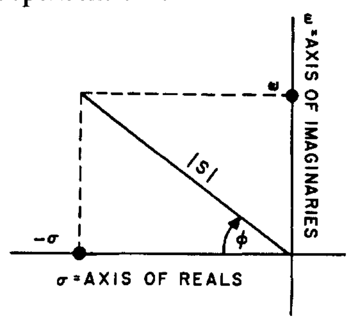


Fig. 2.9. Complex s plane for representing complex frequency.

Now consider the case of a response transform of a given system in which the polynomials in the numerator and denominator can be factored as shown in equation (2.39).

$$F(s) = K\frac{(s + \gamma_1)(s + \gamma_2)}{(s + \alpha_1)(s + \alpha_2)(s + \alpha_3)} \tag{2.39}$$

The five values of s, denoted α_1, α_2, α_3, γ_1, and γ_2 are unique. For example, if s is equal to either $-\gamma_1$ or $-\gamma_2$, the numerator becomes zero and the whole response transform is zero. Hence, $-\gamma_1$ and $-\gamma_2$ are called the *zeros of the response transform* $F(s)$. Because they are unique values of s with specific values of σ and ω, they can be located in the complex s plane. Zeros are usually denoted by small circles as shown in figure (2.10).

When s is equal to $-\alpha_1$, $-\alpha_2$, or $-\alpha_3$, the denominator of the response transform is zero, so that the over-all function is infinite. Values of s

that make the characteristic transform zero, or the response transform infinite, are called the *poles* of $F(s)$.[7] They are indicated on the complex s plane by small crosses as shown in figure (2.10).

In some cases the poles and zeros are of higher order than the *first-order* types illustrated in the foregoing example. The response transform might be

$$F(s) = K\frac{(s+\gamma)^m}{(s+\alpha)^n}$$

This function has an mth order zero at $-\gamma$ and an nth order pole at $-\alpha$.

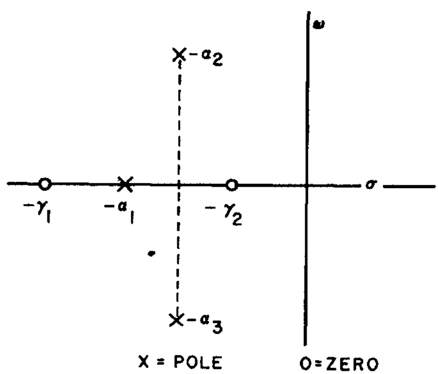


Fig. 2.10. Poles and zeros in the complex s plane.

The poles and zeros are unique values of s, and as such, uniquely determine the response characteristics of the physical system. Some aspects of this idea are covered in later sections.

2.13. Transient Response Deduced from Pole Locations

It has been shown that the response transform of a given physical system always has the form of a rational ratio of polynomials in s, where the polynomials can be factored. Such a transform might appear as

$$F(s) = K\frac{(s+\gamma)}{(s+\alpha_1)(s+\alpha_2)(s+\alpha_3)}$$

This function has a zero at $-\gamma$ and three poles at $-\alpha_1$, $-\alpha_2$, and $-\alpha_3$.

[7] These are only the simplest kind of poles, and the definition given is not precise in all cases. While this definition is sufficient for the purposes of this book, more advanced readers desiring more information should refer to R. V. Churchill, *Modern Operational Mathematics in Engineering*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1944, pp. 139–144.

The function can be expanded into its partial fractions as

$$F(s) = \frac{A_1}{s + \alpha_1} + \frac{A_2}{s + \alpha_2} + \frac{A_3}{s + \alpha_3}$$

where the A_1, A_2, and A_3 are constants that can be evaluated by standard methods. It is important to note that the *poles* retain their identity through this expansion, whereas the zeros are absorbed in the coefficients. The values of the three coefficients depend upon both the zeros and the poles.

The transient response, which is the inverse transform of $F(s)$, is easily written from the table of function-transform pairs as

$$f(t) = A_1\varepsilon^{-\alpha_1 t} + A_2\varepsilon^{-\alpha_2 t} + A_3\varepsilon^{-\alpha_3 t}$$

It is important to note that the *nature* of the transient response depends only upon the nature of the poles because only the poles affect the form of the time dependent factors in the transient response. Of course, here there are three distinct poles and each contributes a simple exponential factor to the transient response. The constants multiplying the exponentials have *not* been specified, so it is not possible to say which exponential is the largest or smallest. Nevertheless, there are many cases in which it is necessary only to know the *nature* of the response without regard to the relative magnitudes of the component response functions. In such cases it is necessary only to examine the poles of the response transform and their location in the complex s plane, and the nature of the transient response can be quickly deduced from the table of function-transform pairs.

This idea is illustrated in figure (2.11). This figure and the table of function-transform pairs should be studied at the same time. The point being made reduces to the following:

(1) Any response transform can be expanded into a sum of transforms of types found in the table of function-transform pairs. This is accomplished by a partial fraction expansion.

(2) Each such partial transform in one expansion involves one or moie poles of the original response transform.

(3) Thus each such pole combination contributes a single term to the inverse transform, or time response.

(4) There are many possibilities for the individual partial transforms; those shown in figure (2.11) are typical. A little experience with pole diagrams and some familiarity with the table of function-transform pairs will enable you to estimate quickly the character of the transient response.

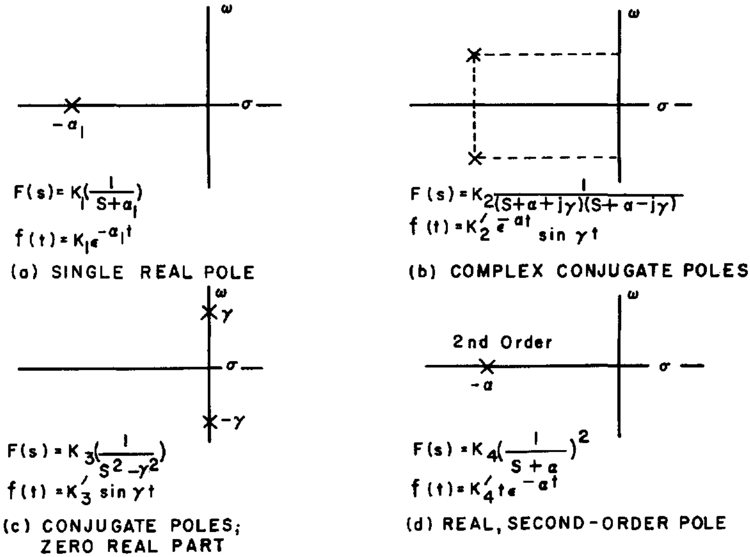


Fig. 2.11. Relationship between pole location and transient response.

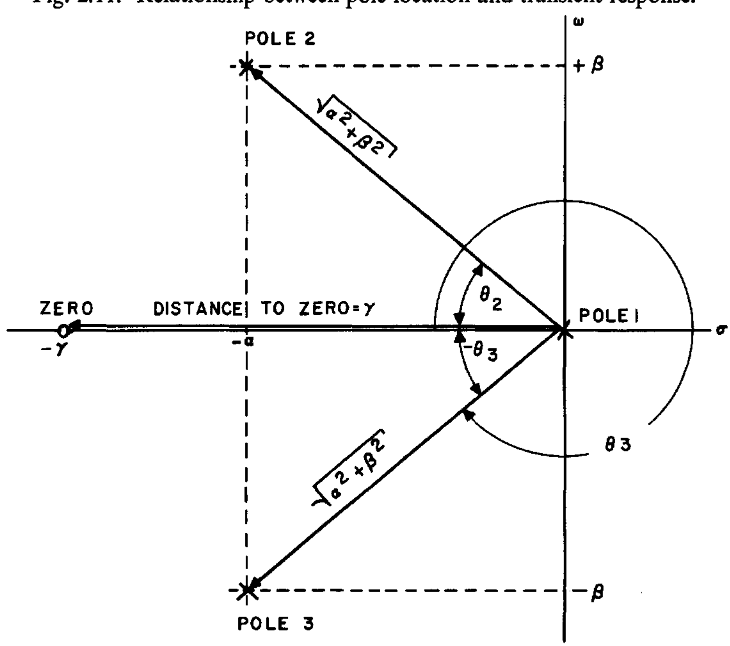


Fig. 2.12. Method of evaluating the partial fraction coefficient associated with pole 1, which is at the origin in this example.

An important observation that has considerable use later is the fact that the response is non-oscillatory as long as all poles are real. Oscillation results whenever complex conjugate poles appear.

The partial fraction coefficients are easily computed from the pole-zero diagram for any function of s. Thus the entire character of the transient response can be determined. For example, consider the following function:

$$F(s) = \frac{s + \gamma}{s(s + \alpha + j\beta)(s + \alpha - j\beta)}$$
$$= \frac{A}{s} + \frac{B}{s + \alpha + j\beta} + \frac{B'}{s + \alpha - j\beta}$$

The corresponding pole-zero diagram is given in figure (2.12).

The partial fraction coefficients are easily evaluated by the method previously given and expressed as follows:

$$A = \frac{\gamma}{(\alpha + j\beta)(\alpha - j\beta)}$$
$$= \frac{\gamma \underline{/0^\circ}}{\sqrt{\alpha^2 + \beta^2}\sqrt{\alpha^2 + \beta^2} \underline{/\theta_2 + \theta_3}}$$
$$B = \frac{(\gamma - \alpha) - j\beta}{(-\alpha - j\beta)(-2j\beta)}$$
$$B' = \frac{(\gamma - \alpha) + j\beta}{(-\alpha + j\beta)(+2j\beta)}$$

To evaluate A, draw lines from the pole at $s = 0$ to all other poles and zeros as shown in figure (2.12). The coefficient A is associated with the pole at $s = 0$ in the partial fraction expansion. You will observe from figure (2.12) that:

(1) The line drawn to the zero at $-\gamma$ has a length γ and this is equal to the numerator in the equation for A.

(2) The lines drawn to the conjugate poles have lengths $\sqrt{\alpha^2 + \beta^2}$ and angles θ_2 and θ_3. These are the components in the denominator of the equation for A expressed in polar form. Hence the magnitude and angle of the partial fraction coefficient A associated with the pole at $s = 0$ is

$$A = \frac{\text{product of line lengths from this pole to all zeros}}{\text{product of line lengths from this pole to all other poles}}$$
$$\theta_A = \sum \text{(angles of lines to zeros)} - \sum \text{(angles of lines to poles)}$$

The same technique is used to compute B and B' except that the lines are drawn from the pole associated with the coefficient being evaluated. Thus, to evaluate B, lines are drawn from the pole at $(-\alpha - j\beta)$ to the zero at $-\gamma$ and the two poles at $(-\alpha + j\beta)$ and 0. Then B is computed in polar form from the formula previously given.

2.14. Determination of the Steady State Response

Steady state response as used here refers to the network response to constant amplitude, constant frequency sinusoidal excitation functions, the signals of a-c circuit theory. It is customary to represent the steady

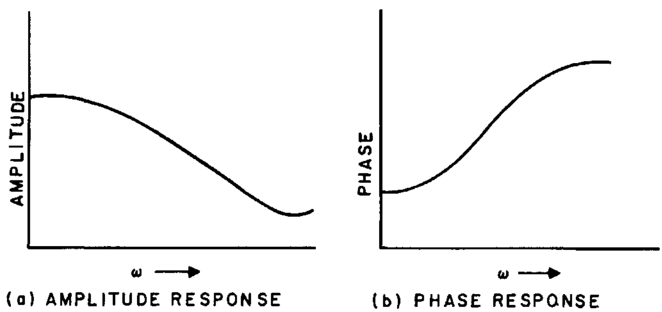
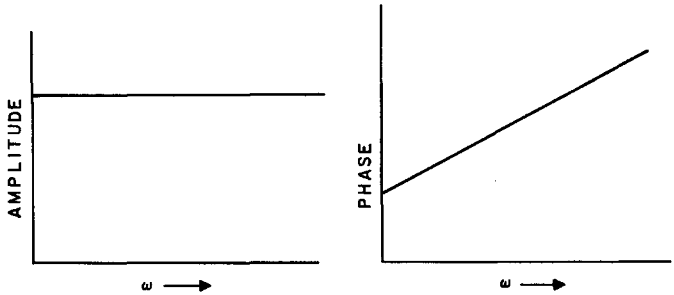


Fig. 2.13. Representative frequency response characteristics.

state response to such excitation by means of two graphs: (1) *amplitude characteristics*, a plot of the absolute value of the transfer function vs. ω or frequency; (2) *phase characteristic*, phase shift of the transfer function *vs*, ω or frequency. These two graphs are called the *frequency response* characteristics of the transfer function. Typical examples of such response characteristics are shown in figure (2.13).

In some special cases the amplitude and phase characteristics are straight lines as shown in figure (2.14). When these conditions exist, the signal is transmitted through the network in such a way that it is an undistorted replica of the input, but delayed in time. Consequently, these frequency response characteristics are the conditions for distortionless transmission.

If the amplitude response deviates from a horizontal straight line, the conditions for distortionless transmission are violated and the output will not have exactly the same shape as the input. Hence the network introduces *amplitude distortion*. If the phase characteristic is nonlinear, *phase distortion* results. Thus the frequency response characteristics of a transfer function can be useful in judging the extent to which distortionless transmission is violated. They provide convenient criteria for comparing the relative suitability of various networks for use in a given application. Because of this usefulness and corresponding



(a) AMPLITUDE RESPONSE (b) PHASE RESPONSE

Fig. 2.14. Frequency response for distortionless transmission.

importance, the balance of this chapter is devoted to a summary of a simple method of computing the frequency response characteristics from the pole-zero diagrams of transfer functions.

It will be recalled that the transfer function of any physical system could be expressed as

$$\text{transfer function} = \frac{\text{response function}}{\text{excitation function}}$$

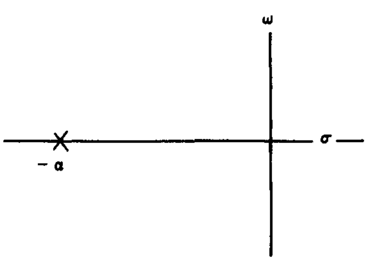


Fig. 2.15. Single pole in the s plane.

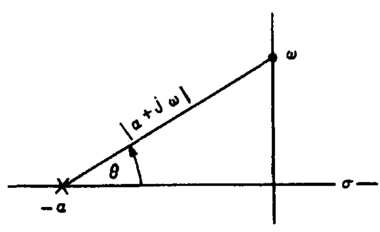


Fig. 2.16. Determination of amplitude and phase angle of $F(j\omega)$.

When expressed in terms of complex frequency this was shown to have the form of a ratio of polynomials multiplied by a scale factor. A typical example would be

$$F(s) = K\frac{1}{s+\alpha} \tag{2.40}$$

This is a simple case in which the transform of the transfer function has a single, first-order pole and no zero. The pole is a real number, which is a necessary condition for a one-pole system. Hence, the pole is located in the complex s plane as shown in figure (2.15).

Suppose that the steady state amplitude and phase response characteristics are to be determined. In this case ω is the independent variable. In practical problems ω can never be negative. As a result, ω can vary along the imaginary axis of the s plane from zero to positive infinity. Therefore the s in equation (2.40) should be replaced by $j\omega$, and the transform of the transfer function becomes

$$F(j\omega) = \frac{K}{(\alpha + j\omega)} \tag{2.41}$$

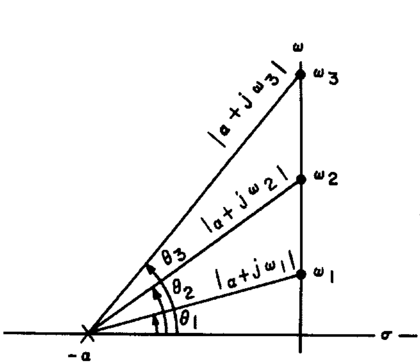


Fig. 2.17. Determination of steady state response.

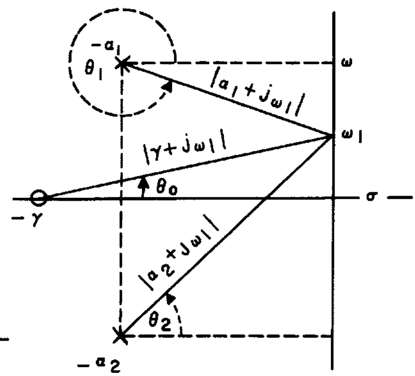


Fig. 2.18. Two pole, one zero case; evaluation of frequency response.

The function can be written in polar form as

$$F(j\omega) = K \left|\frac{1}{\alpha + j\omega}\right| \frac{1}{\underline{/\theta}} \tag{2.42}$$

The magnitude factor $|\alpha + j\omega|$ is simply the length of the line drawn from the pole at $-\alpha$ to the assumed value for ω located on the positive imaginary axis. The angle θ is the angle that this line makes with a line drawn parallel to the real axis. These statements are illustrated in figure (2.16).

It is evident that the data needed to plot the magnitude and phase of $F(j\omega)$ as a function of the frequency ω can be obtained directly by following the procedure shown in figure (2.16) for a series of different frequencies. This process is illustrated in figure (2.17).

In the one-pole case just outlined the method has little advantage to offer over direct numerical calculation. However, consider a more complicated case in which the transform of the transfer function has two poles and one zero. That is, let

$$F(s) = K\frac{(s+\gamma)}{(s+\alpha_1)(s+\alpha_2)}$$

In the steady state case this becomes

$$F(j\omega) = K\frac{|\gamma + j\omega|\,/\underline{\theta_0}}{|\alpha + j\omega|\,|\alpha + j\omega|\,/\underline{\theta_1 + \theta_2}} \tag{2.43}$$

Polar form has been used for the complex numbers. The frequency characteristics of a function of this type are relatively difficult to evaluate by direct numerical substitution. However, it is a relatively easy process using the graphical method in the complex s plane. The procedure involves only evaluating the magnitude and angle factor separately for each assumed frequency and then substituting these into equation (2.43). A typical case is shown in figure (2.18). A protractor and a pair of dividers, plus a little elementary slide rule work are all that is required to compute the desired characteristics.

A general formula can be written to describe this method of determining the steady state characteristics. The form is clearly

$$|F(j\omega)| = K\frac{\text{product of line lengths from zeros to }\omega}{\text{product of line lengths from poles to }\omega}$$
$$= K\frac{\text{product of the zero distances}}{\text{product of the pole distances}}$$

and the angle function is

$$\theta(j\omega) = (\text{sum of the zero angles}) - (\text{sum of pole angles})$$

2.15. Physically Realizable and Minimum Phase Shift Circuits

It was shown in section (2.4) that voltages and currents involving a positive real part of the complex frequency s exhibit the properties of a continuously increasing sinusoid. It is clear that such a response would be impossible in practice because practical circuit elements are not linear for all currents and voltages. Eventually, if the current is continuously increasing, the circuit elements will exhibit a nonlinearity that will prevent further increases in the current. This nonlinearity might be caused by a burnout, vacuum tube overload, or resistance change from overheating. In any case, a stable network can never permit the continued existence of an ever increasing current. Thus the transfer function of a *physically realizable* circuit cannot have a pole in the right half of the complex s plane.

Networks used in electronic circuits are often of the *minimum phase shift* type. A minimum phase shift circuit is defined as *the* circuit which has the least phase shift of all circuits which have its particular amplitude characteristic.[8] The transfer functions of minimum phase shift circuits cannot have zeros in the right half of the complex s plane. You can prove this to yourself pretty easily by working out some sample cases.

PROBLEMS

2.1. Derive the equations for the mutual impedances, as functions of the complex frequency s, for each of the networks shown in figure (2.19). In each case the mutual impedance is considered to be the ratio of the output voltage to the input current. It will be shown in a later chapter that these are the equivalent plate circits of a number of important vacuum tube amplifiers.

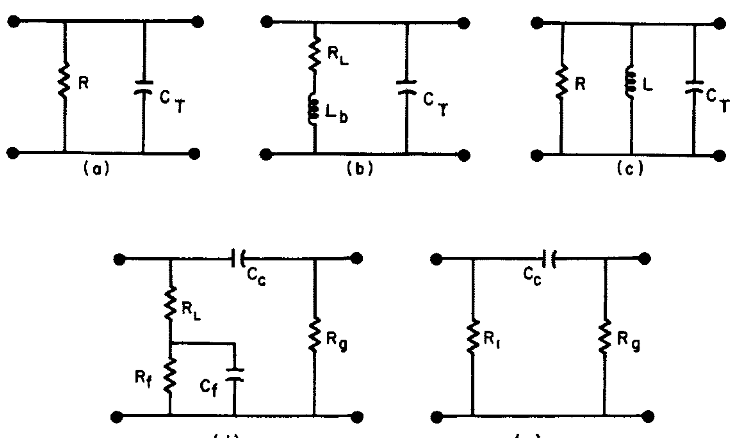


Fig. 2.19.

2.2. For each of the mutual impedance functions derived in problem (2.1), make sketches showing the relative locations of the poles and zeros in the complex s plane. If more than one possibility exists, indicate each possibility by a separate pole-zero diagram.

2.3. From the impedance functions of problem (2.1) and the pole-zero diagrams of problem (2.2), discuss the transient response characteristics of the various networks.

[8] See H. W. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand Company, Inc., New York, 1945, Art. 11.5.

2.4. For circuit b of figure (2.19), assume that

$$R_L = 7000 \text{ ohms}; \qquad L_b = 500\ \mu\text{h}; \qquad C_T = 30\ \mu\mu\text{f}$$

Using the method outlined in section (2.14), determine the steady state response characteristics of the circuit over a frequency range from $\omega = 0$ to $\omega = 14 \times 10^6$ rps. Plot your results on semilog paper with ω on the logarithmic scale.

2.5. Derive the Laplace transform of the function $f(t) = t$. See pair number 6 in the table.

2.6. Derive the Laplace transform of $f(t) = \varepsilon^{-\alpha t}$. See pair number 9 in the table.

2.7. Using partial fractions, obtain the inverse transforms of

$$(a) \quad F(s) = \frac{1}{(s^2 + \alpha^2)s^3} \qquad (b) \quad F(s) = \frac{(s^2 - \alpha^2)}{(s^2 + \alpha^2)^2}$$

Evaluate all the partial fraction coefficients.

2.8. Using partial fractions, calculate the inverse transform of

$$F(s) = \frac{(s + 1)(s + 4)}{s(s + 2)^2 (s + 3)^2}$$

Calculate all the partial fraction coefficients.

Part II

CLASS A CIRCUITS

Chapter 3

PRINCIPLES OF VACUUM TUBE AMPLIFIERS

An *amplifier* is a circuit whose output is an enlarged reproduction of the input, but the power developed in the output is drawn from a source other than the signal input.

Vacuum tubes are frequently operated as *voltage* amplifiers so that the output voltage is an enlarged reproduction of the signal input voltage. The tube is generally operated in class A when serving as a voltage amplifier, and only this type of operation is covered in this sequence of chapters.

The chapter commences with a discussion of the method of determining the proper operating potentials on the tube required for class A operation. Then the class A equivalent circuit developed in chapter 1 is used to derive the general theory of vacuum tube voltage amplifiers.

It will be shown that there are three basically different vacuum tube amplifier circuits: (1) grounded cathode; (2) grounded plate; (3) grounded grid. In all three cases the factors of primary interest are (1) voltage amplification; (2) output impedance; (3) input impedance or admittance.

It will be shown that the voltage amplification equation for *any* vacuum tube amplifier can be written

$$A = \pm g'_m Z'_m = \frac{\text{output voltage}}{\text{input voltage}}$$

where g'_m = effective transconductance of the circuit and tube; Z_m = mutual impedance of the passive circuit.

Expressions will also be derived for the input and output impedances of the three basic amplifier circuits.

3.1. Direct Current Circuit Connections

A vacuum tube will not function properly unless there is a complete circuit for direct current running from the plate circuit power supply, through any impedance connected in series with the plate, through the tube to the cathode, and through any cathode impedance to ground. The negative terminal of the plate power supply, which is designated E_{bb}, provides the d-c ground. This requirement is indicated schematically in figure (3.1).

It is necessary to have also a d-c path between the grid of the tube and the source of constant grid voltage. This is also shown in figure (3.1).

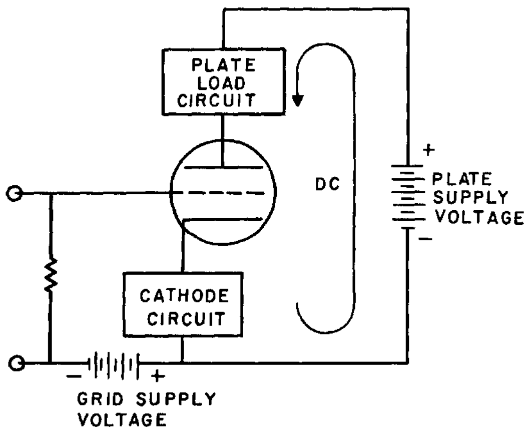


Fig. 3.1. Essential connections for a vacuum tube amplifier.

Other requirements also exist, though their importance will not be established until later. It will be shown in section (3.3) that the impedance in the plate circuit, whether in series with the plate or cathode lead of the tube, cannot be zero at the frequency of the input signal. If it is zero, the circuit will not amplify.

Finally, in most cases only the signal frequency component of the output from an amplifier is of interest and some method must usually be provided for removing any d-c component that may exist. This is mainly important if the amplifier output is fed to the input of another amplifier. If the d-c component is not removed from the output, the grid voltage on the succeeding amplifier stage will be altered and improper operation could result.

Four general requirements that must be met by nearly all amplifier circuits have been listed. The necessity of removing the d-c component from the output is the only one of the four that can *ever* be violated.

3.2. Load Lines and Operating Points

The circuit connections for one type of elementary amplifier are shown in figure (3.2). The nomenclature used is standard, with terms specified as follows:

E_{cc} = d-c grid voltage = grid bias
E_{bb} = plate supply voltage
R_L = plate load resistor
i_b = plate current
e_g = grid signal voltage
e_c = $e_g - E_{cc}$ = grid-to-cathode voltage = grid voltage
e_b = plate-to-cathode voltage = plate voltage

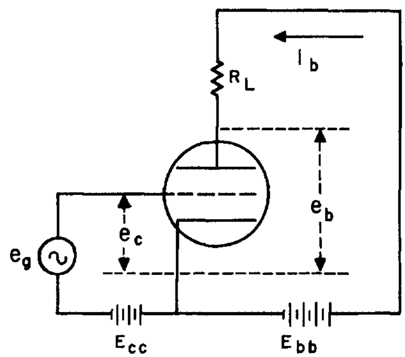


Fig. 3.2. Elementary vacuum tube amplifier.

This circuit meets all the requirements specified in the preceding section except that no provision is made for removing the d-c component from the output.

The Kirchhoff loop equation for the plate circuit can be written directly from figure (3.2) as

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

or, in an alternative form,

$$e_b = E_{bb} - i_b R_L \qquad (3.1)$$

By assuming values for e_b and calculating i_b from equation (3.1), a graph of i_b vs. e_b for given values of E_{bb} and R_L is easily constructed. The resulting graphical form of equation (3.1) is called the *load line* and is shown in figure (3.3). The terminology is appropriate because the slope is determined by the load resistance R_L.

This graphical construction would be sufficient to determine the plate current i_b corresponding to any given plate voltage e_b if the equation were complete as it stood. However, the circuit contains a vacuum triode, and the plate current is a complex function of the plate and grid voltages. Thus, to obtain a solution for the plate current through the tube corresponding to particular values of plate voltage e_b, power supply voltage E_{bb}, load resistance R_L, and grid voltage e_c, it is necessary to obtain a simultaneous solution of: (1) the load line equation; (2) the equation for the plate current through the tube as a function of the grid and plate voltages. This would be a difficult thing to do analytically because of the complex form of the equation relating i_b, e_b, and e_c. However, it is relatively easy to compute the solution by graphical methods.

Typical static plate characteristics for a triode are shown in figure (3.4). Superimpose the load line graph of figure (3.3) on this plot of the tube

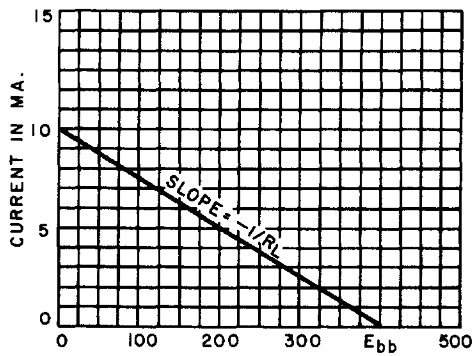


Fig. 3.3. The load line. $R_L = 40K$, $E_{bb} = 400v$ in this example.

characteristics, using the same current and voltage scales. A simultaneous solution of the two functions can be obtained from this construction.

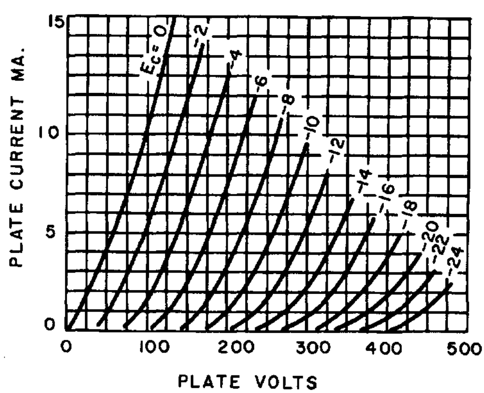


Fig. 3.4. Triode characteristics.

The actual solution, which is the *one* point at which the two equations are simultaneously satisfied, is called the *operating point*. It is determined by finding the point of intersection of the load line with the particular $i_b - e_b$ curve corresponding to a specified grid voltage.

The operating point determined in the *quiescent* state, which means that there is no grid signal voltage, is called the *Q point*. The location of the Q point is shown in figure (3.5), where it appears as the point of intersection of the load line with the tube characteristic curve for the value of grid voltage equal to the grid bias. Then the effect of a signal

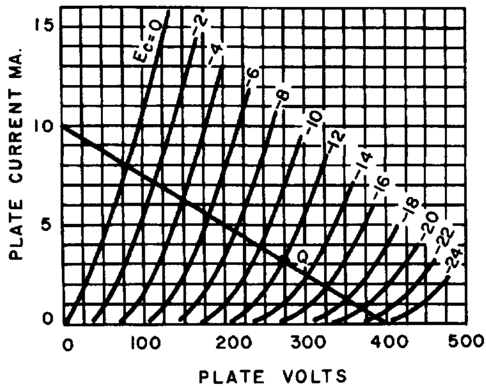


Fig. 3.5. Determination of the operating point of a triode.

applied to the grid is to cause the actual operating point to swing about the Q point and along the load line from one point of intersection to another.

3.3. Operation of an Elementary Amplifier

Vacuum tube amplification is easily understood from this graphical presentation. In figure (3.5) for example, with the value of load resistance used, a 2 volt swing in grid voltage above and below the bias E_{cc} will cause the plate voltage to swing over a range of 238 to 302 volts. If voltage amplification is defined as the ratio of the swing in output voltage to the swing in input voltage, then

$$A = \text{voltage amplification}$$
$$= \frac{302 - 238}{4} = \frac{64}{4} = 16$$

Amplification is possible only because the load line has a finite slope. If R_L is zero the load line is vertical and no amount of grid voltage variation produces any change in plate voltage. The amplification is obviously zero. As the load resistance is gradually increased, the slope of the load line decreases. As the load line becomes more nearly horizontal, a given grid voltage swing will cause progressively larger swings in plate voltage, indicating an increase in the circuit voltage amplification. In the limit, when the load line is horizontal, the voltage amplification is equal to the amplification factor μ of the tube. This condition can be brought about by making the load resistance infinite. Of course, this is impractical, because you can see that E_{bb} would have to be infinite. Thus, this is a purely theoretical limiting case. The general nature of the variation in voltage amplification as a function of the load resistance is shown in figure (3.6).

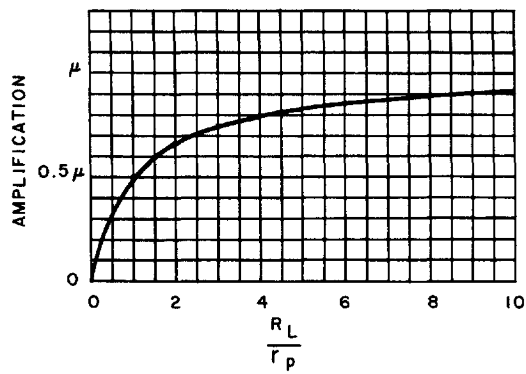


Fig. 3.6. Effect of load resistance on voltage amplification.

Amplification is accomplished through the action of the grid in controlling the flow of current through the load resistance. Variations in grid voltage change the voltage drop across the tube, causing the constant voltage source E_{bb} to supply a varying component of plate current. This in turn produces a varying component of plate voltage. The vacuum tube acts simply as a converter of energy, changing the d-c energy available from the plate supply into variational or signal energy.

From figure (3.5) it is apparent that making the grid more negative than the bias voltage causes: (1) plate current to decrease; (2) plate voltage to increase. Thus the application of a triangular grid signal will cause the plate current and voltage to vary as shown in figure (3.7). Note that the plate voltage is inverted, like a mirror image, with respect to the grid voltage. This is a fundamental property of this type of amplifier. For this reason it is often called a *phase inverter*.

The voltage amplification of the simple amplifier of figure (3.2) can be determined from a graphical analysis of the type shown in figure (3.5). However, when the plate load is anything other than a simple resistance, the graphical determination becomes inconveniently complicated.

Because nearly all practical plate load circuits are more complex than a single resistor, other methods are generally used to find the circuit amplification.

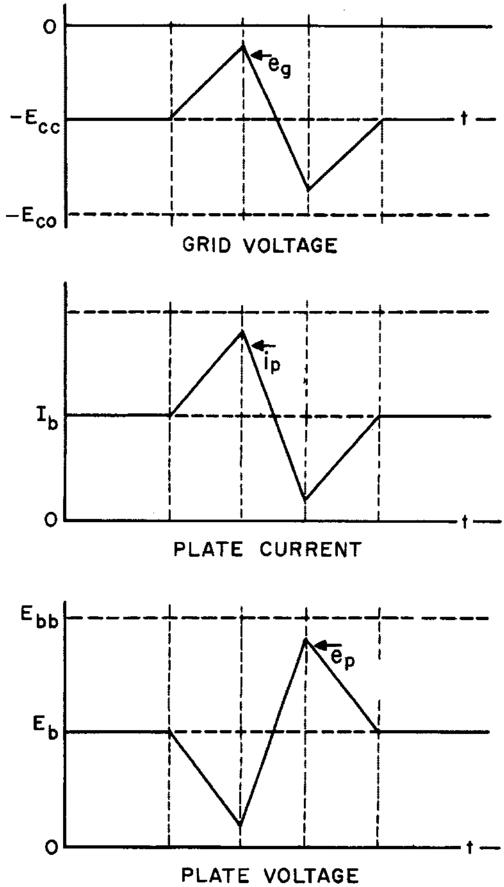


Fig. 3.7. Waveforms in an amplifier with resistance load.

3.4. Polarizing Potentials

The procedure followed in determining the Q point was shown in figure (3.5). The plate voltage at the Q point is E_b and the corresponding plate current is I_b. The two electrode voltages at the Q point, E_b and E_{cc}, are called the *polarizing potentials* of the tube. If the tube is a tetrode or pentode, the screen voltage required is also a polarizing potential.

The polarizing potentials are selected to fix the Q point to a particular place on the tube characteristics, thereby achieving certain operating characteristics from the tube. A graphical construction of the type previously shown in figure (3.5) can be used to determine the polarizing potentials required to establish operation at a specified Q point. This is necessary in some cases. However, for many cases the tube manufacturer has determined the proper polarizing potentials required for certain applications of the tube and this information is published in convenient handbook form. Ordinarily, the published values can be used without recourse to a determination of the type shown in figure (3.5), but care should be exercised to make sure that the proper polarizing potentials are used for the specific application planned for the tube.

In figure (3.2) for one type of basic triode amplifier circuit, all electrode voltages were supplied by batteries. This is not uncommon. However, the availability of a-c power and the inconvenience caused by the size, weight, and necessity of recharging and replacing batteries has led to the use of a-c power in a large proportion of electronic equipment.

For a-c operated units the plate supply voltage E_{bb} is usually supplied by a *rectifier* and *filter* circuit. These circuits are discussed in detail in chapter 14. For present purposes, the effect of the rectifier type of power supply is the same as that of a battery, so a d-c power supply for the plate circuit will be indicated without specifying how that voltage is developed.

The calculation of the value for E_{bb} is usually one of the last steps in the design of an amplifier. When this computation is made, E_b and I_b, which are the Q point plate voltage and current, are usually known, because the Q point is usually specified for the tube in use. The load resistance is determined by other amplifier design requirements and can be presumed known when E_{bb} is being calculated. Hence, the value of E_{bb} required to establish the proper Q point can be computed from the Kirchhoff loop equation in the quiescent state. That is,

$$E_{bb} = E_b + I_b R_L \tag{3.2}$$

There are many cases where an amplifier has a different load resistance for direct current than for signal frequency currents. This results in different load lines in the direct current and signal frequency cases. This problem is treated in section (11.9).

Few amplifiers have pure resistance loads; some reactive component is invariably present. This causes the plate-to-cathode voltage and plate current to be out of phase by an angle ϕ, so that

$$e_p = E_p \sin \omega t \quad \text{and} \quad i_p = I_p \sin (\omega t + \phi)$$

These are the parametric equations for an ellipse, which shows that the operating point traverses an elliptical path instead of the straight line followed for a resistance load. The major axis of the ellipse will fall on the load line drawn for a resistance equal to the magnitude of the load impedance.

3.5. Cathode and Screen Biasing Circuits

Instead of using a battery to provide the grid bias voltage E_{cc}, a parallel *RC* circuit is often placed in series with the cathode of the tube as shown in figure (3.8). This is called a *cathode bias circuit.* The

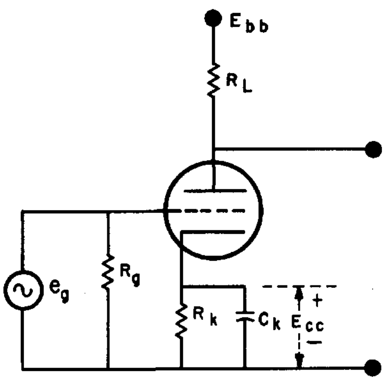


Fig. 3.8. Triode amplifier using cathode bias.

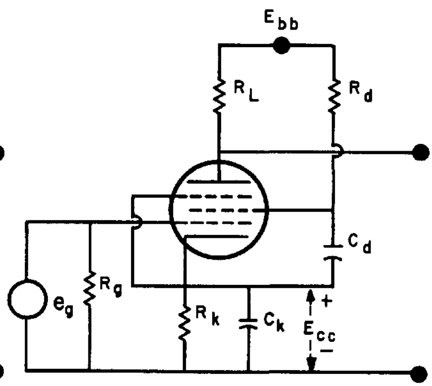


Fig. 3.9. Pentode amplifier showing screen circuit connections.

quiescent plate current I_b flows through the cathode resistance R_k, so that the cathode is positive with respect to ground by a voltage equal to I_bR_k. The grid is at ground potential because no current flows through the *grid leak resistor* R_g; hence the grid is *negative with respect to the cathode* and the purpose of grid bias has been accomplished by the cathode bias circuit. In the ideal case there is no current flow through the grid leak resistor because the grid is negative and does not attract electrons.

The grid bias can be adjusted to the proper point by setting the value of R_k such that

$$E_{cc} = I_bR_k \tag{3.3}$$

So the proper value for the cathode bias resistor is

$$R_k = \frac{E_{cc}}{I_b} \tag{3.4}$$

If the cathode capacitor C_k is omitted, a signal voltage applied to the grid of the tube will cause the plate current to vary so that the cathode bias voltage developed across R_k will vary in a similar manner. This is *not* usually desirable, although it has applications to be discussed later. The bias variation can be prevented if desired by placing the capacitor C_k in parallel with the resistor R_k as shown in figure (3.8). The capacitance of this capacitor should be rather high so that it will have a negligibly small reactance, compared with R_k, for all signal frequencies. Because of its low reactance, this *cathode by-pass capacitor* shunts nearly all the alternating component of plate current around R_k and directly to ground. Therefore, only direct current flows through R_k, and the cathode voltage is constant as desired.

A similar technique is used to obtain the correct screen voltage for tetrodes and pentodes. The screen operates at a positive voltage that should be kept as nearly constant as possible. The screen voltage E_{g_2} can be set to the proper value by connecting the screen to the plate supply through a suitable voltage dropping resistor R_d. A high capacitance capacitor C_d is then added to shunt the variable component of screen current to cathode or ground and keep the screen voltage constant. A typical circuit connection is shown in figure (3.9).

The value of R_d, the screen dropping resistor, is computed from the Kirchhoff loop equation around the screen circuit. That is,

$$E_{g_2} = E_{bb} - I_{g_2}R_d - E_{cc}$$

or, solving for R_d,

$$R_d = \frac{E_{bb} - E_{g_2} - E_{cc}}{I_{g_2}} \tag{3.5}$$

In calculating the value for the *cathode resistor* for pentodes it is well to remember that both the plate *and* screen currents flow through R_k.

3.6. Types of Vacuum Tube Amplifiers

If it is assumed that the polarizing potentials of the amplifier tube and the magnitude of the applied signal voltage are adjusted for *class A operation*, the equivalent circuit of the amplifier tube appears as shown in figure (3.10). As far as variational or signal voltages are concerned, the amplifier tube is a three-terminal circuit element; the three terminals are the grid, cathode, and plate.

Any one of the three electrodes can be used as the reference, or *grounded* electrode, where *grounded* means that it is the reference point for variational or signal voltages. This has led to the designation of three amplifier circuit configurations as: (1) grounded cathode amplifier; (2) grounded plate amplifier (cathode follower); (3) grounded grid amplifier. For a general three-terminal network the point selected as the reference does not affect the operation of the circuit, yet the three

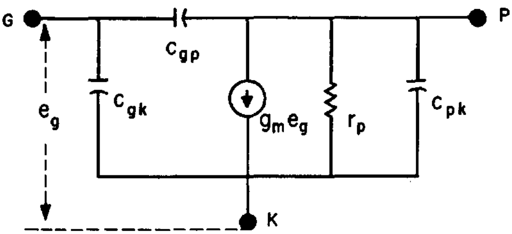


Fig. 3.10. Class A equivalent circuit of a vacuum tube.

amplifier types exhibit dissimilar operating characteristics. Clearly, there is more to the difference in these three circuits than a mere selection of reference point. The true origin of the differences should be clear from the circuit diagrams given in figure (3.11).

In figure (3.11) you will observe that an arbitrary three-terminal

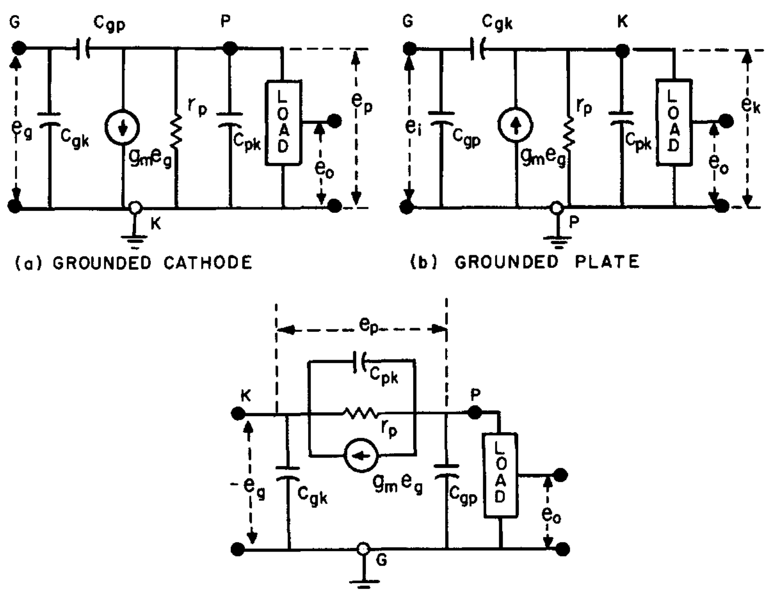


Fig. 3.11. Class A equivalent circuits of vacuum tube amplifiers.

load circuit is shown connected to each amplifier. This is the general case. All practical amplifiers have load circuits with two or three terminals. The two-terminal circuit is a special case of the three-terminal circuit, so we will assume the general case for all discussions in this chapter.

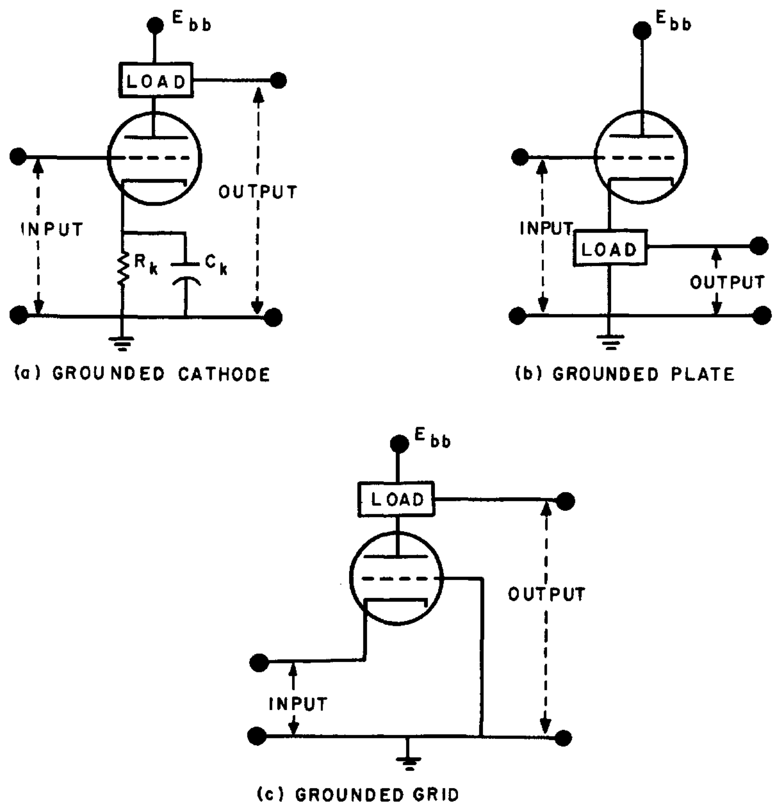


Fig. 3.12. Basic amplifier types showing d-c connections and signal input and output terminals.

When an electrode is said to be *grounded*, the term means *grounded for signal voltages only*; the grounded electrode may actually be at a high direct potential with respect to the circuit ground. This should be clear from figure (3.12) which shows *actual* circuit connections for the three basic amplifier types. The class A equivalent circuits were given in figure (3.11); they were derived from the circuits of figure (3.12) by assuming (1) the internal impedance of the plate power supply E_{bb} is zero, and (2) the reactance of C_k is zero at all frequencies of interest. Although triodes are shown in figure (3.12), pentodes also may be used.

3.7. General Equation for Voltage Amplification

The equivalent circuit of a general class A grounded cathode amplifier is shown in figure (3.11a). For the time being, assume that the grid-to-plate interelectrode capacitance C_{gp} is so small that its reactance at all signal frequencies of interest approaches that of an open circuit.

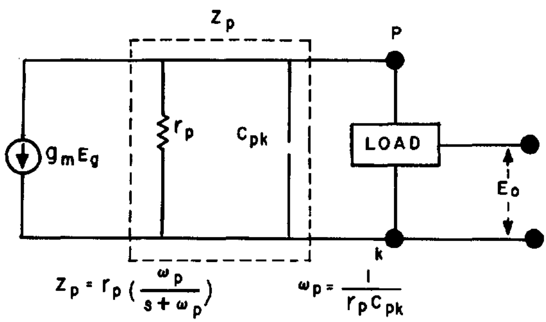


Fig. 3.13. Equivalent plate circuit of a grounded cathode amplifier.

Because of this approximation, the plate circuit is isolated from the grid circuit and can be considered by itself. Thus, the equivalent *plate* circuit of the amplifier appears as shown in figure (3.13).

The amplifier excitation, or signal input, is E_g, while the output is E_o. From chapter 2 it will be recalled that the ratio of the response function to the excitation function is the transfer function of the system. In this case, both the response and excitation are voltages, so that the amplifier has a voltage transfer function called the *voltage amplification*, which is denoted by A. Thus

$$A = \frac{E_o}{E_g} = \text{voltage amplification}$$

This can be computed from the equivalent circuit.

The *passive* circuit elements in the equivalent plate circuit of figure (3.13) arise from two sources: (1) the contributions of the tube, r_p and C_{pk} (or C_o for a pentode); (2) the connected load circuit and its associated wiring capacitance to ground. Designate the parallel combination of r_p and C_{pk} as z_p to indicate that it is an impedance associated with the plate of the tube. Hence

$$z_p = \frac{r_p}{1 + s/\omega_p} = r_p \frac{\omega_p}{s + \omega_p} \tag{3.6}$$

where

$$\omega_p = \frac{1}{r_p C_{pk}} \tag{3.7}$$

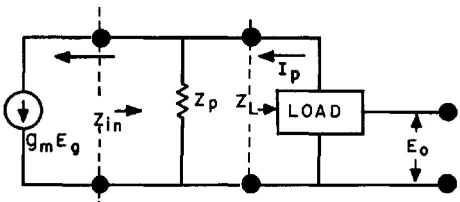


Fig. 3.14. Terminology associated with the equivalent plate circuit.

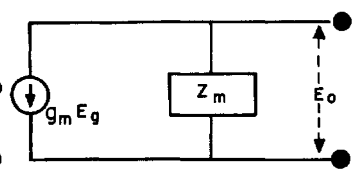


Fig. 3.15. Final form of the equivalent circuit of a grounded cathode amplifier.

Therefore the equivalent plate circuit reduces to the form shown in figure (3.14). The quantity Z_L is defined as the input impedance of the connected load circuit together with its associated distributed wiring capacitance to ground. Two other important terms can be defined from figure (3.14), as follows: Z_m = mutual impedance of the entire passive network in the equivalent plate circuit $= E_o/I$; Z_{in} = input impedance of the entire passive network in the equivalent plate circuit, or $Z_{in} = z_p Z_L/(z_p + Z_L)$.

If the passive network between the marked terminals is replaced by its mutual impedance, the equivalent circuit assumes the simple form shown in figure (3.15). Because the mutual impedance was defined as $Z_m = E_o/I$, then $E_o = IZ_m$. However, $I = -g_m E_g$, so that the output voltage is

$$E_o = -g_m E_g Z_m$$

Divide this equation through by E_g to obtain the voltage amplification as

$$A = \frac{E_o}{E_g} = -g_m Z_m \tag{3.8}$$

This is an *extremely important* equation, for it shows that the voltage amplification of any class A grounded cathode amplifier is given by the product of the tube transconductance and the mutual impedance of the passive network in the equivalent plate circuit. The minus sign signifies the phase inversion of the signal, a point explained in section (3.3). It will be shown later that all three of the basic amplifier types have voltage amplification equations of this same general form.

It has become common practice in the field of electronics to refer to the voltage amplification of a circuit as the *voltage gain,* or even just as the gain when it is understood that voltage is of primary importance. There is some objection to this terminology and with some justification. However, the usage appears well established, and the words *amplification* and *gain* will be used interchangeably here.

The general gain equation for a class A grounded cathode amplifier is

$$A = -g_m Z_m$$

where $g_m =$ mutual transconductance of the tube; $Z_m =$ mutual impedance of the passive network in the equivalent plate circuit.

The study of voltage amplifiers is largely concerned with an examination of the factors affecting this general amplification equation. Thus there are two natural sub-divisions of material: (1) study of the factors affecting g_m, and (2) a determination of the mutual impedance of various networks and the establishment of criteria for comparing them.

This chapter is primarily concerned with an evaluation of the factors affecting the transconductance of the tube. Chapter 4 is a detailed study of the effects of various circuit configurations upon the mutual impedance term in the general gain equation.

A component of the amplifier that should be considered at the very first is the tube itself. What are the effects on g_m and Z_m if a pentode is used instead of a triode? Generally, the transconductances of triodes and pentodes are of the same general order of magnitude, so there is little effect there. However, the plate resistance r_p of a pentode is from 10 to 1000 times larger than the r_p of a triode. The effect of a low value for r_p is a reduction in the size of Z_m, and this causes a corresponding reduction in the voltage gain. Hence, the voltage gain of a triode amplifier is usually less than that of a pentode amplifier even if both tubes have the same mutual transconductance.

It is obvious that the mutual impedance will vary with different network configurations, and we might conclude that any kind of amplifier response could be obtained simply be selecting the particular circuit having the desired response characteristics. However, other factors govern the form of the connected load circuit, and these were discussed in section (3.1).

3.8. Practical Amplifier Circuits

The discussion in the preceding sections was based upon the equivalent plate circuit of a grounded cathode amplifier in which only

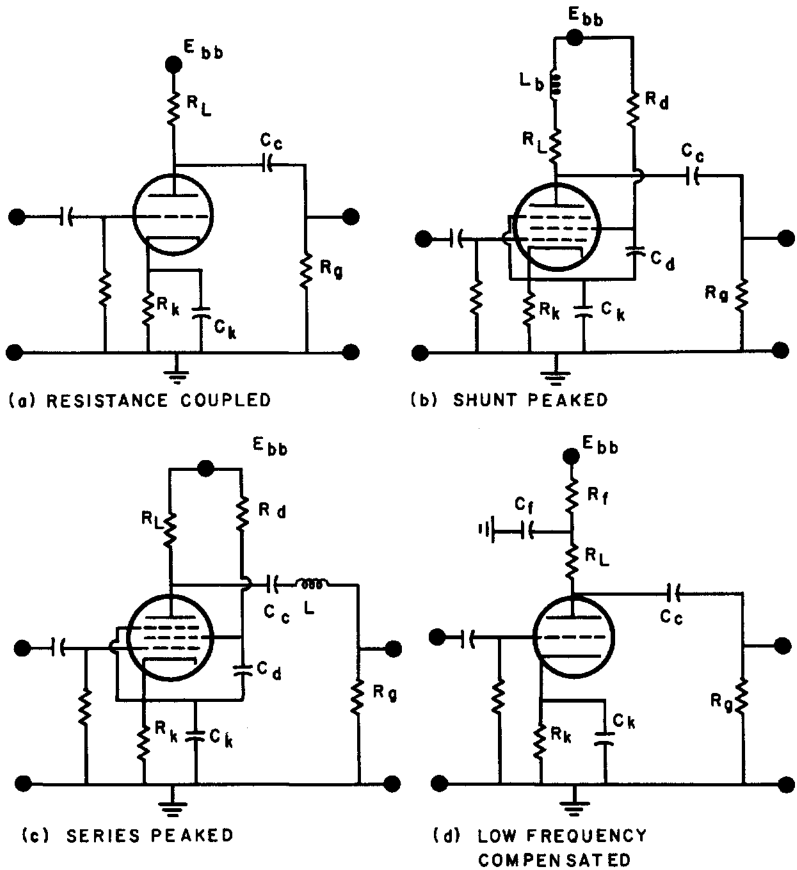


Fig. 3.16. Representative grounded cathode amplifiers.

variational components of plate current were of interest. The d-c aspect of the problem did not enter into this analysis. However, proper operation of the tube requires proper adjustment of the electrode potentials; this means that the tube must be connected as specified in section (3.1).

A number of examples of practical amplifier circuits are shown in figures (3.16) and (3.17) and identified by their most commonly accepted names. You should examine each circuit carefully to see that the four requirements specified in section (3.1) are actually met. Only those circuit elements actually connected in the circuit are shown. Unseen

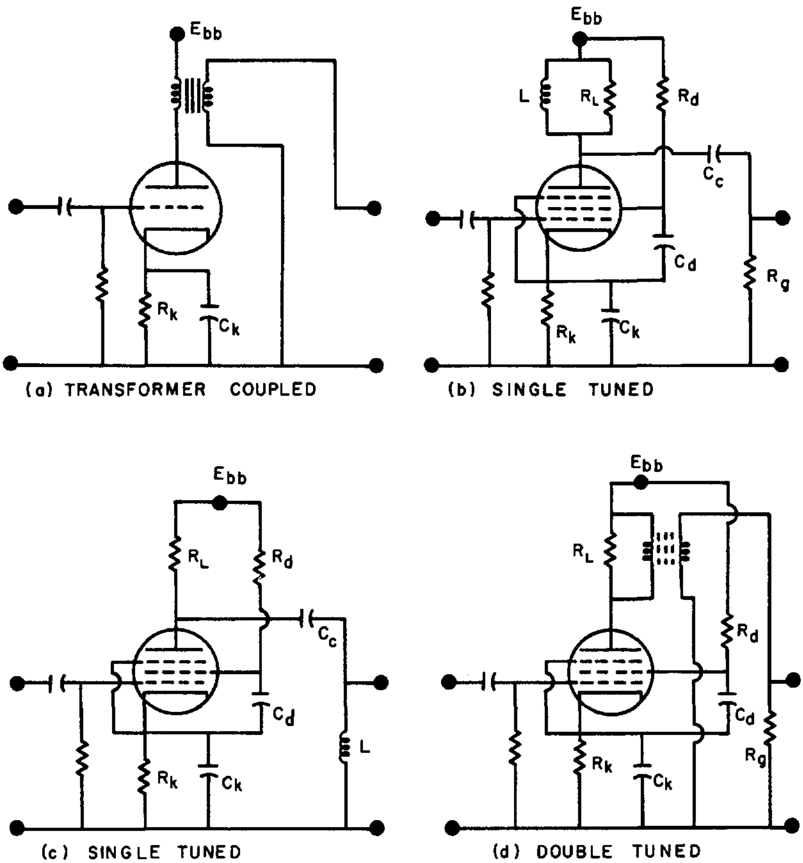


Fig. 3.17. Representative grounded cathode amplifiers.

elements are present as well, principally the input and output capacitances of the tubes and the wiring capacitances to ground caused by the presence of the various network elements.

All the amplifiers in these two figures are of the grounded cathode type.

Most of the amplifiers in these two figures are discussed separately and in detail in the next chapter.

3.9. Input Admittance of Grounded Cathode Amplifiers

In the derivation of the general equation for the voltage amplification of a grounded cathode amplifier, it was assumed that the grid-to-plate interelectrode capacitance C_{gp} was so small that it provided very little coupling between the grid and plate circuits. This is not always a valid approximation, and the effect of C_{gp} should be evaluated.

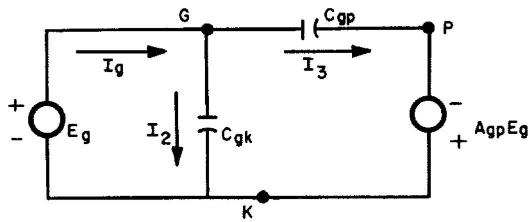


Fig. 3.18. Simplified equivalent circuit of a grounded cathode amplifier.

Reference should be made to figure (3.11a), which showed the general equivalent circuit for a grounded cathode amplifier. Define a new parameter as follows:

$$A_{gp} = -\frac{E_p}{E_g} = -\frac{\text{plate-to-cathode voltage}}{\text{grid-to-cathode voltage}} \tag{3.9}$$

This is *not* necessarily the voltage gain of the amplifier because the output from the amplifier might *not* be taken from the plate terminal of the tube. Rearrange terms slightly and write

$$E_p = -A_{gp}E_g \tag{3.10}$$

The plate circuit of the amplifier can be considered to be equivalent to a generator of voltage $E_p = -A_{gp}E_g$ and the amplifier equivalent circuit can be redrawn as shown in figure (3.18). This is an application of the *compensation theorem.*

In the a-c steady state case, the voltage loop equations can be written directly from figure (3.18) as

$$\frac{I_2}{j\omega C_{gk}} = \frac{I_3}{j\omega C_{gp}} - A_{gp}E_g = E_g \tag{3.11}$$

Consequently, the individual branch currents are

$$I_2 = j\omega C_{gk}E_g \tag{3.12}$$

$$I_3 = j\omega C_{gp}(1 + A_{gp})E_g \tag{3.13}$$

At the current node

$$I_g = I_2 + I_3 = j\omega[C_{gk} + C_{gp}(1 + A_{gp})]E_g \tag{3.14}$$

The input admittance of the amplifier is obtained by dividing this equation through by E_g. Therefore

$$Y_{in} = \frac{I_g}{E_g} = j\omega[C_{gk} + C_{gp}(1 + A_{gp})] \tag{3.15}$$

No restriction is placed upon the nature of A_{gp}. It is a ratio of two voltages, so it is a dimensionless number, but it could be either real or complex. If A_{gp} is a real number, the input admittance is wholly imaginary, or capacitive, and the input capacitance is

$$C_{in} = C_{gk} + C_{gp}(1 + A_{gp}) \tag{3.16}$$

Unfortunately, in some cases the function A_{gp} is a complex number. It can then be expressed in rectangular form as

$$A_{gp} = |A_{gp}|(\cos\theta + j\sin\theta) = |A_{gp}|\underline{/\theta} \tag{3.17}$$

Then the input admittance becomes

$$Y_{in} = j\omega\{C_{gk} + C_{gp}[1 + |A_{gp}|(\cos\theta + j\sin\theta)]\} \tag{3.18}$$

Collect real and imaginary terms and write

$$Y_{in} = G_{in} + j\omega C_{in} \tag{3.19}$$

where

$$G_{in} = -\omega C_{gp}|A_{gp}|\sin\theta \tag{3.20}$$

$$C_{in} = C_{gk} + C_{gp}(1 + |A_{gp}|\cos\theta) \tag{3.21}$$

This change in the input admittance is called the *Miller effect*.[1]

The equivalent input conductance G_{in} is negative if $\sin\theta$ is positive, and this is an important effect. It indicates that energy is being fed back into the grid circuit from the plate circuit through C_{gp}. Under certain conditions this will cause the amplifier to *oscillate*; this results from the production of right half plane poles (see chapter 2). The phenomenon is common in triodes having resonant plate circuits. The effect can be overcome by means of *neutralizing* circuits, which are discussed in a later chapter.

The value of G_{in} can be made smaller by reducing C_{gp}. The pentode was the ultimate result of this effort to reduce C_{gp}, and pentodes, which have small C_{gp}, generally do not need neutralizing circuits to prevent oscillation.

It has been shown in this section that the effect of C_{gp} can be represented by replacing the equivalent circuit of figure (3.11a) with a new

[1] After J. M. Miller, "Dependence of Input Impedance of a Three-Electrode Vacuum Tube Upon the Load in the Plate Circuit," *Bull. 351*, National Bureau of Standards, Washington, 1919.

equivalent circuit of the form shown in figure (3.19). In other words, the coupling between the grid and plate circuits is replaced by an admittance in shunt with the amplifier input terminals and loading the signal source driving the amplifier. Thus, through this method, the *form* of the equivalent *plate* circuit used in deriving the general voltage gain equation is preserved even when the effect of C_{gp} is not negligible.

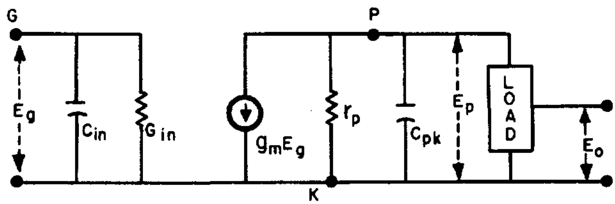


Fig. 3.19. Equivalent circuit of a grounded cathode amplifier in which coupling through C_{gp} is replaced by a shunt input admittance dependent upon C_{gp}.

3.10. Cathode Degeneration

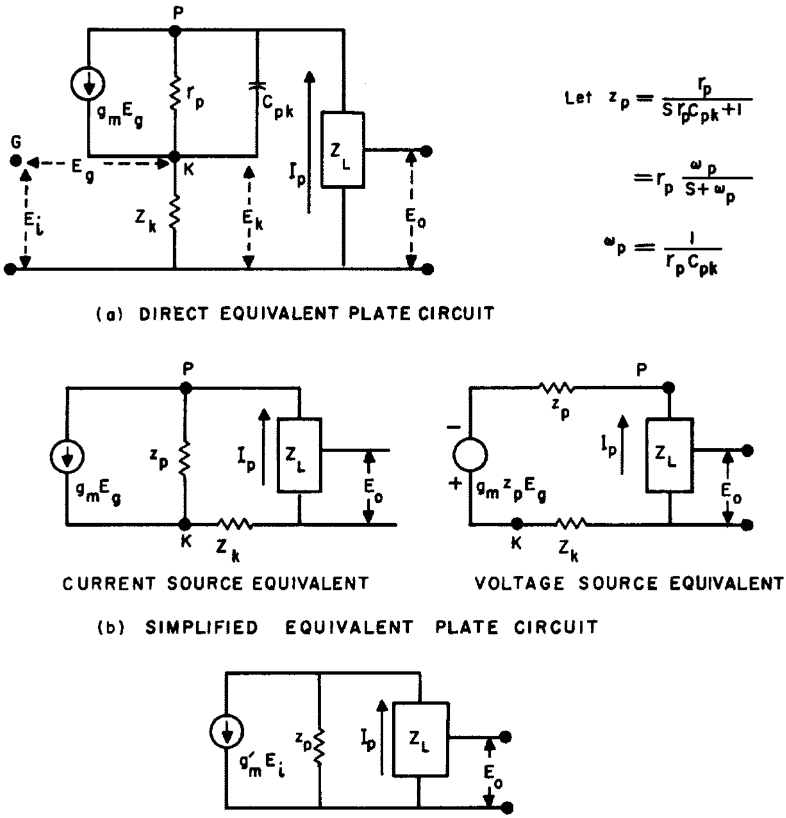
In the derivation of the general voltage gain equation for the grounded cathode amplifier it was assumed that the cathode capacitor provided perfect bypassing so that the cathode voltage was always constant. The cathode circuit thereby appeared as a signal frequency short circuit, effectively grounding the cathode for all signal currents. This ideal condition does not exist at low frequencies. Also, it is not uncommon to leave a part or all of the cathode resistance unbypassed to achieve certain operating characteristics. Hence, the effects produced are quite important when the cathode impedance is not negligible.

The equivalent plate circuit of an amplifier having an appreciable cathode impedance is shown in figure (3.20a), where, as before, Z_L is the input impedance of the connected load circuit. Here r_p and C_{pk} are combined into the single impedance z_p, so that the equivalent circuit of figure (3.20b) results. These are actual, or direct, equivalent circuits.

The circuit of figure (3.20c) shows the form *desired* for the equivalent circuit. It is desired because it has precisely the same *form* as that of the amplifier with perfect bypassing, and the voltage gain equation is clearly

$$A = -g'_m Z_m \tag{3.22}$$

where Z_m has exactly the same meaning as before and g'_m = *effective* transconductance of the tube and amplifier circuit. The derivation of



(c) PROPOSED EQUIVALENT CIRCUIT

Fig. 3.20. Equivalent plate circuits of an amplifier having appreciable cathode impedance Z_k.

this voltage gain equation requires the evaluation of g'_m. This is easily computed by determining the currents I_p flowing in circuits (3.20b) and (3.20c). If these circuits are to be equivalent, these currents must be equal.

The Kirchhoff voltage loop equation for the actual voltage source equivalent circuit of figure (3.20b) is

$$g_m z_p E_g = I_p(z_p + Z_L + Z_K) \tag{3.23}$$

However, from figure (3.20a) it is clear that

$$E_g = E_i - E_k = E_i - I_p Z_k \tag{3.24}$$

Therefore the voltage loop equation can be written

$$g_m z_p E_i = I_p[z_p + Z_L + Z_k(g_m z_p + 1)] \tag{3.25}$$

and the loop current is

$$I_p = g_m \frac{z_p}{z_p + Z_L + Z_k(g_m z_p + 1)} E_i \tag{3.26}$$

The current I_p in the proposed equivalent circuit is easily computed to be

$$I_p = g_m' \frac{z_p}{z_p + Z_L} E_i \tag{3.27}$$

Now equate these two expressions for I_p and solve for g_m'. The result is

$$g_m' = \frac{g_m}{1 + Z_k(g_m z_p + 1)/(z_p + Z_L)} \tag{3.28}$$

Therefore the voltage gain of the amplifier is

$$A = -g_m' Z_m = -\frac{g_m Z_m}{1 + Z_k(g_m z_p + 1)/(z_p + Z_L)} \tag{3.29}$$

Because the denominator of the equation for the voltage gain is always larger than unity, the gain is always less than that obtained when the cathode resistance is perfectly bypassed. In other words, the effective transconductance g_m' is always less than the actual transconductance g_m of the tube. This results in a reduction in the gain of the amplifier. It is called *cathode degeneration* because it is produced by the cathode impedance.

There are a number of special cases of equation (3.28) for g_m'. If the cathode impedance is the rather common parallel RC circuit, the impedance Z_k is negligible at all frequencies except those in the low frequency range. At the low frequencies,

$$z_p = r_p \frac{1}{1 + s/\omega_p} \doteq r_p \tag{3.30}$$

because $s/\omega_p \doteq 0$. Therefore, the effective transconductance is

$$g_m' = \frac{g_m}{1 + Z_k(\mu + 1)/(r_p + Z_L)} \tag{3.31}$$

where $\mu = g_m r_p$. This is a very useful form.

The general equation for g_m' given in (3.28) must be used when an unbypassed cathode resistor is used, because the low frequency approximation just used is invalidated.

When pentodes are used for the amplifier tube, another approximate form for g_m' can be used. In such cases it is generally true that

$$z_p \geqq 10Z_L \quad \text{and} \quad z_p \geqq 10Z_k$$

As a result, the effective transconductance is approximately

$$g_m' \doteq \frac{g_m}{1 + g_m Z_k} \tag{3.32}$$

As long as the stated approximations hold, this expression is valid for any cathode impedance Z_k in any frequency range.

The locations of the pole and zero of the effective transconductance function are of interest because they affect the over-all response of the amplifier. For simplicity, and because of its use in many practical cases, equation (3.32) for g_m' will be used in analyzing the pole and zero locations.

If the common case of a parallel *RC* cathode circuit is assumed, the impedance of the cathode circuit is

$$Z_k(s) = \frac{R_k}{s(R_k C_k) + 1} \tag{3.33}$$

Hence the approximate form for the effective transconductance becomes

$$g_m' = \frac{g_m}{1 + g_m Z_k} = \frac{g_m}{1 + g_m R_k/[s(R_k C_k) + 1]} \tag{3.34}$$

Define a new term as follows:

$$\omega_k = \frac{1}{R_k C_k} \tag{3.35}$$

Substitute this into equation (3.34) and rearrange terms until the effective transconductance is found to be

$$g_m' = g_m \frac{(s + \omega_k)}{s + (1 + g_m R_k)\omega_k} \tag{3.36}$$

Hence the function has a pole at

$$s = -(1 + g_m R_k)\omega_k \tag{3.37}$$

and a zero at

$$s = -\omega_k. \tag{3.38}$$

The relative locations of the pole and zero are shown in the complex s plane given in figure (3.21).

From the nature of the terms it is clear that the pole is always further out along the real axis than is the zero. Therefore the effective transconductance introduces additional phase shift and amplitude reduction in the voltage amplification of the amplifier. It will have a unity amplitude factor and zero phase shift in the effective transconductance only when the pole and zero coincide.

Coincidence of the pole and zero is possible only if the capacitance of C_k approaches infinity or if R_k becomes zero. This indicates that the

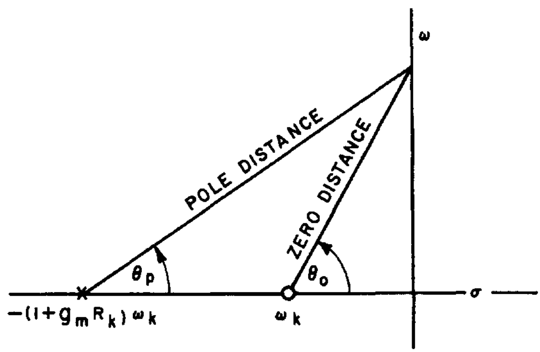


Fig. 3.21. Pole and zero of the effective transconductance function.

degenerative effect of the cathode impedance can be reduced by making C_k as large as possible and R_k as small as possible. Coincidence would also be produced if g_m were zero. This makes the gain zero and is a trivial case, but it does show that high g_m tubes are more susceptible to cathode degeneration.

Cathode degeneration may be desirable in some cases. For example, it is often necessary to have amplifiers with gain characteristics that are not noticeably changed when tubes are replaced. Because a 2-to-1 range in g_m may occur among tubes of the same type, cathode degeneration can be used to make the effective transconductance largely independent of tube variations.

3.11. Screen Degeneration

A degenerative effect, similar to that produced by the cathode bias circuit, results from imperfect screen bypassing. When the screen circuit impedance Z_s is appreciable, the signal component of screen current I_s produces a voltage drop $E_s = I_sZ_s$ in this impedance, and this causes the *plate* current to decrease. The effect is best analyzed from an equivalent circuit representation of the phenomenon. For this purpose, define the following terms:

$$r_s \overset{\Delta}{=} \frac{\partial e_d}{\partial i_d} = \text{variational screen resistance} \tag{3.39}$$

$$\doteq \text{plate resistance when the tube is triode connected}$$

$$g_s \overset{\Delta}{=} \frac{\partial i_b}{\partial e_d} = \text{screen grid-to-plate transconductance} \tag{3.40}$$

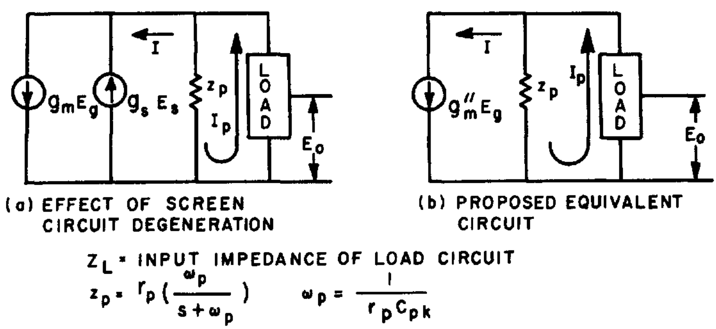


Fig. 3.22. Equivalent plate circuit relationships for screen circuit degeneration.

The terminology and symbolism used for the currents and voltages are specified as follows:

$e_d = E_d + E_s \sin(\omega t + \theta_1)$; $E_d = Q$ point screen voltage

$i_d = I_d + I_s \sin(\omega t + \theta_2)$; $I_d = Q$ point screen current

$i_b = I_b + I_p \sin(\omega t + \theta_3)$; $I_b = Q$ point plate current

The effect of screen circuit degeneration can be described in terms of these definitions as shown by the equivalent circuit given in figure (3.22a). To preserve the same form for the equivalent circuit as used in all previous cases, a proposed equivalent circuit is shown in figure (3.22b), where a new effective transconductance g''_m has been specified to include the effect of screen degeneration. The following derivation simply establishes an equivalence between the plate currents in the actual and proposed equivalent circuits, and this defines the effective transconductance in terms of the circuit constants.

From figure (3.22a) it is clear that

$$I = g_mE_g - g_sE_s \tag{3.41}$$

In the *absence of cathode degeneration*, you can see that

$$E_s = I_s Z_s \tag{3.42}$$

Consequently, equation (3.41) becomes

$$I = g_m E_g - I_s g_s Z_s \tag{3.43}$$

In pentodes, the screen and plate currents bear an almost constant relationship to one another. Thus.

$$k \triangleq \frac{\partial i_d}{\partial i_b} = \frac{I_s}{I_p} \tag{3.44}$$

In most pentodes the value of k will be about 0.3. From equation (3.44),

$$I_s = kI_p \tag{3.45}$$

Therefore, equation (3.43) can be written

$$I = g_m E_g - kI_p g_s Z_s \tag{3.46}$$

According to the proposed equivalent circuit of figure (3.22b), the plate current I_p is

$$I_p = g_m'' E_g \frac{z_p}{z_p + Z_L} \tag{3.47}$$

where

$$g_m'' = I/E_g \tag{3.48}$$

Thus, equation (3.46) becomes

$$I = g_m E_g - g_m'' E_g \frac{z_p}{z_p + Z_L} kg_s Z_s \tag{3.49}$$

Divide this equation through by E_g, replace I/E_g with g_m'', and solve the result for g_m''. This procedure gives

$$g_m'' = \frac{g_m}{1 + (kg_s)Z_s z_p/(z_p + Z_L)} \tag{3.50}$$

The kg_s factor can be simplified because

$$kg_s = \left(\frac{\partial i_d}{\partial i_b}\right)\left(\frac{\partial i_b}{\partial e_d}\right) = \frac{\partial i_d}{\partial e_d} = \frac{1}{r_s} \tag{3.51}$$

Thus the effective transconductance of a grounded cathode amplifier with *only* screen circuit degeneration is

$$g_m'' = \frac{g_m}{1 + \dfrac{Z_s}{r_s} \cdot \dfrac{z_p}{z_p + Z_L}} \tag{3.52}$$

There are several important approximate forms of this equation.

Because pentodes are involved, in most cases the plate impedance z_p is much larger than the connected load impedance Z_L so that

$$z_p \gg Z_L \tag{3.53}$$

Therefore

$$g_m \doteq \frac{g_m}{1 + Z_s/r_s} \tag{3.54}$$

This approximation is valid for nearly all practical cases.

An alternative form of equation (3.52) can be used because the degenerative effect of the screen circuit is usually most pronounced at the low frequencies. In the low frequency range it was shown in section (3.10) that $z_p \doteq r_p$. Hence

$$g_m'' \text{ (low frequency)} = \frac{g_m}{1 + \dfrac{Z_s}{r_s} \cdot \dfrac{r_p}{r_p + Z_L}} \tag{3.55}$$

3.12. Simultaneous Cathode and Screen Degeneration

When pentodes are used in amplifiers it is possible to connect the screen bypass capacitor either to the cathode of the tube or to the circuit ground. When both the screen and cathode circuits have appreciable impedance because of imperfect bypassing, it is evident that the operating characteristics of these two circuit connections will be different. Only the low frequency case will be considered here, so that $z_p = r_p$.

Consider the case where the screen bypass capacitor is grounded. The various equivalent circuits that apply to this case are shown in figure (3.23). As a result of this connection, *both* the plate and screen currents flow through the cathode impedance. Thus

$$E_k = (I_p + I_s)Z_k \tag{3.56}$$

and

$$E_g = E_i - E_k = E_i - (I_p + I_s)Z_k \tag{3.57}$$

or

$$E_g = E_i - I_p(1 + k)Z_k \tag{3.58}$$

where

$$k = \frac{I_s}{I_p} \tag{3.59}$$

The equivalent voltage source in figure (3.23b) arising from screen circuit degeneration is

$$g_s r_p E_s = g_s r_p (I_s Z_s + E_k) \tag{3.60}$$

or

$$g_s r_p E_s = I_p r_p \left[\frac{Z_s}{r_s} + (1 + k) g_s Z_k\right] \tag{3.61}$$

because

$$k g_s = \frac{1}{r_s} \tag{3.62}$$

Consequently, the voltage loop equation for the circuit of figure (3.23b) is

$$\mu[E_i - I_p(1+k)Z_k] - I_p r_p\left[\frac{Z_s}{r_s} + (1+k)g_s Z_k\right]$$
$$= I_p[r_p + Z_L + (1+k)Z_k] \qquad (3.63)$$

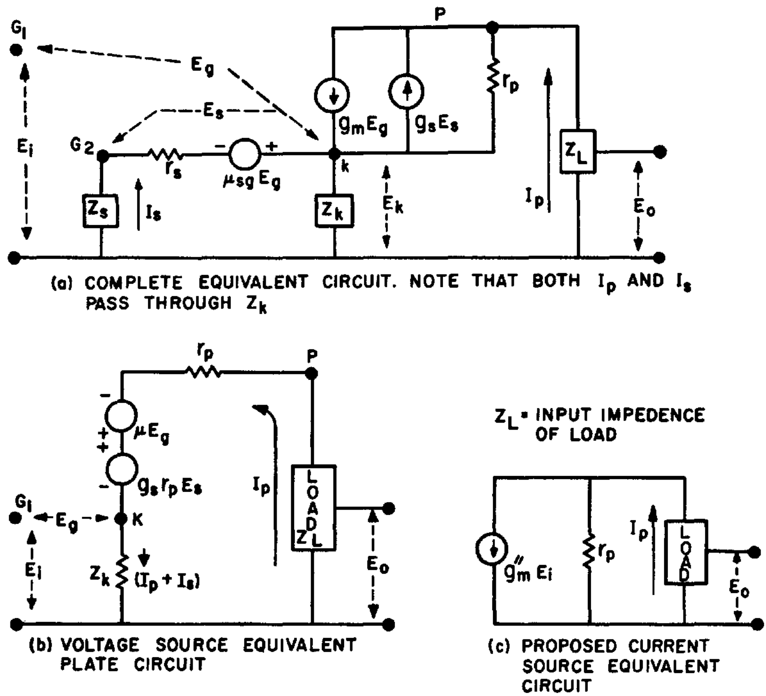


Fig. 3.23. Equivalent circuits for the case of simultaneous cathode and screen degeneration; cathode and screen bypass capacitors both grounded. Low frequency case assumed so $z_p = r_p$.

Collect all terms involving I_p and solve for μE_i.

$$\mu E_i = I_p\left\{(r_p + Z_L) + r_p\frac{Z_s}{r_s} + (1+k)\,[(\mu+1) + g_s r_p]Z_k\right\} \qquad (3.64)$$

Now, for the proposed equivalent circuit,

$$I_p = g_m'' \frac{r_p}{r_p + Z_L} E_i \qquad (3.65)$$

Substitute this relationship into equation (3.64), cancel the common E_i term, and solve for g_m''. The result is

$$g_m'' = \frac{g_m}{1 + [r_p Z_s/r_s + (\mu + 1 + g_s r_p)(1 + k)Z_k]/(r_p + Z_L)} \tag{3.66}$$

Because a pentode is involved, it is generally true that

$$r_p \gg Z_L; \qquad \mu \gg 1 \tag{3.67}$$

and so the equation for the effective transconductance reduces to the approximate form:

$$g_m'' \doteq \frac{g_m}{(1 + Z_s/r_s) + (1 + k)(g_m + g_s)Z_k} \tag{3.68}$$

A similar derivation can be carried through for the case when the screen bypass capacitor is connected to the cathode instead of to ground.

You will find it very informative to compare this result with those obtained for the individual cases of cathode and screen degeneration.

3.13. Grounded Plate Amplifier—Cathode Follower

The circuit diagram, voltage source equivalent plate circuit, and the proposed equivalent circuit of a grounded plate amplifier, or *cathode follower*, are shown in figure (3.24). In drawing the equivalent circuits, it is assumed that the impedances of the cathode and screen biasing circuits are negligibly small at all frequencies of interest. If these assumptions are not valid, it is relatively easy to compute an appropriate value for the effective transconductance, following the method outlined in the preceding section.

From the voltage source equivalent circuit of figure (3.24b), the circuit loop equation is

$$g_m z_p E_g = I_p(z_p + Z_L) \tag{3.69}$$

However, from the actual circuit diagram in figure (3.24a) you can see that

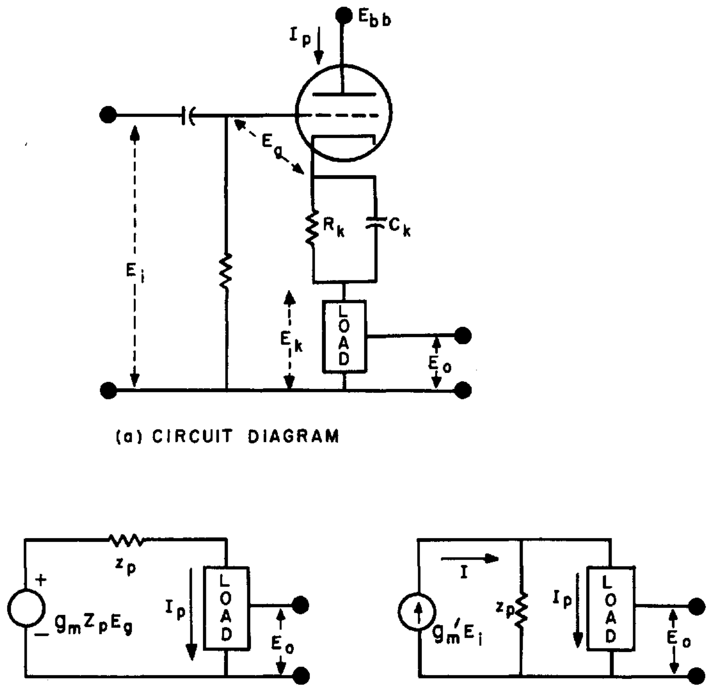
$$E_g = E_i - E_k = E_i - I_p Z_L \tag{3.70}$$

so that the voltage loop equation becomes

$$g_m z_p E_g = g_m z_p(E_i - I_p Z_L) = I_p(z_p + Z_L) \tag{3.71}$$

Solve this equation for I_p.

$$I_p = \frac{g_m z_p}{z_p + Z_L + g_m z_p Z_L} E_i \tag{3.72}$$



(a) CIRCUIT DIAGRAM

Z_L = INPUT IMPEDANCE OF LOAD

$z_p = r_p \left(\frac{\omega_p}{\omega_p + s}\right) \quad \omega_p = \frac{1}{r_p C_{pk}}$

(b) ACTUAL VOLTAGE SOURCE EQUIVALENT CIRCUIT

(c) PROPOSED CURRENT SOURCE EQUIVALENT CIRCUIT

Fig. 3.24. Circuit relationships for the cathode follower.

According to the proposed equivalent circuit this same plate current can be written in terms of the unknown effective transconductance as

$$I_p = g_m' \frac{z_p}{z_p + Z_L} E_i \tag{3.73}$$

Therefore, if the two circuits are to be equivalent to each other, the two values of plate current just computed must be equal. Set equation

(3.72) equal to (3.73), cancel the common E_i term, and then solve for the effective transconductance g'_m.

$$g'_m = \frac{g_m}{1 + g_m z_p Z_L/(z_p + Z_L)} = \frac{g_m}{1 + g_m Z_L/(1 + Z_L/z_p)}$$
$$= \frac{g_m}{1 + g_m Z_{in}} \tag{3.74}$$

because $Z_{in} = z_p Z_L/(z_p + Z_L)$. Equation (3.74) is the effective transconductance of a grounded plate amplifier.

If the entire passive impedance network connected between the cathode and ground terminals of the proposed equivalent circuit has a mutual impedance Z_m, the amplifier output voltage is clearly

$$E_o = IZ_m = g'_m E_i Z_m \tag{3.75}$$

and the voltage amplification is

$$A = \frac{E_o}{E_i} = g'_m Z_m = \frac{g_m Z_m}{1 + g_m Z_{in}} \tag{3.76}$$

Therefore the equation for the voltage amplification of a cathode follower is exactly the same as that for the grounded cathode amplifier with two exceptions:

(1) There is no minus sign on the gain equation so that there is no phase inversion of the signal.

(2) The transconductance of the amplifier is given by equation (3.74) and the value obtained is considerably less than the transconductance of the tube.

In many cases the plate impedance z_p of the tube is much larger than the magnitude of the connected load impedance Z_L. If this is true, then equation (3.74) has the approximate form:

$$g'_m \doteq \frac{g_m}{1 + g_m Z_L} \tag{3.77}$$

Under these conditions, which are common for pentodes, the voltage amplification is approximately

$$A \doteq \frac{g_m Z_m}{1 + g_m Z_L} \tag{3.78}$$

The mutual impedance Z_m will always be less than the input impedance of the load circuit Z_L, so that the voltage gain of a cathode follower can never exceed unity.

Because the grounded cathode amplifier discussed in the previous sections can provide relatively large voltage gains while the cathode follower never has a gain exceeding unity, you might wonder if the cathode follower is merely an academic curiosity. It does have a wide field of use, but the applications depend upon *other* characteristics of the circuit. It turns out that the cathode follower has a low output impedance and a high input impedance, both relative to the grounded cathode amplifier. These characteristics are discussed in the next two sections.

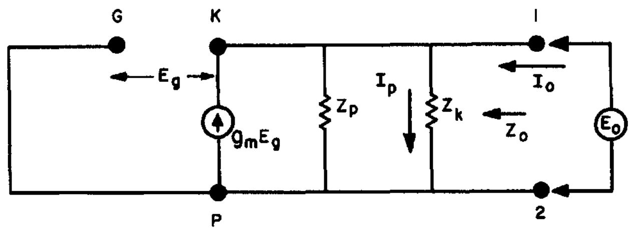


Fig. 3.25. Measuring the output impedance of a cathode follower.

3.14. Output Impedance of the Cathode Follower

The equivalent circuit of a cathode follower, neglecting the grid-to-cathode capacitance, is shown in figure (3.25). The current source equivalent of the tube is used. The load circuit is assumed to be a two-terminal network with an input impedance of Z_k. This symbol is used rather than Z_L mainly to indicate that this is *not* the general three-terminal load circuit case.

The output impedance of the cathode follower is to be computed. By *output impedance* we shall mean the impedance presented to a physical generator connected across the output terminals 1-2, with all generators to the left of these terminals replaced by their internal impedances. This is a general definition of the output impedance of any network containing both active and passive circuit elements.

The procedure for measuring the output impedance experimentally or analytically is clear from the definition. The signal source at the input E_i is replaced by its internal impedance, a short circuit in this case. Then a generator of E_o volts is applied across the output terminals 1-2 of the amplifier. The ratio of this voltage to the resulting input current I_o is the output impedance of the cathode follower. That is,

$$Z_o = \text{output impedance} = \frac{E_o}{I_o}$$

The circuit connections just discussed are shown in figure (3.25).

The applied voltage E_o is equal to the cathode voltage E_k because it is connected directly from cathode to ground. In a cathode follower, the grid-to-cathode voltage is

$$E_g = E_i - E_k \tag{3.79}$$

But in this case, the input voltage is zero because the input terminals are short circuited and the cathode voltage is equal to E_o. Therefore

$$E_g = -E_o \tag{3.80}$$

As a result, the three branch currents shown in figure (3.25) are

$$I_1 = \frac{E_o}{Z_k} \tag{3.81}$$

$$I_2 = \frac{E_o}{z_p} \tag{3.82}$$

$$I_3 = -g_m E_g = +g_m E_o \tag{3.83}$$

The total current supplied by the connected generator is the sum of these three branch currents, or

$$I_o = I_1 + I_2 + I_3 \tag{3.84}$$

$$I_o = \left(\frac{1}{Z_k} + \frac{1}{z_p} + g_m\right) E_o \tag{3.85}$$

Solve this equation for the output impedance.

$$Z_o = \frac{E_o}{I_o} = \frac{1}{1/Z_k + 1/z_p + g_m} = \frac{1}{1/Z_{in} + g_m} \tag{3.86}$$

With a little algebraic manipulation, this equation is readily expressed in any of the following forms:

$$Z_o = \frac{Z_{in}}{1 + g_m Z_{in}} = \frac{Z_k}{1 + (g_m z_p + 1)(Z_k/z_p)} \tag{3.87}$$

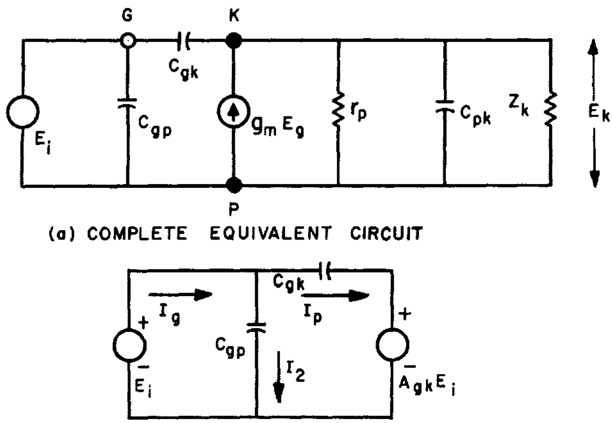
$$= \frac{Z_k}{1 + g_m Z_k + Z_k/z_p} \tag{3.88}$$

These are the general equations for the output impedance of a cathode follower. You can see that Z_o is always less than the actual cathode load impedance Z_k. To see this more clearly, some approximate relationships will be of help. For example, for many of the tubes used in cathode followers, $Z_k \ll z_p$, so that the output impedance is approximately

$$Z_o \doteq \frac{Z_k}{1 + g_m Z_k} \tag{3.89}$$

Regardless of the values of g_m and Z_k, you can see now that the output impedance will always be less than Z_k.

Because of the low value for the output impedance of a cathode follower, the circuit finds a wide field of application as a matching device between electronic systems of high impedance and load elements of low impedance such as transmission lines.



(a) COMPLETE EQUIVALENT CIRCUIT

(b) SIMPLIFIED EQUIVALENT CIRCUIT

Fig. 3.26. Circuits for computing the input admittance of a cathode follower.

3.15. Input Admittance of the Cathode Follower

The purpose of this section is to show that the input *admittance* of a cathode follower is much lower than the input admittance of a grounded cathode amplifier using the same tube. If the admittance is lower, the input impedance will be higher.

The equivalent plate circuit of a cathode follower showing all tube interelectrode capacitances is given in figure (3.26). Follow the same procedure as that used in section (3.9). Define a new parameter A_{gk} to be the ratio of the cathode voltage to the grid signal voltage; both voltages are measured with respect to ground. Therefore

$$A_{gk} = \frac{E_k}{E_i} \tag{3.90}$$

and the cathode voltage is

$$E_k = A_{gk}E_i \tag{3.91}$$

The cathode circuit can be replaced with an equivalent generator, $A_{gk}E_i$, as shown in figure (3.26b). Because the grid-to-ground voltage

E_i and cathode-to-ground voltage $A_{gk}E_i$ are in phase, the relative polarity markings on the generator must be as shown in the figure.

In terms of a-c steady state currents, the two Kirchhoff voltage loop equations for the circuit of figure (3.26b) are

$$E_i = \frac{I_2}{j\omega C_{gp}} = A_{gk}E_i + \frac{I_p}{j\omega C_{gk}} \tag{3.92}$$

Solve this for the two branch currents.

$$I_2 = j\omega C_{gp}E_i \tag{3.93}$$

$$I_p = j\omega C_{gk}(1 - A_{gk})E_i \tag{3.94}$$

The total input grid current is then

$$I_g = I_2 + I_p = j\omega[C_{gp} + C_{gk}(1 - A_{gk})]E_i \tag{3.95}$$

and the input admittance of the tube is

$$Y_{in} = \frac{I_g}{E_i} = j\omega[C_{gp} + C_{gk}(1 - A_{gk})] \tag{3.96}$$

In the most common case generally discussed, A_{gk} is a real number, so that the input admittance is purely capacitive. The parameter A_{gk} is closely related to the voltage amplification of the cathode follower, and it is generally quite close to unity. Hence the inside bracketed term in equation (3.96) is virtually zero, and the input capacitance is virtually equal to C_{gp}. If a pentode is used, this is a very small capacitance and leads to a much smaller value for the input admittance than could be obtained from a grounded cathode amplifier.

If A_{gk} is a complex number, the input admittance has both conductance and susceptance components, so that

$$Y_{in} = G_{in} + j\omega C_{in} \tag{3.97}$$

where

$$G_{in} = \omega C_{gk} \,|A_{gk}|\, \sin\theta \tag{3.98}$$

$$C_{in} = C_{gp} + C_{gk}(1 - |A_{gk}| \cos\theta) \tag{3.99}$$

The input conductance can be positive or negative depending upon the sign associated with $\sin\theta$. Usually A_{gk} is much less than A_{gp} of the grounded cathode circuit, so that the input conductance is usually less for the cathode follower. The input capacitance is also less than that for the grounded cathode amplifier.

3.16. Grounded Grid Amplifier

The circuit diagram of a typical grounded grid amplifier is shown in figure (3.27). The voltage source equivalent circuits are given in figure

(3.28). As in all previous cases the load circuit is shown as a general three-terminal network of input impedance Z_L.

The Kirchhoff voltage loop equation around the grid-to-cathode loop of the equivalent circuit is

$$E_g = E_i - I_p Z_k \tag{3.100}$$

The equation for the plate circuit loop is

$$E_i + g_m z_p E_g = I_p(z_p + Z_k + Z_L) \tag{3.101}$$

Substitute equation (3.100) for E_g in equation (3.101).

$$(g_m z_p + 1)E_i - g_m z_p Z_k I_p = I_p(z_p + Z_k + Z_L) \tag{3.102}$$

Solve for I_p.

$$I_p = \frac{g_m z_p + 1}{z_p + Z_L + Z_k(g_m z_p + 1)} E_i \tag{3.103}$$

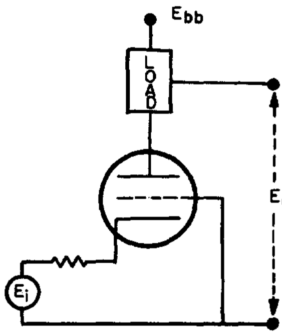


Fig. 3.27. Circuit diagram of a grounded grid amplifier.

To maintain the same general form for the gain function $A = g''_m Z_m$ as used in all previous cases, we must define a new effective transconductance to account for the effects produced by the circuit configuration. This should not alter the mutual impedance Z_m. To maintain this equality of form, the proposed equivalent circuit of figure (3.29) must exist. If it does exist, the plate current I_p must be equal to the plate current computed in equation (3.103) from the actual amplifier equivalent circuit.

From figure (3.29) the plate current can be expressed in terms of the effective transconductance as

$$I_p = g''_m \frac{z_p}{z_p + Z_L} E_i \tag{3.104}$$

Set this equation equal to equation (3.103), cancel the common E_i term, and solve for g''_m. The result is

$$g''_m = \frac{g_m z_p + 1}{z_p + \left(\dfrac{z_p Z_k}{z_p + Z_L}\right)(g_m z_p + 1)} \tag{3.105}$$

$$= \frac{g_m + 1/z_p}{1 + \left(\dfrac{Z_k}{z_p + Z_L}\right)(g_m z_p + 1)} \tag{3.106}$$

Because of the presence of the $1/z_p$ term in the numerator, the effective transconductance of a grounded grid amplifier is greater than the transconductance of a grounded cathode amplifier subject to a corresponding amount of cathode degeneration. The comparison is

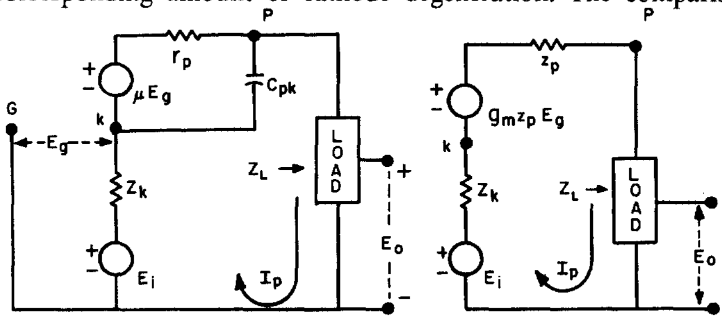


Fig. 3.28. Voltage source equivalent plate circuits of a grounded grid amplifier.

slightly more obvious if the cathode impedance Z_k is made equal to zero. Then for the grounded grid amplifier,

$$g_m'' = g_m + \frac{1}{z_p} \tag{3.107}$$

while in the grounded cathode case,

$$g_m' = g_m \tag{3.108}$$

You can immediately see the increase in transconductance and the corresponding increase in voltage amplification. It must be equally clear that the increase will be small when the plate impedance of the tube is large.

The input impedance of the grounded grid amplifier, neglecting the effects of interelectrode capacitance, is of interest. From figure (3.28) you can see that this is

$$\begin{aligned} Z_{in} = \frac{E_i}{I_p} &= Z_k + \frac{z_p + Z_L}{g_m z_p + 1} \\ &= Z_k + \text{(input impedance of tube alone)} \end{aligned} \tag{3.109}$$

In some cases, particularly for pentodes, the second term in equation (3.109) is approximately $1/g_m$. Under such conditions, the input impedance of the grounded grid amplifier is quite small. This is an advantage in some cases because the signal frequency must become very high before the shunting effects of the input capacitance of the amplifier become objectionable. Because of its low input impedance, the circuit is often used as an impedance transforming device connecting a low impedance generator, such as a transmission line, to a high impedance circuit, such as a grounded cathode amplifier.

Tubes designed for use in grounded grid amplifiers are constructed so that the grid provides excellent electrostatic shielding of the cathode from the plate. This reduces the Miller effect by reducing the coupling between input and output circuits. This makes the grounded grid

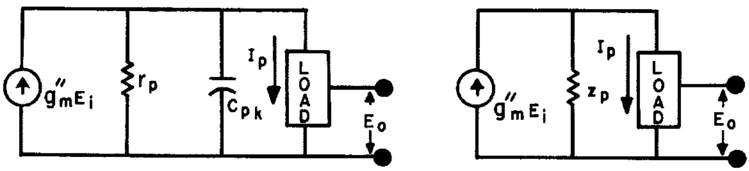


Fig. 3.29. Proposed equivalent circuits for a grounded grid amplifier.

amplifier more stable than the grounded cathode connection. As a result, triodes can be used as grounded grid amplifiers at much higher frequencies than when connected as grounded cathode circuits. This is particularly important in the design of *low noise* circuits where triodes are preferred because they have smaller noise parameters (see chapter 9).

3.17. Decibels and the Use of Relative Magnitudes

There are many cases in electrical engineering in which it is desirable to express the performance characteristics of a device or circuit in terms of *relative* magnitude rather than absolute magnitude. In the study of amplifiers, the amplification is generally more important, from the theoretical viewpoint, than the actual voltage output. The amplification is dimensionless because it is a ratio of the output voltage to the input voltage. The same considerations apply in any energy or signal transmission system in which transmission losses or gains are best expressed in terms of ratios.

Aside from the generality obtained by using relative rather than absolute magnitudes to express performance characteristics, it is also generally desirable to use logarithmic units rather than numerical ratios; this is so because log units reduce multiplication and division of numerical ratios to simple addition and subtraction of logarithmic units.

The computational advantages associated with the use of relative magnitudes expressed in terms of logarithmic units was recognized a long time ago. A special unit called the *decibel* (abbreviated db) was defined in terms of a *power ratio* as

$$\text{number of db} = 10 \log_{10} \left(\frac{P_o}{P_i}\right) \tag{3.110}$$

where 1 decibel corresponds to a power ratio that can be computed from equation (3.110) to be

$$1 \text{ db} = 10 \log_{10} \left(\frac{P_o}{P_i}\right) \quad \text{or} \quad 1 = \log_{10} \left(\frac{P_o}{P_i}\right)^{10}$$

so that

$$\frac{P_o}{P_i} = 10^{0.1} = 1.25893^{-} \tag{3.111}$$

Thus a power ratio of 1.25893 corresponds to 1 db in logarithmic units.

When the powers used in the evaluation of the decibel power loss or gain are dissipated in resistances, equation (3.110) can be rewritten as

$$\begin{aligned} \text{db} &= 10 \log_{10} \frac{E_o^2/R_o}{E_i^2/R_i} = 10 \log_{10} \frac{E_o^2}{E_i^2} \cdot \frac{R_i}{R_o} \\ &= 20 \log_{10} \frac{E_o}{E_i} + 10 \log_{10} \frac{R_i}{R_o} \end{aligned} \tag{3.112}$$

If the input and output resistances are equal, the last term in equation (3.112) is zero and the *power gain* in decibels can be computed directly from the voltage ratio as

$$\text{db} = 20 \log_{10} \frac{E_o}{E_i} \tag{3.113}$$

In many respects this is an unfortunate result because it has led to widespread incorrect use of decibels. Equation (3.113) is often and *incorrectly* used to compute voltage amplification even when the input and output resistances are unequal. This is a violation of the definition of the decibel that specifically required a power ratio. The error has perpetuated itself and expanded into other areas of incorrect usage until the significance of the decibel as a power ratio has been all but lost. This alarming development was noted by Horton[1], and recommendations were made for correction. As nearly as possible I will follow these recommendations with occasional reference to usual, if incorrect, practice.

The incorrect usage of the decibel arose from the great desire to express relative magnitudes in terms of equivalent logarithmic units.

[1] J. W. Horton, "Fundamental Considerations Regarding the Use of Relative Magnitudes," *Proc. I.R.E.*, vol. 40, April, 1952, pp. 440–444.

Essentially, the problem reduces to specifying a general formula for the use of relative magnitudes that uses a decimal system; such a general expression will include the decibel as a specific case of a power ratio.

Horton proposes that a new and general unit, called the *logit*, be defined as follows:

$$1 \text{ logit} = 10 \log_{10} (1.25893)$$

and

$$\text{number of logits} = 10 \log_{10} (\text{relative magnitude})$$

In other words, a 1 logit change in relative magnitude corresponds to a change of 1.25893 in numerical ratio. Clearly then, a *decibel* is simply a *power logit*. Voltage amplification can be expressed in terms of *volt logits* as

$$\text{gain in volt logits} = 10 \log_{10} \frac{E_o}{E_i}$$

Similarly, a change in relative length can be expressed in length logits and so on for any type of units.

It is proposed by Horton that the symbol ℓ be used to designate a logit and the type of logit designated by a prefix. Thus:

pℓ = power logit = decibel
vℓ = volt logit
iℓ = current logit
dℓ = length logit
mℓ = mass logit

The use of a new and as yet nonstandard unit is a little dangerous because it could lead to confusion in evaluating current, though incorrect, terminology. However, the main difficulty will arise in this book in specifying the voltage gain of vacuum tube amplifiers because it is generally given in decibels in present practice and computed from the formula

$$\text{voltage gain in db} = 20 \log_{10} \frac{E_o}{E_i} \longleftarrow \text{INCORRECT}$$

Because the voltage gain in volt logits is properly expressed as

$$\text{voltage gain in v}\ell = 10 \log_{10} \frac{E_o}{E_i} \longleftarrow \text{CORRECT}$$

it is clear that the improper designations in decibels are easily converted to volt logits by simply dividing the number of decibels by 2.

TABLE 3

PROPERTIES OF CLASS A VOLTAGE AMPLIFIERS

Grounded cathode	Grounded plate	Grounded grid
$A = -g_m Z_m$ with perfect bypassing $A = -g'_m Z_m$ with imperfect bypassing	$A = +g'_m Z_m$	$A = +g''_m Z_m$
g_m = handbook value $g'_m = \dfrac{g_m}{1 + Z_k \dfrac{g_m z_p + 1}{z_p + Z_L}}$	$g'_m = \dfrac{g_m}{1 + g_m Z_{in}}$	$g''_m = \dfrac{g_m + 1/z_p}{1 + Z_k \dfrac{g_m z_p + 1}{z_p + Z_L}}$
$C_{in} = C_{gk} + C_{gp}(1 + A_{gp})$	$C_{in} = C_{gp} + C_{gk}(1 - A_{gk})$	$C_{in} = C_{gk} + C_{pk}(1 - A_{pk})$
$Z_o = Z_m$	$Z_o = \dfrac{Z_L}{1 + \dfrac{Z_L(g_m z_p + 1)}{z_p}}$	$Z_o = Z_m$

PROBLEMS

3.1. Derive the equation given in Table 3 for the input capacitance of a grounded grid amplifier.

3.2. Using the class A equivalent plate circuit of the circuit shown in figure (3.30), derive an equation for the transconductance of the tube in terms of R_L, R_g, and R_x. Assume that proper positioning of the controls is obtained when the headphone current is zero. Assume the headphone impedance is zero and that C_c and C_k are perfect short circuits at the signal frequency. You may also assume that r_p is much larger than R_L.

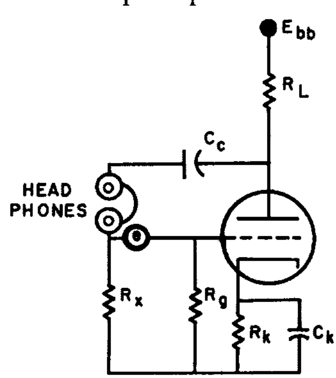


Fig. 3.30. Circuit for measuring the g_m of a vacuum tube.

3.3. A typical triode, for which $g_m = 1250\ \mu$mhos, $r_p = 10{,}000$ ohms, $R_k = 600$ ohms, $R_L = 10{,}000$ ohms, operates as a grounded cathode amplifier with R_k unbypassed. Calculate the effective transconductance and amplification factor in this mode of operation.

3.4. Design a pentode cathode follower circuit to have an output impedance of 72 ohms. The data for the tube to be used are as follows: $r_p = 100{,}000$ ohms; $E_{cc} = -2$ v; $E_d = 140$ v; $I_d = 7$ ma; $E_b = 200$ v; $g_m = 7700$ μmhos; $I_b = 21$ ma.

(a) Draw the circuit diagram of the amplifier.

(b) Calculate the cathode bias resistor and screen dropping resistor.

(c) Determine the values of the cathode and screen bypass capacitors if their reactance is not to exceed 1/10 of R_k and R_d at 20 c.

(d) Calculate the amount of unbypassed cathode resistance required.

(e) Calculate the voltage gain of the circuit and express it in volt logits.

3.5. Calculate the input capacitance of the preceding circuit if $C_{gk} = 5.0\ \mu\mu$f, $C_{gp} = 0.02\ \mu\mu$f, $C_{pk} = 8.0\ \mu\mu$f.

3.6. The circuit diagram in figure (3.31) shows a single tube phase splitting circuit that is frequently used with the push-pull amplifiers discussed in

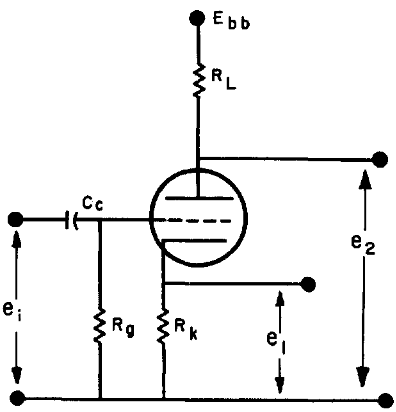


Fig. 3.31. Single tube phase splitting circuit.

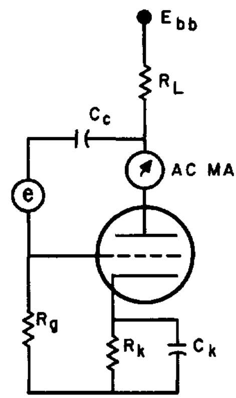


Fig. 3.32. Circuit for measuring the μ of a vacuum tube.

chapter 11. It provides two outputs from a single input that are 180° out of phase with each other. Draw the class A equivalent circuit, neglecting all interelectrode capacitances and assuming that C_c is a signal frequency short circuit, and assuming r_p to be much larger than R_L and R_k. Derive the two voltage gain equations. State the condition necessary to make the voltage gains equal.

3.7. The circuit diagram in figure (3.32) shows a method of measuring the voltage amplification factor μ of a vacuum tube. Proper operation is obtained when the signal current through the plate ammeter is zero. From the class A equivalent circuit derive an equation for μ in terms of R_L and R_g. Neglect interelectrode capacitances and assume that C_k and C_c appear as signal frequency short circuits.

3.8. Construct the class A equivalent circuits of all the amplifiers given in figures (3.16) and (3.17). Indicate all interelectrode and distributed wiring capacitances.

3.9. For a given pentode connected as a grounded cathode amplifier, the following data apply: $E_b = 285$ v; $E_d = 220$ v; $E_{cc} = -5$ v; $I_b = 2$ ma; $I_d = 0.5$ ma; $R_L = 5000$ ohms; $k = 0.31$; $g_m = 4000\ \mu$mhos; $r_p = 700{,}000$ ohms; $\mu_s = 10$; $r_s = 8000$ ohms.

(a) Calculate the values required for the screen dropping resistor and cathode bias resistor.

(b) Calculate the value required for E_{bb}.

(c) Calculate the values for the bypass capacitors if they are to have a reactance of 1/10 the resistance they are bypassing at 40 c.

(d) Calculate the effective transconductance at 20 c.

(e) Assume the screen resistor is perfectly bypassed at 20 c and calculate the effective transconductance in the presence of cathode degeneration only.

(f) Calculate the effective transconductance as in (e) if the amplifier is connected in the grounded grid configuration.

3.10. The diagram of a reactance tube is shown in figure (13.22a). This circuit is frequently used in frequency modulated transmitters. The input capacitance of the circuit is strongly capacitive by an amount approximately given by $C_{in} = g_m R_g C_g$. This is valid if the reactance of C_g is many times larger than R_g at the operating frequency. Assume that the radio frequency choke (RFC) is an open circuit and that C_c, C_k, and C_d are signal frequency short circuits. Derive the equation for C_{in}. How could C_{in} be made almost linearly dependent upon grid voltage?

Chapter 4

SINGLE STAGE VACUUM TUBE AMPLIFIERS

Most of the amplifiers in this chapter are of the nondegenerative, grounded cathode type. Although these circuits are common in practice, it might appear to be an unwarranted restriction of coverage in a general study of amplifiers. However, it was shown in chapter 3 that the gain functions of cathode followers, grounded grid, and degenerative amplifiers could all be expressed in terms of the gain function of the corresponding nondegenerative grounded cathode amplifier. Therefore the results of chapter 3 can be used to extend the results of this chapter to cover nearly any vacuum tube circuit that might come up for discussion.

The general gain function for any single vacuum tube amplifier was shown to be of the form

$$A = \pm g_m' Z_m$$

In essence, chapter 3 was a detailed study of the factors affecting g_m'. Chapter 4 is concerned exclusively with various circuit configurations that govern the characteristics of Z_m.

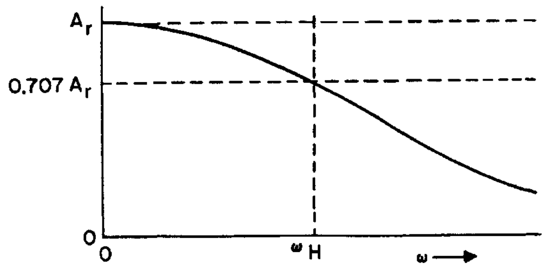
4.1. Criteria for Comparing Amplifiers, Steady State

From the material presented in chapter 3, and particularly from an examination of figures (3.16) and (3.17), it is clear that you have a wide variety of vacuum tube amplifier circuits at your disposal when designing an electronic system. At some point in your work you must make a choice between these circuits. For the selection to be made on a basis other than that of a purely random guess, criteria must exist for comparing the suitability of one circuit with that of another when a particular application is under consideration.

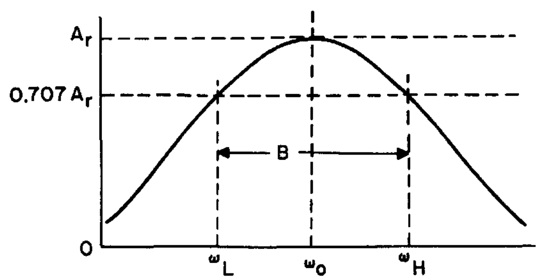
Generally speaking, class A voltage amplifiers are used in applications that permit the specification of various *figures of merit* for the circuits. These figures of merit provide the main criteria for comparing different amplifiers. However, the nature of the criteria so specified will obviously depend upon the performance characteristics of interest.

For example, in a great many cases, amplifiers are required to have a certain voltage gain vs. frequency characteristic over a specified frequency range. The nature of the phase response as a function of frequency is also of interest. In these cases involving the *steady state* frequency response characteristics, the relative merits of different amplifiers could be judged from the shapes of the gain and phase shift characteristic curves.

Fortunately, most amplifier circuits are of the minimum phase shift



(a) LOW PASS CHARACTERISTIC



(b) BAND PASS CHARACTERISTIC

Fig. 4.1. Steady state amplitude response characteristics.

type. As a consequence, the phase response can be determined directly from the amplitude response, and it is not necessary to determine a separate phase characteristic.[1] Therefore in nearly all practical cases it is sufficient to know the amplitude response alone, and criteria for steady state amplifier performance can be established in terms of this curve.

Two typical amplifier amplitude response characteristics are shown in figure (4.1), where $B =$ bandwidth in radians/sec $= \omega_H - \omega_L$; $\omega_H =$ upper cutoff frequency in radians/sec; $\omega_L =$ lower cutoff

[1] D. E. Thomas, "Tables of Phase Associated With a Semi-Infinite Unit Slope of Attenuation," *Bell System Tech. J.*, vol. 26, October, 1947, pp. 870–899.

frequency in radians/sec; A_r = reference gain or amplification. The two cutoff frequencies are nearly always defined as the frequencies at which the voltage gain is reduced to $1/\sqrt{2}$, or 0.707, of the reference gain A_r. Alternatively, this corresponds to the frequencies at which the voltage gain is down 1.5 vℓ (or 3 db in incorrect notation) from the reference value. Frequencies ω_H and ω_L are frequently called the *half power* frequencies, but these are nearly always incorrect designations despite their widespread use.

The bandwidth B is defined as the frequency difference, expressed in radians, between the upper and lower cutoff frequencies. This is a general definition, but if the amplitude response exhibits peaks and dips it is necessary to stipulate that no dip shall be more than 1.5 vℓ below any peak between the two cutoff frequencies.

Most amplifiers have a lower cutoff frequency. However, it is often so small compared with ω_H that the amplifier is essentially a low pass circuit. In such a case,

$$B = \omega_H - \omega_L \doteq \omega_H$$

Therefore the bandwidth of a low pass amplifier is practically equal to the upper cutoff frequency. Thus the terms *bandwidth* and *upper cutoff frequency* are frequently used interchangeably when discussing low pass amplifiers.

The other factor of interest noted on figure (4.1) and mentioned in the bandwidth definition, is the reference gain A_r of the amplifier. This is *not always* the maximum gain of the amplifier. In nearly every practical case the plate circuit contains reactive circuit elements. As a result, the voltage amplification function is nearly always frequency dependent and is a complex number. At some particular frequency the amplification function becomes a purely real or purely imaginary number. The particular value of the voltage gain at this frequency is designated in this text by the term *reference gain.*

Amplifier performance is often described by the factors just defined. In other words, the over-all quality of a bandpass amplifier can be represented in a general way by an amplifier figure of merit defined as F_a = bandpass amplifier figure of merit = (reference gain) $\times$ (bandwidth), or

$$F_a = A_r B \tag{4.1}$$

This is generally called the *gain-bandwidth product.* A similar figure of merit is used for the low pass amplifier. That is, F_a = low pass amplifier figure of merit = (reference gain) $\times$ (upper cutoff frequency),

or
$$F_a = A_r \omega_H \tag{4.2}$$

A *high pass* amplifier characteristic is shown in figure (4.2). In this case there is no commonly accepted figure of merit though one could be defined as $F_a' =$ high pass amplifier figure of merit $=$ (reference gain)/(lower cutoff frequency), or

$$F_a' = \frac{A_r}{\omega_L} \tag{4.3}$$

This is the *gain-cutoff frequency ratio.*

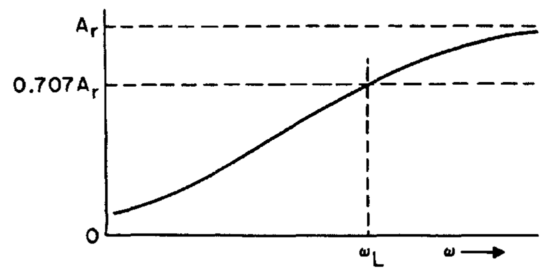


Fig. 4.2. High pass amplifier characteristic.

4.2. Transient Response Criteria

The figures of merit discussed in the preceding section are not suitable when the performance of pulse amplifiers is evaluated. Amplifiers of this type are found in radar and television receivers, facsimile, pulse-time modulation equipment, and many other devices used in

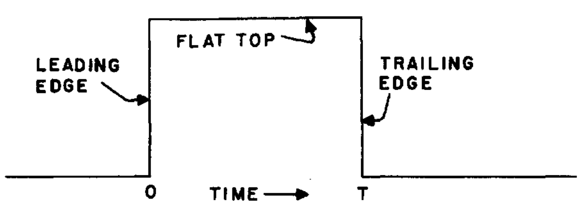


Fig. 4.3. Standard rectangular pulse.

communication and control services. In these cases the *time response* of the amplifier is of interest, and emphasis falls primarily upon the *shape* of the output signal relative to the input.

It is convenient to establish one particular type of pulse as a standard, and it is customary to use a rectangular pulse of the form shown in figure (4.3). Actually it does not make any difference what pulse shape is adopted as the standard, because the time response of a linear circuit to any excitation is determined by the poles of the transfer function and not by the excitation.

A comparison of the performance of one amplifier with that of another is generally based upon the nature of the response to the *edges* of the pulse and the response during the *flat top* of the pulse. The output of an amplifier resulting from the efforts of the circuit to reproduce an edge of a negative input pulse might appear as shown in figure (4.4). In addition to the *final value* of the output voltage from the amplifier,

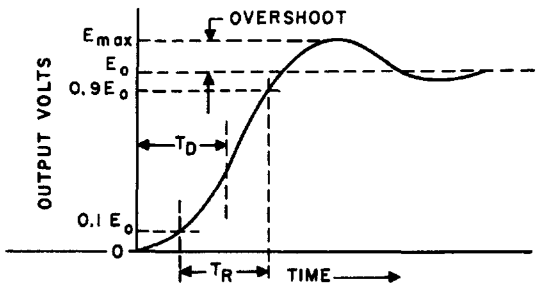


Fig. 4.4. Possible edge response.

other important factors shown on the figure are the *overshoot* γ, the *rise time* T_R, and the *delay time* T_D.

The overshoot is simply a measure of the extent to which the output exceeds its limiting final value. It is usually expressed as the *fractional* or *percentage overshoot* as

$$\% \gamma = \left(\frac{e_{\text{max}}}{e_{\text{final}}} - 1\right) 100\%$$

The rise time can be defined in a great many ways. One common specification is the 10–90% rise time.[2] This is simply a measure of the time required for the response to rise from 10% to 90% of its final value. This appears to be the most generally used definition. There is nothing particularly significant about these percentages; 5–95% could be used just as well, or any other combination of percentages adapted to the problem at hand. However, once a definition is specified, it should be used in all cases, otherwise comparative figures on the rise time will mean little. In any definition of this type, the actual transient response must be computed before the rise time can be determined.

There is another definition of rise time that has some use at times.

[2] See Valley and Wallman, *Vacuum Tube Amplifiers*, vol. 18, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, pp. 71–84.

It was developed by Elmore[3] and applies only to amplifiers that do *not* overshoot. In these cases, Elmore's rise time can be calculated directly from the gain function of the amplifier without actually plotting the time response. Also, the resulting values for the rise time are fairly close to the values computed from the 10–90% definition.

According to Elmore, the general gain function for any linear amplifier can be written in the general form

$$A(s) = \frac{1 + a_1 s + a_2 s^2 + \ldots}{1 + b_1 s + b_2 s^2 + \ldots} \tag{4.4}$$

The rise time given by Elmore is

$$T_R = \sqrt{2\pi[b_1^2 - a_1^2 + 2(a_2 - b_2)]} \tag{4.5}$$

Because the leading edge of the amplifier output pulse is not vertical, but has some finite slope, it is difficult to establish any method of indicating the *time delay* of the pulse. Several schemes could be and have been used, but none have achieved widespread adoption so far as I know. Fortunately, there are many cases in which this delay time is not important, and the topic will not be treated here. In cases where it is significant, it will be largely up to the designer to define delay time in a way that will permit an evaluation of the comparative merits of various amplifiers for use in a projected application.

It is clear that four factors are required to completely evaluate the edge response of one amplifier relative to another. These factors are: (1) reference gain A_r; (2) rise time T_R; (3) delay time T_D; (4) overshoot γ. It is convenient to lump the first two factors together into a figure of merit f_a, called the *gain/rise time ratio*. That is

$$f_a = \frac{\text{gain}}{\text{rise time}} = \frac{A_r}{T_R} \tag{4.6}$$

In this book the rise time used in this figure of merit will be the 10–90% rise time.

After the gain/rise time ratio has been given, the overshoot and time delay are specified separately. In some cases, figures of merit have been defined to include the gain, rise time, and overshoot.[4]

[3] W. C. Elmore, "Transient Response of Damped Linear Networks With Particular Regard to Wideband Amplifiers," *J. Appl. Phys.*, vol. 19, January, 1948, pp. 55–62.

[4] R. C. Palmer and L. Mautner, "A New Figure of Merit For the Transient Response of Video Amplifiers," *Proc. IRE*, vol. 37, September, 1949, pp. 1073–1077.

High fidelity transmission of the flat top of the pulse is opposed by three main circuits: (1) cathode bias circuit; (2) screen bias circuit; (3) gridleak-coupling capacitor circuit. The effects produced by these three circuits cause the output from the amplifier to *sag* away from its

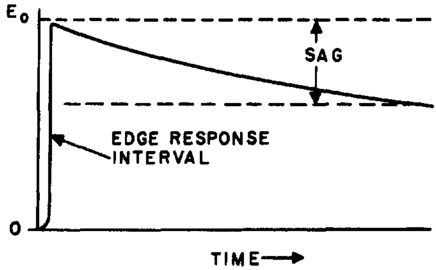


Fig. 4.5. Possible flat top response.

initial value. The performance of the amplifier is then described in terms of the *sag* and this is used in evaluating performance. A typical flat top response is shown in figure (4.5).

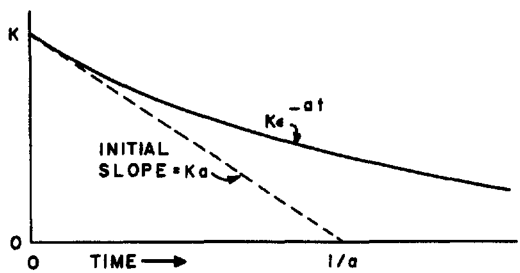


Fig. 4.6. Time response resulting in sag.

All the factors contributing toward the total sag have the form of simple exponential decay transients. For example, assume that a certain circuit has a voltage transfer function of the form

$$A(s) = K\frac{s}{s + a}$$

Now assume that the input, or driving function, is a unit step voltage, so that

$$e_i(t) = 1 \quad \text{for all} \quad t > 0$$

The Laplace transform of 1, according to pair number 1 in table 1, is $1/s$. Therefore the response function is

$$E_o(s) = E_i(s)A(s)$$
$$= \frac{K}{s} \cdot \frac{s}{s+a} = K\frac{1}{s+a}$$

The time response function $e_o(t)$ is the inverse transform of $E_o(s)$. It is given by pair number 8 in table 1. Hence

$$e_o(t) = K\varepsilon^{-at}$$

This is plotted in figure (4.6).

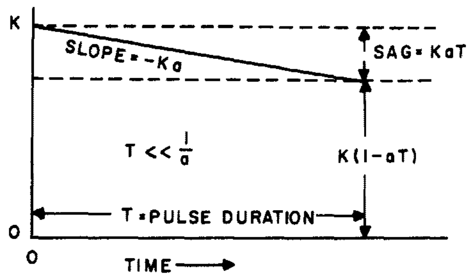


Fig. 4.7. Computation of sag; time scale from figure (4.6) greatly expanded.

The initial slope of this function, which is the slope of the curve at $t = 0$, is the derivative of $e_o(t)$ evaluated at $t = 0$. Therefore

$$\text{initial slope} = -aK \tag{4.7}$$

If the time interval of interest T is much less than the time constant $1/a$ the decay is nearly linear, with a slope equal to $-aK$ over the interval T. This is illustrated in figure (4.7), which is the same as figure (4.6) except that the time scale has been expanded to show the interval of interest. Under this assumption you can see that the sag is given approximately by sag $= KaT$.

$$\text{fractional sag} = \frac{\text{actual sag}}{K} = aT \tag{4.8}$$

4.3. Mutual Impedance

The labor required to determine the mutual impedances of networks used in the plate circuits of grounded cathode amplifiers can be simplified considerably by deriving the equation for the mutual impedance of a general, unloaded, pi section. Nearly all plate circuit impedance networks can be put into this form, or into the form of a single shunt impedance, or L section. The last two circuits are special cases of the general pi network. The equation to be derived merely provides a convenient method for systematizing the work following this section.

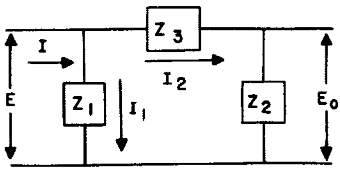


Fig. 4.8. General unloaded pi network.

Consider the general pi network shown in figure (4.8). The Kirchhoff voltage loop and current node equations are

$$E = I_1 Z_1 = I_2(Z_2 + Z_3) \quad (4.9)$$

$$I = I_1 + I_2 \quad (4.10)$$

Solve for the two branch currents from equation (4.9).

$$I_1 = E/Z_1; \qquad I_2 = \frac{E}{Z_2 + Z_3}$$

According to equation (4.10) the total current is

$$I = E\left(\frac{1}{Z_1} + \frac{1}{Z_2 + Z_3}\right) \quad (4.11)$$

and the output voltage E_o is

$$E_o = I_2 Z_2 = E\frac{Z_2}{Z_2 + Z_3} \quad (4.12)$$

Therefore, the mutual impedance is the ratio of this output voltage to the input current or

$$Z_m = \frac{E_o}{I} = \frac{Z_2/(Z_2 + Z_3)}{1/Z_1 + 1/(Z_2 + Z_3)}$$

Obtain the common denominator of this expression and then simplify. The result is

$$Z_m = \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3} \quad (4.13)$$

This is the general equation for the mutual impedance of an unloaded pi section, and will be used as the starting point in all the succeeding work connected with mutual impedance calculations.

The use of the preceding equation is conveniently illustrated by an example. Consider the high pass circuit shown in figure (4.9). By analogy to figure (4.8) you can see that

$$Z_1 = R_1; \qquad Z_2 = R_g; \qquad Z_3 = 1/sC_c$$

Therefore the mutual impedance in terms of complex frequency is

$$Z_m(s) = \frac{R_1 R_g}{(R_1 + R_g) + (1/sC_c)} = \frac{R_1 R_g s/(R_1 + R_g)}{s + 1/[(R_1 + R_g)C_c}$$

Now define some auxiliary terms as follows:

$$R = \frac{R_1 R_g}{R_1 + R_g}; \qquad \omega_1 = \frac{1}{(R_1 + R_g)C_c} \tag{4.14}$$

Consequently, the mutual impedance of the high pass network can be written

$$Z_m(s) = R\frac{s}{s + \omega_1} \tag{4.15}$$

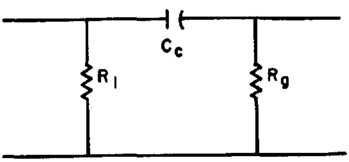


Fig. 4.9. High pass circuit.

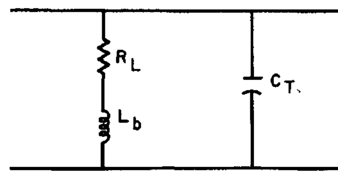


Fig. 4.10. Shunt peaking circuit.

One more example should be sufficient to illustrate the procedure. Consider a two-terminal network composed of shunt and series elements. In such a case the driving point impedance and mutual impedance are identically the same. The *shunt peaking* circuit shown in figure (4.10) is a typical case.

For the circuit diagram of figure (4.10),

$$Z_1(s) = R_L + sL_b; \qquad Z_2(s) = 1/sC_T; \qquad Z_3(s) = 0$$

Therefore the mutual impedance is

$$Z_m(s) = \frac{(R_L + sL_b)(1/sC_T)}{R_L + sL_b + (1/sC_T)} \tag{4.16}$$

Rearrange terms slightly so that the mutual impedance is

$$Z_m(s) = \frac{1}{C_T} \cdot \frac{s + R_L/L_b}{s^2 + R_L s/L_b + 1/L_b C_T}$$

Define two parameters as follows:

$$\omega_2 = \frac{1}{R_L C_T} \tag{4.17}$$

$$m = \frac{L_b}{R_L^2 C_T} = \frac{\omega_2 L_b}{R_L} = \text{circuit } Q \text{ at } \omega_2 \tag{4.18}$$

Therefore the mutual impedance can be expressed as

$$Z_m(s) = \frac{1}{C_T} \cdot \frac{s + \omega_2/m}{s^2 + \omega_2 s/m + \omega_2^2/m} \tag{4.19}$$

or

$$Z_m(s) = R_L \frac{1 + ms/\omega_2}{1 + s/\omega_2 + ms^2/\omega_2^2} \tag{4.20}$$

4.4. Resistance Coupled Amplifier, Equivalent Circuits

The circuit diagram of a grounded cathode *resistance coupled* amplifier is shown in figure (4.11). In the most general case the output

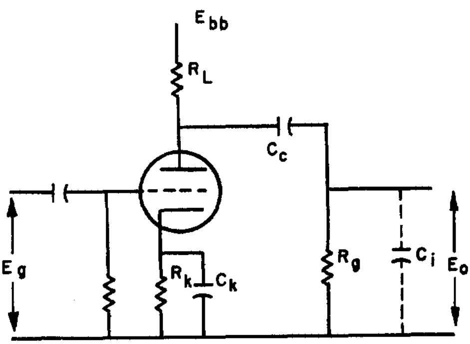


Fig. 4.11. Resistance coupled amplifier.

of the amplifier is connected to the input of some other circuit, another amplifier for example. The input impedance of the loading circuit is usually predominantly capacitive; therefore a capacitance C_i is shown across the output terminals of the amplifier in figure (4.11) to account for this effect.

The equivalent plate circuit of the amplifier is constructed by the methods given in chapter 3. Replace the tube by its current source equivalent, assume perfect bypassing, and that the internal impedance of the power supply is either negligible or is included with R_L, the load resistance. As a result, the equivalent plate circuit of the amplifier appears in the form shown in figure (4.12). The dotted capacitances account for the distributed wiring capacitance C_w, the output capacitance C_o of the tube, and the input capacitance C_i of the succeeding circuit.

Figure (4.12) is too complicated to be analyzed conveniently by a direct evaluation of the mutual impedance. Instead, it is usually simplified into three different forms.

For example, assume that the frequency of the input signal is so high that the reactance of the coupling condenser C_c is very small compared with the resistance of the grid leak R_g. Hence C_c can be

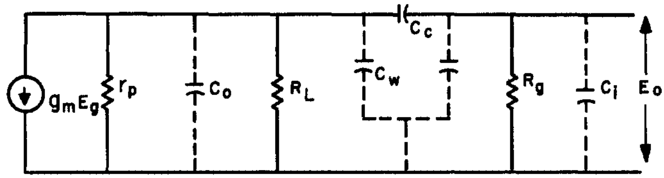


Fig. 4.12. Equivalent plate circuit of a resistance coupled amplifier.

replaced by a short circuit and the equivalent plate circuit simplified as shown in figure (4.13). This is a simple circuit, and you can easily calculate the mutual impedance for this *high frequency* case.

The circuit in figure (4.13) is also useful in determining the transient edge response of the amplifier when the input signal is a standard

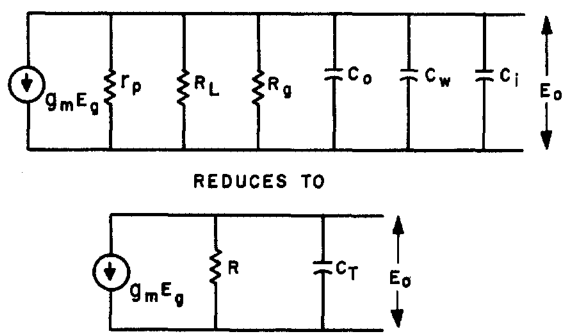


Fig. 4.13. High frequency and edge response equivalent circuit for a resistance coupled amplifier.

rectangular pulse. As the output voltage attempts to rise in response to the sudden change in grid voltage e_g, the shunt capacitances oppose the increase in output voltage, delaying its rise towards its final value. Therefore the high frequency equivalent circuit is also used to determine the transient edge response of the amplifier.

Another special form of the general equivalent circuit can be derived by assuming that the signal frequency is so low that the shunting effects of the three shunt capacitances are negligible compared with the effects produced by the three shunt resistances; thus the shunt capacitances can be treated as open circuits and omitted from the equivalent circuit. The coupling capacitor C_c now has a reactance of the same order of magnitude as the grid leak R_g. Hence the *low frequency* equivalent circuit has the form shown in figure (4.14). The mutual impedance of this circuit was computed as an example in section (4.3) and the result is given in equation (4.15).

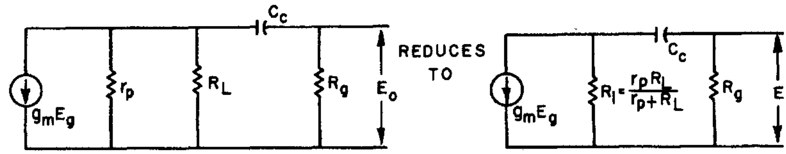


Fig. 4.14. Resistance coupled amplifier equivalent circuit for low frequency and flat top response.

This equivalent circuit is also useful in evaluating the transient response characteristics. After the shunt capacitance is fully charged during the initial edge response transient, the coupling capacitor commences charging. The extent to which the flat top of the pulse is faithfully delivered to the output terminals depends upon the rate of

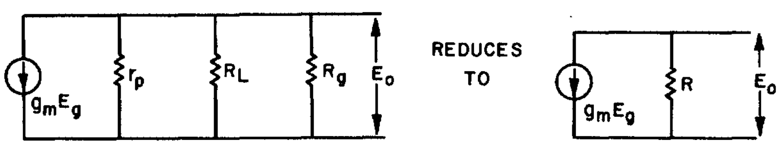


Fig. 4.15. Mid-frequency equivalent circuit of a resistance coupled amplifier.

charge of the coupling capacitor. If the rate is high (rapid charging), the pulse top sags away from its initial value. Therefore this low frequency equivalent circuit is useful in investigating the flat top response of the amplifier when the signal input is a standard rectangular pulse.

The third and final equivalent circuit can be derived by assuming that the frequency of the input signal is so high that the reactance of the coupling capacitor is virtually zero. At the same time, the frequency is assumed to be so low that the reactance of the shunt capacitance approaches that of an open circuit. The resulting equivalent circuit for this *mid-frequency* condition is also used to calculate the reference gain for the transient response. The circuit is shown in figure (4.15).

The mid-frequency equivalent circuit does not contain any reactive elements, and the gain calculated from this circuit is called the *reference gain* A_r.

4.5. Resistance Coupled Amplifier, Performance Characteristics

The three equivalent circuits of a resistance coupled amplifier were derived in section (4.4). The mutual impedances for each of these three circuits are computed with the aid of equation (4.13), the general equation for the mutual impedance of an unloaded pi section. Simply multiply the mutual impedances by $-g_m$ and the gain equation is obtained. The results of this process are summarized in table 4. You should be able to fill in the algebraic details without any difficulty.

TABLE 4

RESPONSE OF RESISTANCE COUPLED AMPLIFIERS

Edge response	Final value	Flat top response
High frequency response	Mid-frequency response	Low frequency response
$A(s) = -\frac{g_m}{C_T} \cdot \frac{1}{s + \omega_2}$ $A(s) = -g_m R \frac{1}{1 + s/\omega_2}$	$A(s) = -g_m R$	$A(s) = -g_m R \frac{s}{s + \omega_1}$ $A(s) = -g_m R \frac{1}{1 + \omega_1/s}$
where $\omega_2 = \frac{1}{RC_T}$ $C_T = C_o + C_w + C_i$	where $R = \frac{1}{1/r_p + 1/R_L + 1/R_g}$	where $\omega_1 = \frac{1}{(R_1 + R_g)C_c}$ $R_1 = \frac{1}{1/r_p + 1/R_L}$

The pole-zero diagrams corresponding to these gain functions are shown in figure (4.16). Both the high and low frequency gain functions have single poles so the transient responses of the two circuits are simple exponential functions.

Many of the factors used to specify the various performance criteria can now be determined. For example, from table 4 it is known that the reference gain of the amplifier in all cases is the mid-frequency value

$$A_r = g_m R \tag{4.21}$$

The edge response function has only a single pole, so the overshoot γ must be zero.

Because the amplifier does not overshoot, the rise time for the edge response can be calculated from the general gain function, using

Elmore's definition. The gain function coefficients are $a_1 = 0$; $a_2 = 0$; $b_1 = 1/\omega_2$; $b_2 = 0$. Therefore Elmore's rise time can be calculated from

$$T_R = \sqrt{2\pi[b_1^2 - a_1^2 + 2(a_2 - b_2)]}$$

or

$$T_R = 2.51RC_T = \frac{2.51}{\omega_2} \tag{4.22}$$

If you actually plot the transient response as shown in figure (4.17), you will find that the 10–90% rise time is

$$T_R = 2.2RC_T \tag{4.23}$$

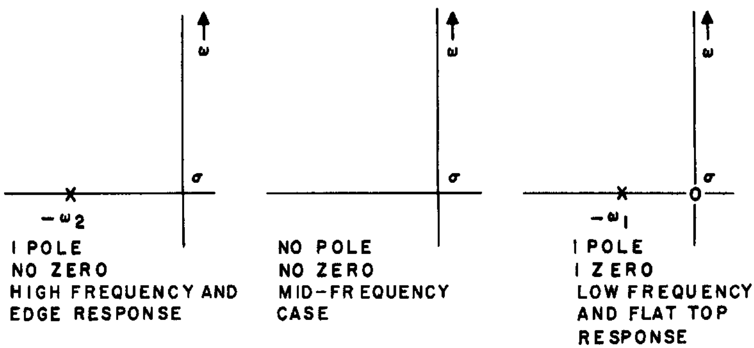


Fig. 4.16. Pole-zero diagrams for the gain function of a resistance coupled amplifier.

The cutoff frequencies are calculated directly from the steady state gain functions. Substitute $j\omega$ for s in the gain equations of table 4 and write

$$A_H(j\omega) = \frac{-g_mR}{1 + j\omega/\omega_2} = \frac{g_mR}{\sqrt{1 + (\omega/\omega_2)^2}}\underline{/\theta_H} \tag{4.24}$$

for the high frequency case and

$$A_L(j\omega) = \frac{-g_mR}{1 + \omega_1/j\omega} = \frac{g_mR}{\sqrt{1 + (\omega_1/\omega)^2}}\underline{/\theta_L} \tag{4.25}$$

for the low frequency case. At the cutoff frequencies, ω_H and ω_L, the gain will be down to

$$\frac{A_r}{\sqrt{2}} = \frac{g_mR}{\sqrt{2}} \tag{4.26}$$

Thus at these frequencies the radicals in equations (4.24) and (4.25) must be equal to $\sqrt{2}$, so that

$$1 + \left(\frac{\omega_H}{\omega_2}\right)^2 = 2 \quad \text{and} \quad 1 + \left(\frac{\omega_1}{\omega_L}\right)^2 = 2$$

Solve these two equations for ω_H and ω_L; the results are

$$\omega_H = \omega_2 = \frac{1}{RC_T} = \text{upper cutoff frequency} \tag{4.27}$$

$$\omega_L = \omega_1 = \frac{1}{(R_1 + R_g)C_c} = \text{lower cutoff frequency} \tag{4.28}$$

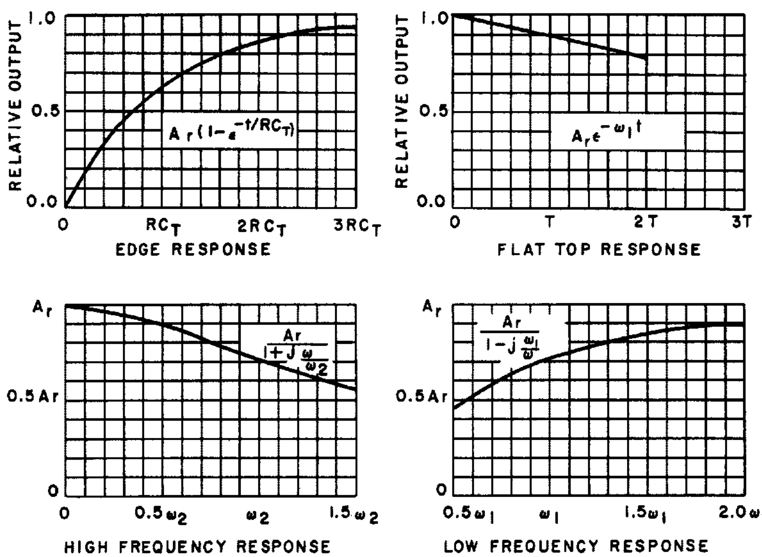


Fig. 4.17. Response characteristics of a resistance coupled amplifier.

According to the equation for the sag, as developed in section (4.2), sag $= aT$, where $a =$ time constant and $T =$ pulse duration. In the resistance coupled amplifier the sag produced by the gridleak-coupling capacitor circuit is

$$\text{sag} = \omega_1 T = \frac{T}{(R_1 + R_g)C_c} \tag{4.29}$$

The factors computed in this section, along with the corresponding figures of merit, are tabulated in table 5. The actual response characteristics are given in figure (4.17).

TABLE 5

RESISTANCE COUPLED AMPLIFIER PERFORMANCE DATA

Factor	Symbol	Value	Equation
Reference gain	A_r	$g_m R$	4.21
Upper cutoff frequency	ω_2	$\frac{1}{RC_T}$	4.27
Lower cutoff frequency	ω_1	$\frac{1}{(R_1 + R_g)C_c}$	4.28
Overshoot	γ	0	
Elmore's rise time	T_R	$2.51RC_T$	4.22
10–90% rise time	T_R	$2.2RC_T$	4.23
Gain-bandwidth product	F_a	$\frac{g_m}{C_T}$	
Gain/rise time ratio	f_a	$\frac{g_m}{2.2C_T}$	
Sag	sag	$\frac{T}{(R_1 + R_g)C_c}$	4.29

4.6. Figure of Merit for Tubes

If you are to design an amplifier you must take two preliminary steps before any actual computations are made. First, you must determine the kind of amplifier circuit to be used. The problem is more acute and your selection has the greater effect when the high frequency and edge response characteristics of the amplifier are of interest. In this case you should select a tube and circuit to provide a large enough figure of merit to meet the design requirements. The selection of the tube to be used is the second preliminary step.

From table 5, which listed the performance data for resistance coupled amplifiers, you can see that both the gain-bandwidth product and gain/rise time ratio depend upon the transconductance g_m of the tube and the total shunt capacitance C_T of the amplifier plate circuit. Except for the distributed wiring capacitances, the ratio of g_m to C_T is largely governed by the tube. It is true that the wiring capacitance may account for a large percentage of the total shunt capacitance, but the *variation* in wiring capacitance for the various interstage networks that might be used is *not* large compared with the total capacitance present.

Consequently, the high frequency and edge response figures of merit for the amplifier might be expressed as

$$F_a = k\frac{g_m}{C_t}$$

The first factor, k, is a constant, and is determined mainly by the amplifier *circuit*. The second factor is governed primarily by the *tube*, where C_t refers to the tube interelectrode capacitances. Such a breakdown of the figure of merit is based upon the assumption just discussed, that the variation in shunt capacitance from circuit to circuit is relatively small even though the total wiring capacitance may be quite large.

To aid you in selecting the proper tube for a given application, the second factor, g_m/C_t, is defined as the *tube figure of merit* F_t. Values of this figure for a number of different tubes selected at random from the tube manual are given in table 6. The Miller effect was approxi-

TABLE 6

FIGURES OF MERIT FOR REPRESENTATIVE TUBES

Tube type	Kind	C_{gk} or C_i ($\mu\mu$f)	C_{gp} ($\mu\mu$f)	C_{pk} or C_o ($\mu\mu$f)	g_m (μmhos)	g_m/C_t (10^6rad/sec)
6AB7	pentode	8.0	0.015	5.0	5000	385
6AC7	pentode	11.0	0.015	5.0	9000	562
6AG7	pentode	12.5	0.06	7.5	7700	385
6AK5	pentode	4.0	0.02	2.8	4300	630
6C6	pentode	5.0	0.007	6.5	1225	107
6J7	pentode	7.0	0.005	12.0	1225	61
6SJ7	pentode	6.0	0.005	7.0	1650	127
6SK7	pentode	6.0	0.003	7.0	2000	154
6U7	pentode	5.0	0.007	9.0	1600	114
6C5G	triode	4.4	2.2	12.0	2000	52
6J5	triode	3.4	3.4	3.6	2600	63
6K5	triode	2.4	2.0	3.6	1400	54

mated in these calculations by assuming that the gain of the stage following the one for which the calculation was made was 10. This has little effect when pentodes are involved, but the effect is quite pronounced for triodes. Thus the triode figures of merit can be considered to be only representative of what might actually be obtained.

4.7. Design of Resistance Coupled Amplifiers

Generally, you should make the shunt capacitance C_T of the amplifier as small as possible, because this increases the upper cutoff frequency ω_2, decreases the rise time T_R, and thereby increases the amplifier figure of merit. All these effects are usually desirable although there are cases where they are minor considerations.

The shunt capacitance can be minimized in several ways. The tube capacitances can be reduced by proper tube selection, and where possible, by using pentodes in preference to triodes. The wiring capacitance can be reduced by making all signal leads as short as possible, locating the plate load resistor, coupling capacitor, and gridleak away from the chassis, and by locating the bypass capacitors as close to the tube socket as possible.

Little advantage is gained from the high frequency and edge response viewpoints by reducing R. Although such a reduction increases the upper cutoff frequency and reduces the rise time, these advantages are exactly offset by a decrease in the reference gain. In other words, the amplifier figure of merit is not dependent upon R. In some cases it might be desirable to reduce the value of ω_2 by increasing R to get more gain. This is typical in audio amplifiers.

It is generally desirable, though not always, to use high transconductance tubes, because this results in larger figures of merit. However, high g_m tubes are more susceptible to cathode degeneration, and you must be careful that this does not overcome the anticipated gain improvement.

It is fairly customary to use the largest possible values for the gridleak resistor and coupling capacitor to reduce the sag and the lower cutoff frequency. A steady component of grid current flows at all times for three main reasons: (1) primary and secondary grid emission; (2) ionization of residual gases; (3) leakage resistance. Therefore the manufacturer of the tube generally specifies a maximum value for the grid leak resistor. Values of R_g in excess of this recommended value may cause a change in the operating point of the tube because of the change in grid bias brought about by the increased voltage drop across the grid leak. This might affect operation adversely.

The coupling capacitor is usually selected to give a certain sag or lower cutoff frequency ω_L. However, it cannot be made indefinitely larger for two reasons:

(1) The leakage resistance of the coupling capacitor increases as the capacitance is increased; therefore large values for the coupling capacitance can cause a change in grid bias as a result of d-c coupling through the leakage resistance.

(2) An increase in capacitance of the coupling capacitor is usually accompanied by an increase in its physical size if the same voltage rating is maintained. This, in turn, increases the wiring capacitance to ground and reduces the amplifier figure of merit.

The procedure for designing a resistance coupled amplifier can be summarized in a very general way as follows:

(1) Select a tube. The choice will depend upon a great many factors such as ruggedness, cost, uniformity of characteristics, as well as its electrical characteristics. If the high frequency and edge response characteristics are of interest, the tube figure of merit can be used as one basis for selection.

(2) Use the recommended value for R_g given by the tube manufacturer.

(3) The plate load resistor can be calculated from the rise time or upper cutoff frequency requirements.

(4) The coupling capacitor is determined from the specifications concerning the allowable sag or lower cutoff frequency.

(5) The cathode and screen resistors are determined from the bias requirements after the plate supply voltage has been found.

(6) The bypass capacitors are determined from the allowable sag and lower cutoff frequency requirements.

(7) The plate power supply voltage required is found by writing the Kirchhoff voltage loop equation around the plate circuit.

4.8. Shunt Peaked Video Amplifier

The resistance coupled amplifier discussed in the preceding sections is extensively used in voltage amplifiers. However, there are areas of use in which it will not provide a sufficiently wide bandwidth for a given value of reference gain, regardless of the tube used. The gain-bandwidth product and gain/rise time ratio of a resistance coupled amplifier are comparatively small, and this sets an upper limit on the high frequency and edge response characteristics obtainable. When this situation exists, the only thing left is to try other circuits that might have better performance characteristics.

Low pass amplifiers having upper cutoff frequencies in the megacycle range are called *video amplifiers.* The term developed because the early work on this type of amplifier was directed toward the amplification of the video, or picture, signal in television systems. The term persists even though the amplifier may be used in a system having little resemblance to television.

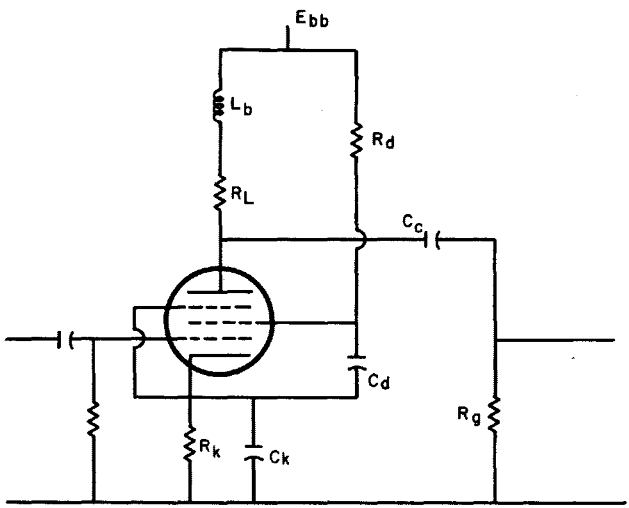


Fig. 4.18. Shunt peaked video amplifier.

Resistance coupled amplifiers can be, and are, used as video amplifiers. However, the relatively small values for the figure of merit restricts their use considerably and other kinds of amplifiers are often used.

There are many circuits with figures of merit exceeding those of resistance coupled amplifiers. The simplest such circuit is the *shunt peaked,* or *shunt compensated,* amplifier. The circuit diagram of such an amplifier is shown in figure (4.18). This is a conventional resistance coupled amplifier with the addition of a coil L_b in series with the plate load resistor R_L. This coil is called the *peaking inductor.*

Pentodes are nearly always used in shunt peaked amplifiers because they have much larger figures of merit than triodes. Because this new amplifier is to provide a larger figure of merit, you can see that tubes with high figures of merit should be used.

The equivalent plate circuit of the amplifier is constructed by the method outlined in the discussion of resistance coupled amplifiers.

The result, including the effects of interelectrode and shunt capacitances, appears as shown in figure (4.19).

For low and medium frequencies, the reactance of the peaking inductance L_b is negligible in comparison with the resistance of R_L.

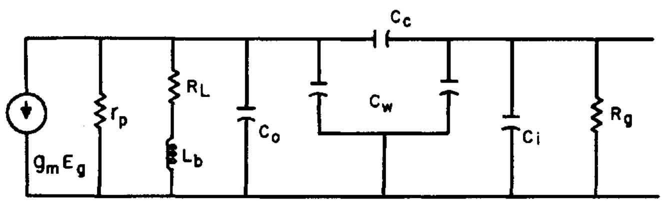
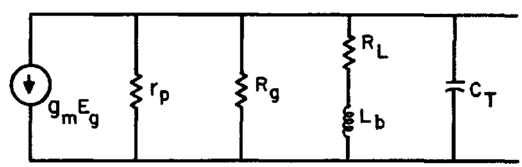


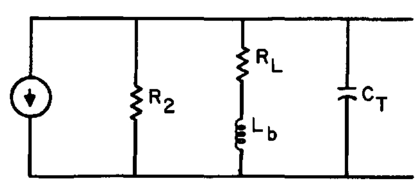
Fig. 4.19. Equivalent plate circuit of a shunt peaked video amplifier.

Therefore the coil can be replaced by a short circuit at these frequencies, and the circuit reduces to the same equivalent circuit as that for the resistance coupled amplifier in the low and mid-frequency ranges.



$C_T = C_o + C_w + C_i$ $\qquad R_2 = \dfrac{r_p R_g}{r_p + R_g}$

(a) FIRST EQUIVALENT CIRCUIT



(b) SECOND EQUIVALENT CIRCUIT

Fig. 4.20. High frequency and edge response equivalent circuits of a shunt peaked amplifier.

Thus the low frequency, mid-frequency, and flat top response characteristics of this amplifier are precisely the same as those of the resistance coupled amplifier and need not be discussed further. The only change of any interest lies in the nearly exclusive use of pentodes for shunt peaked amplifiers, and this does not change the form of the equations previously derived for the resistance coupled amplifier.

The main factors of interest are the high frequency and edge response characteristics of the circuit. If the operating frequency is high enough so that the reactance of the coupling capacitor is practically zero, the equivalent plate circuit of figure (4.20a) is obtained from figure (4.19). The circuit can be further simplified as shown in figure (4.20b) by combining r_p and R_g into an equivalent resistance R_2.

The mutual impedance of this network is easily written as the parallel combination of R_2, $1/sC_T$, and $(R_L + sL_b)$. After a little manipulation you should have no trouble in writing

$$Z_m(s) = \frac{1}{C_T} \cdot \frac{s + R_L L_b}{s^2 + (R_L/L_b + 1/R_2 C_T)s + (1 + R_L/R_2)/L_b C_T} \tag{4.30}$$

or
$$Z_m(s) = R \frac{1 + sL_b/R_L}{1 + s[RC_T + L_b/(R_2 + R_L)] + s^2 L_b R C_T/R_L} \tag{4.31}$$

where
$$R = \frac{R_2 R_L}{R_2 + R_L} = \frac{1}{1/r_p + 1/R_L + 1/R_g} \tag{4.32}$$

$$RC_T = \frac{1}{\omega_2} \tag{4.33}$$

where $\omega_2 =$ upper cutoff frequency of the amplifier without L_b. Now define a new term, to be called the *peaking parameter m*, as

$$m = \frac{\omega_2 L_b}{R_L} = Q \tag{4.34}$$

of R_L and L_b at the frequency ω_2. Therefore the preceding equations can be written in terms of this peaking parameter as

$$Z_m(s) = \frac{1}{C_T} \cdot \frac{s + \omega_2/m}{s^2 + (\omega_2/m + R\omega_2/R_2)s + \omega_2^2/m} \tag{4.35}$$

$$Z_m(s) = R \frac{1 + ms/\omega_2}{1 + (1/\omega_2 + mR/\omega_2 R_2)s + ms^2/\omega_2} \tag{4.36}$$

In a great many cases the following inequality exists:

$$m \ll \frac{R_2}{R} = 1 + \frac{R_2}{R_L} \tag{4.37}$$

When this inequality is valid, the preceding equations for the mutual impedance reduce to

$$Z_m(s) \doteq \frac{1}{C_T} \cdot \frac{s + \omega_2/m}{s^2 + \omega_2 s/m + \omega_2^2/m} \tag{4.38}$$

$$Z_m(s) \doteq R \frac{1 + ms/\omega_2}{1 + s/\omega_2 + ms^2/\omega_2^2} \tag{4.39}$$

The approximation is valid in a large number of cases and the simplified forms for the impedance function will be used in all the subsequent work in this text. However, cases may arise where the assumption is invalid, so you should be careful in using formulas and graphs developed for the approximate case.

The gain function of the amplifier is quickly obtained by the simple process of multiplying the impedance functions by $-g_m$. The basic information necessary to evaluate the various performance characteristics of the amplifier is now on hand.

4.9. Shunt Peaked Amplifier, Performance Characteristics

The transient response characteristics of the shunt peaked amplifier are governed by the locations of the poles of the amplifier transfer function. The poles are equal to the roots of the denominator of equation (4.38), or the roots of

$$s^2 + \frac{\omega_2}{m} s + \frac{\omega_2^2}{m} = 0$$

By the quadratic formula, the roots of this equation are

$$s_{1,2} = \frac{-\omega_2}{2m} \pm \sqrt{\left(\frac{\omega_2}{2m}\right)^2 - \frac{\omega_2^2}{m}}$$

Factor out the common $\omega_2/2m$ factor and write

$$s_{1,2} = -\frac{\omega_2}{2m}(1 \pm \sqrt{1 - 4m}) \tag{4.41}$$

If you examine equation (4.41) you will see that the poles will be *real* and *different* as long as the peaking parameter m is less than 0.25. In this case the transient response is overdamped and does not exhibit overshoot.

When the peaking parameter is exactly equal to 0.25, a *second-order pole* exists at $s = -\omega_2/2m$. Therefore the time response is critically damped and still does not overshoot.

However, if m is greater than 0.25, the two poles are *complex conjugates*; the response is then oscillatory and exhibits overshoot.

In every case there is a zero at $s = -\omega_2/m$.

The pole-zero diagrams and corresponding transient response characteristics for three different values of the peaking parameter are shown in figure (4.21).

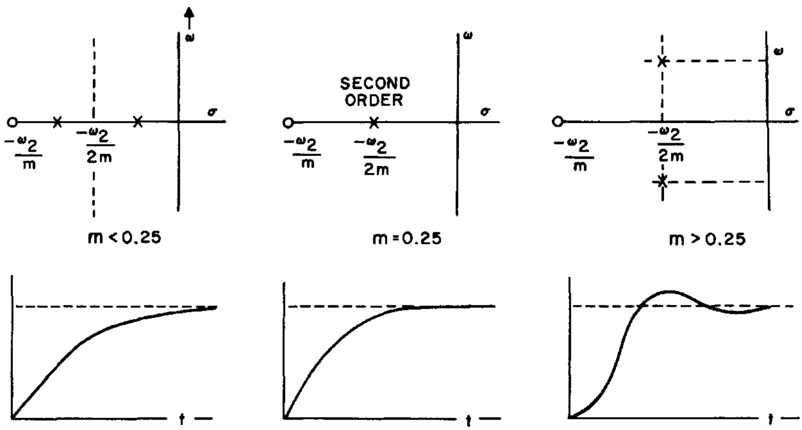


Fig. 4.21. Edge response characteristics of a shunt peaked amplifier when driven by a negative going stop function.

There is no overshoot as long as the peaking parameter is equal to or less than 0.25. Therefore Elmore's rise time formula can be applied in these few special cases. The most effective method of determining the rise time is actually to calculate the transient response of the circuit for a number of different values of the peaking parameter and then compute the 10–90% rise time from plots of the time response. The overshoot can be determined at the same time. This can hardly be called convenient, but it is about the only procedure that can be followed.

A number of responses for various values of the peaking parameter are shown in figure (4.22). From this, the data are taken to plot the curves in figure (4.23). This figure shows the rise time and overshoot as a function of the peaking parameter.

Because the overshoot becomes excessive for most of the usual applications when the peaking parameter exceeds 0.6, values of the overshoot and rise time are not plotted beyond this point.

From figure (4.23) you can see that the rise time of a shunt peaked amplifier can be made much less than that of a resistance coupled amplifier. A value of m of about 0.55 will cut the rise time to half of that of a resistance coupled amplifier. Of course, this is accompanied by an overshoot of about 9%, and this may be excessive. However, as long as the peaking parameter has a value of 0.3 to 0.4, the rise time

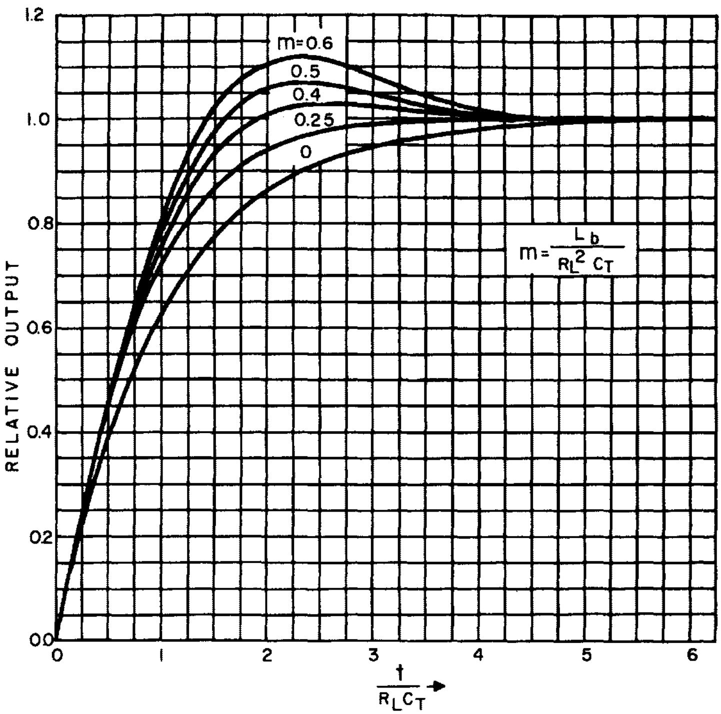


Fig. 4.22. Transient response of a shunt peaked video amplifier. (Courtesy R. C. Palmer, Allen B. DuMont Labs.)

is considerably less than for a resistance coupled amplifier, and the overshoot is only about 2% or less. Hence a shunt peaked amplifier can be designed to have a superior high frequency and edge response characteristic. The edge response superiority over the resistance coupled amplifier is clear from figure (4.23), but the steady state high frequency response improvement remains to be demonstrated.

The upper cutoff frequency of the shunt peaked amplifier, corresponding to any given value for the peaking parameter, can be obtained fairly easily. The general form of the mutual impedance was given in equation (4.39) as a ratio of two polynomials in s. Multiply this

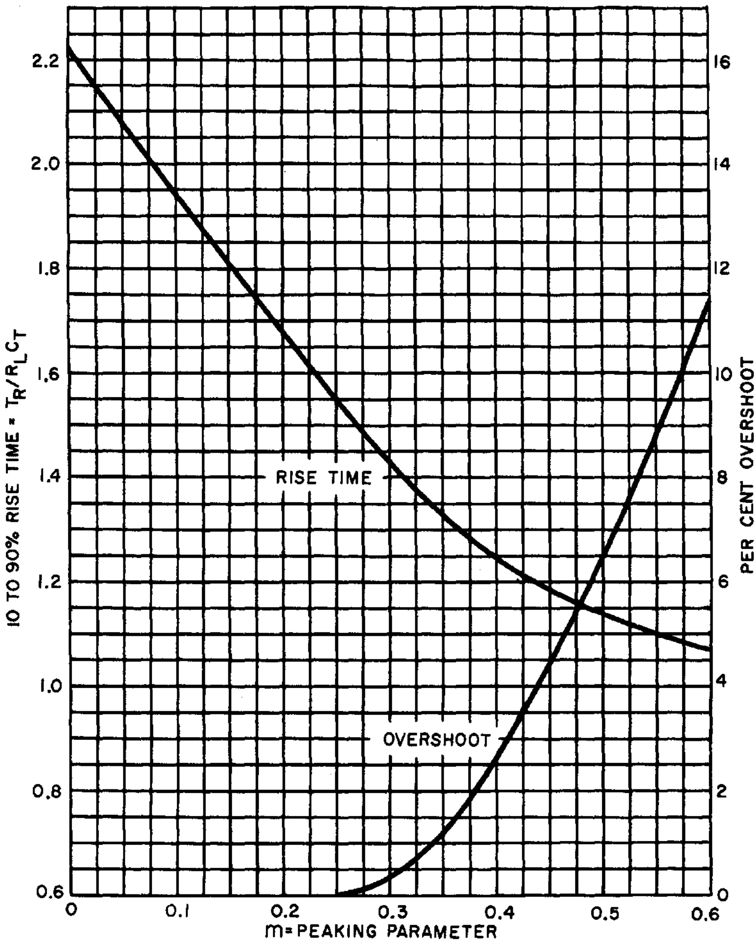


Fig. 4.23. Overshoot and rise time as a function of the peaking parameter in a shunt peaked video amplifier. (From R. C. Palmer and L. Mautner, "Transient Response of Video Amplifiers," *Proc. IRE*, vol. 39, Sept. 1949, p. 1075, courtesy R. C. Palmer, Allen B. DuMont Labs.)

function by $-g_m$ so that the amplifier gain function is

$$A(s) = -g_m R_L \frac{1 + ms/\omega_2}{1 + s/\omega_2 + ms^2/\omega_2^2} \tag{4.42}$$

This can be rewritten in terms of general coefficients as

$$A(s) = -g_m R_L \frac{1 + a_1 s}{1 + b_1 s + b_2 s^2} \tag{4.43}$$

In terms of the steady state frequency $j\omega$, this reduces to

$$A(j\omega) = -g_m R_L \frac{1 + j\omega a_1}{1 - b_2\omega^2 + j\omega b_1} \tag{4.44}$$

The magnitude of the amplifier gain function can then be written

$$|A(j\omega)| = g_m R_L \sqrt{\frac{1 + (\omega a_2)^2}{(1 - b_2\omega^2)^2 + (b_1\omega)^2}} \tag{4.45}$$

At zero frequency, the gain of the amplifier is $A_r = g_m R_L$. At the cutoff frequency ω_H, the gain must become $0.707A_r$. Because 0.707 is equal to $1/\sqrt{2}$, then when the operating frequency is equal to the cutoff frequency, the equation for the gain reduces to

$$\frac{1}{2} = \frac{1 + (a_1\omega_H)^2}{(1 - b_2\omega_H^2)^2 + (b_1\omega_H)^2} \tag{4.46}$$

Carry out all multiplications and collect like powers of ω_H. A quadratic equation of the following form should result:

$$\omega_H^4 - \frac{2b_2 + 2a_1^2 - b_1^2}{b_2^2}\omega_H^2 - \frac{1}{b_2^2} = 0$$

or

$$\omega_H^4 - X\omega_H^2 - \frac{1}{b_2^2} = 0$$

You can solve this for the upper cutoff frequency, so that

$$\omega_H = \sqrt{\frac{X}{2} + \sqrt{\frac{X^2}{4} + \frac{1}{b_2^2}}} \tag{4.47}$$

This is the general equation for the upper cutoff frequency where

$$X = \frac{2b_2 + 2a_1^2 - b_1^2}{b_2^2}$$

and $\qquad a_1 = m/\omega_2; \qquad b_1 = 1/\omega_2; \qquad b_2 = m/\omega_2^2$

When the values for all these coefficients are substituted into equation (4.47), the result is

$$\omega_H = 0.707\frac{\omega_2}{m}\sqrt{2m + 2m^2 - 1 + \sqrt{(2m + 2m^2 - 1)^2 + 4m^2}}$$

A plot of this equation in terms of m and the parameter ω_2 appears in figure (4.24). It should be noted that ω_2 is the upper cutoff frequency of the amplifier before the peaking coil is added to the circuit. Hence it is the same as the ω_H of a resistance coupled amplifier having the same tube, R_L, and C_T. Figure (4.24) makes it clear that extensive improvements in the upper cutoff frequency can be made through the use of a peaking inductance. However, it appears that there is little advantage, as far as bandwidth is concerned, in making the peaking

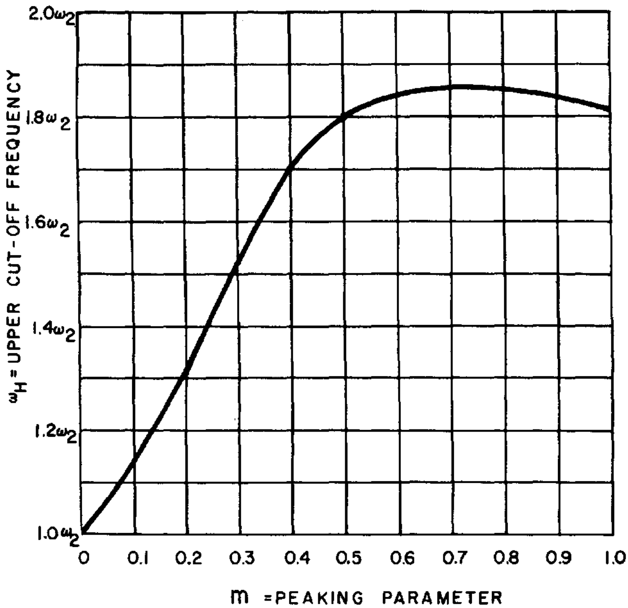


Fig. 4.24. Upper cutoff frequency of a shunt peaked amplifier as a function of the peaking parameter.

parameter much larger than about 0.5, because there is little increase in the upper cutoff frequency beyond this point.

This improvement in the upper cutoff frequency was accomplished without any reduction in the reference gain. Hence, the gain-bandwidth product is increased in the same degree as the upper cutoff frequency. From figure (4.24) it is clear that the figure of merit of a shunt peaked amplifier can be very nearly twice that of a corresponding resistance coupled amplifier.

The procedure for designing a shunt peaked amplifier can be summarized in a general way as follows:

(1) Select a tube with a high enough figure of merit to satisfy the design requirements.

(2) Evaluate C_T.

(3) Calculate the load resistance required to give the necessary reference gain.

(4) Determine the proper value for the peaking parameter by comparing the design requirements against figures (4.23) and (4.24).

(5) Calculate the peaking inductance $L_b = mR_L^2C_T$.

(6) The remainder of the circuit is designed as a resistance coupled amplifier.

4.10. Series Peaked Amplifier, Gain Equation

As long as two-terminal networks are used in the plate circuits of video amplifiers, the shunt peaked amplifier provides about the best high frequency and edge response characteristics available. However,

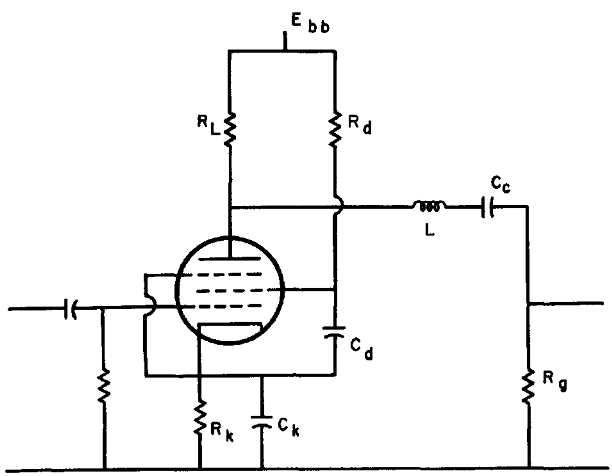


Fig. 4.25. Series peaked video amplifier.

if the simplicity of the shunt peaked circuit is without particular advantage and if more complex three- or four-terminal networks are permissible, it is possible to develop amplifiers with higher figures of merit than were obtained from the shunt peaked amplifier.

One of the simplest three-terminal low pass networks used in amplifier plate circuits is obtained by inserting an inductance in series with the coupling capacitor of an otherwise conventional resistance coupled amplifier. The resulting circuit, shown in figure (4.25), is called a *series peaked* video amplifier.

The reactance of the peaking inductance is negligible, compared to the resistance of R_L, in the middle and low frequency ranges. Therefore the equivalent circuits in these frequency ranges are exactly the same as those for the resistance coupled amplifier and no further discussion is required.

The reference gain is given, as in all previous cases, by the general equation

$$A_r = g_m R \tag{4.48}$$

where

$$R = \frac{1}{1/r_p + 1/R_L + 1/R_g} \tag{4.49}$$

Because a pentode is nearly always used in series peaked video amplifiers,

$$r_p \gg R_L \quad \text{and} \quad R_g \gg R_L \tag{4.50}$$

Consequently, the reference gain is approximately

$$A_r \doteq g_m R_L \tag{4.51}$$

The high frequency and edge response characteristics are calculated from the high frequency equivalent circuit shown in figure (4.26).

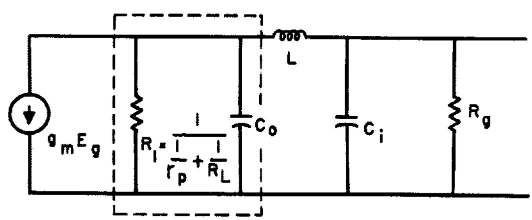


Fig. 4.26. High frequency and edge response equivalent circuit of a series peaked amplifier.

This circuit was constructed by replacing the coupling capacitor with a short circuit, combining r_p and R_L to form an equivalent resistor R_1, and showing the output and distributed capacitances together with whatever input capacitance may be associated with the following amplifier stage. The capacitance C_o includes the output capacitance of the tube plus the distributed wiring capacitance on the plate side of the peaking coil. Capacitance C_i includes the input capacitance of the circuit connected across the output terminals of the amplifier plus the distributed wiring capacitance on the output side of the peaking inductance L.

The voltage amplification can now be evaluated in terms of the mutual impedance of this passive network. According to the general equation for the mutual impedance of an unloaded pi section,

$$Z_m = \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3} \tag{4.52}$$

For the equivalent plate circuit in figure (4.26), these impedances have the following values:

$$Z_1(s) = \frac{R_1(1/sC_o)}{R_1 + 1/sC_o} = \frac{R_1}{s(R_1 C_o) + 1} \tag{4.53}$$

$$Z_2(s) = \frac{R_g(1/sC_i)}{R_g + 1/sC_i} = \frac{R_g}{s(R_g C_i) + 1} \tag{4.54}$$

$$Z_3(s) = sL \tag{4.55}$$

In nearly all practical cases of interest, the magnitude of the gridleak resistance R_g is very large compared with the reactance of the shunt capacitance C_i. Therefore

$$Z_2(s) \doteq 1/sC_i \tag{4.56}$$

Substitute equations (4.53), (4.55), and (4.56) into the general equation for the mutual impedance; multiply the result by $-g_m$, and the gain function is

$$A(s) \doteq -g_m \frac{[R_1/(sR_1C_o + 1)]\,(1/sC_i)}{R_1/(sR_1C_o + 1) + 1/sC_i + sL} \tag{4.57}$$

This equation can be put into a more convenient form; obtain the common denominator and collect like powers of s. The result is then

$$A(s) = -g_m \frac{1/LC_oC_i}{s^3 + (1/R_1C_o)s^2 + (1/LC_o + 1/LC_i)s + (1/LC_i)\,(1/R_1C_o)} \tag{4.58}$$

In some cases the input capacitance C_i is approximately twice as large as the output capacitance C_o. In other cases they are approximately equal, and sometimes the relationship is just the reverse. However, the initial statement is possibly the most common. The capacitance ratio is affected somewhat by the position of the coupling capacitor relative to the peaking coil.

Assume that

$$C_i = 2C_o \tag{4.59}$$

so that the preceding gain equation reduces to

$$A(s) \doteq -g_m \frac{1/2LC_o^2}{s^3 + (1/R_1C_o)s^2 + (3/2LC_o)s + (1/LC_o)\,(1/2R_1C_o)} \tag{4.60}$$

A familiar parameter can now be defined: ω_2 = upper cutoff frequency of the resistance coupled amplifier before the peaking inductance is added to the circuit.

$$\omega_2 = \frac{1}{R_1C_T} = \frac{1}{R_1(C_o + C_i)} = \frac{1}{3R_1C_o} \tag{4.61}$$

Now define a *peaking parameter K*, similar to the *m* used for the shunt peaked amplifier.

$$K \overset{\Delta}{=} \frac{3}{2} \cdot \frac{\omega_2 L}{R_1} \quad \text{where} \quad R_1 = \frac{r_p R_L}{r_p + R_L} \tag{4.62}$$

If these two factors are substituted into equation (4.60) for the voltage gain of the circuit, the result is

$$A(s) \doteq -g_m R_1 \frac{27\omega_2^3/4K}{s^3 + 3\omega_2 s^2 + 27\omega_2^2 s/4K + 27\omega_2^3/4K} \tag{4.63}$$

This is the general equation for the voltage amplification of a series peaked amplifier for which the assumed 2 : 1 capacitance ratio exists. If the ratio is reversed, a similar equation results.

The high frequency and edge response characteristics of the amplifier are discussed in the next section.

4.11. Series Peaked Amplifier, Response Characteristics

The main reason for developing the series peaked amplifier is to increase the figure of merit above that obtainable from a shunt peaked or resistance coupled amplifier. To accomplish this increase, the peaking parameter K must be selected so that the circuit is slightly oscillatory and overshoots somewhat. Therefore Elmore's rise time cannot be calculated from the general gain function derived in the preceding section.

The foregoing statement may be verified by arbitrarily assuming a representative value for the peaking parameter and then determining the pole locations. The amplifier has three poles and *one* of them *must be real.* The real pole can be found by trial and error synthetic division. The two remaining poles, whether real or complex, are found by using the quadratic formula on the quotient remaining after the real root has been divided out.

For example, when $K = 2$, this process indicates that the gain function has the following poles:

$$s_1 = -2.161\omega_2; \qquad s_2 = -\omega_2(0.419 - j1.18);$$
$$s_3 = -\omega_2(0.419 + j1.18)$$

A plot of these poles in the complex s plane is shown in figure (4.27a). Because two of the poles are complex conjugates, the transient response of the amplifier will be oscillatory and some overshoot will result.

As another example, assume that $K = 1.4$. In this case the poles turn out to be

$$s_1 = -1.808\omega_2; \qquad s_2 = -\omega_2(0.596 - j1.52);$$
$$s_3 = -\omega_2(0.596 + j1.52)$$

They are plotted in the complex plane of figure (4.27b). The response will again be oscillatory.

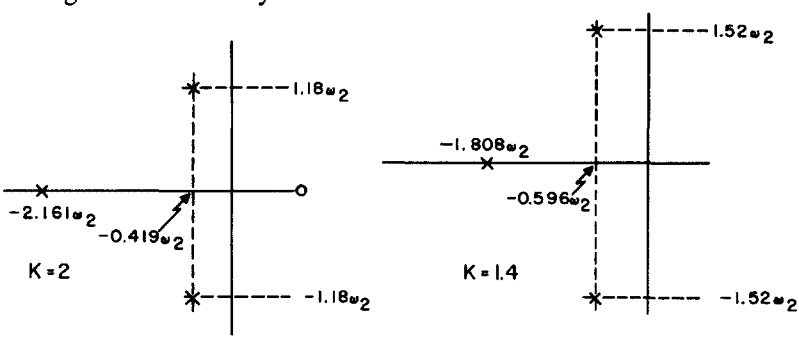


Fig. 4.27. Examples of pole locations of the series peaked amplifier.

Because of the presence of overshoot in the amplifier response, the 10–90% rise time must be determined from an actual plot of the amplifier response to a standard rectangular pulse input. The three poles are of the general form

$$s_1 = -\gamma; \qquad s_2 = -\alpha + j\beta; \qquad s_3 = -\alpha - j\beta$$

so that the general form of the response transform, when the excitation is a unit step function, can be written

$$E_o(s) = -g_m R_1 \frac{27\omega_2^3/4K}{s(s+\gamma)(s+\alpha+j\beta)(s+\alpha-j\beta)} \tag{4.64}$$

or, alternatively,

$$E_o(s) = -g_m R_1 \frac{27\omega_2^3/4K}{s(s+\gamma)[(s+\alpha)^2+\beta^2]} \tag{4.65}$$

The inverse transform of this function can be obtained from a partial fraction expansion or directly from a table of function-transform pairs.[5] The general form of the result is given in the following equation:

$$e_o(t) = -A_0[A_1 - A_2\varepsilon^{-\gamma t} + A_3\varepsilon^{-\alpha t}\sin(\beta t - \phi)] \tag{4.66}$$

The scale factor is

$$A_0 = g_m R_1 \frac{27}{4K} = \frac{6.75}{K} A_r \tag{4.67}$$

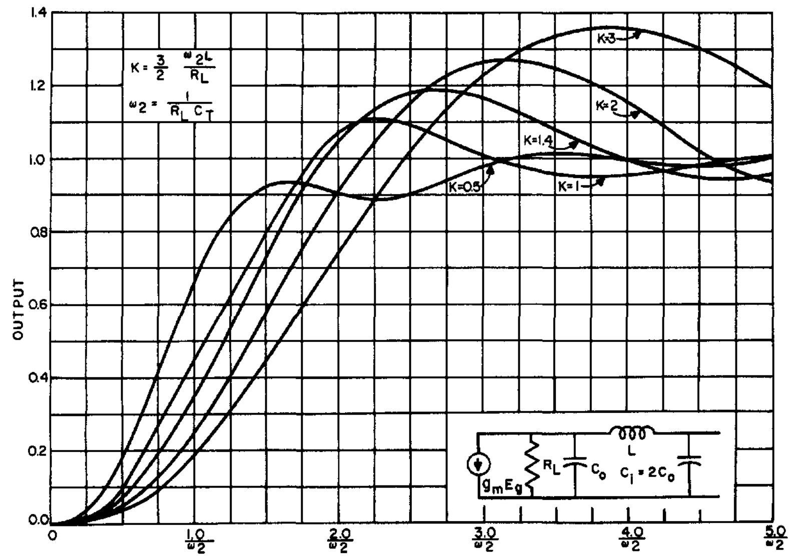


Fig. 4.28. Transient response of a series peaked amplifier with a 2 : 1 capacity ratio. (Computed by 1953 Electrical Engineering class, University of New Mexico.)

If the time response is actually plotted for various values of the peaking parameter K, as shown in figure (4.28), the corresponding rise times and overshoots can be evaluated. Typical results are given in figure (4.29). If you compare this figure with the corresponding one for the shunt peaked amplifier you will see that the series peaked amplifier will have a shorter rise time for the same reference gain.

The upper cutoff frequency ω_H of the series peaked amplifier is calculated by the same process as that followed for the shunt peaked

[5] See Gardner and Barnes, *Transients in Linear Systems*, John Wiley & Sons, Inc., New York, 1942, pair no. 1.319, p. 343.

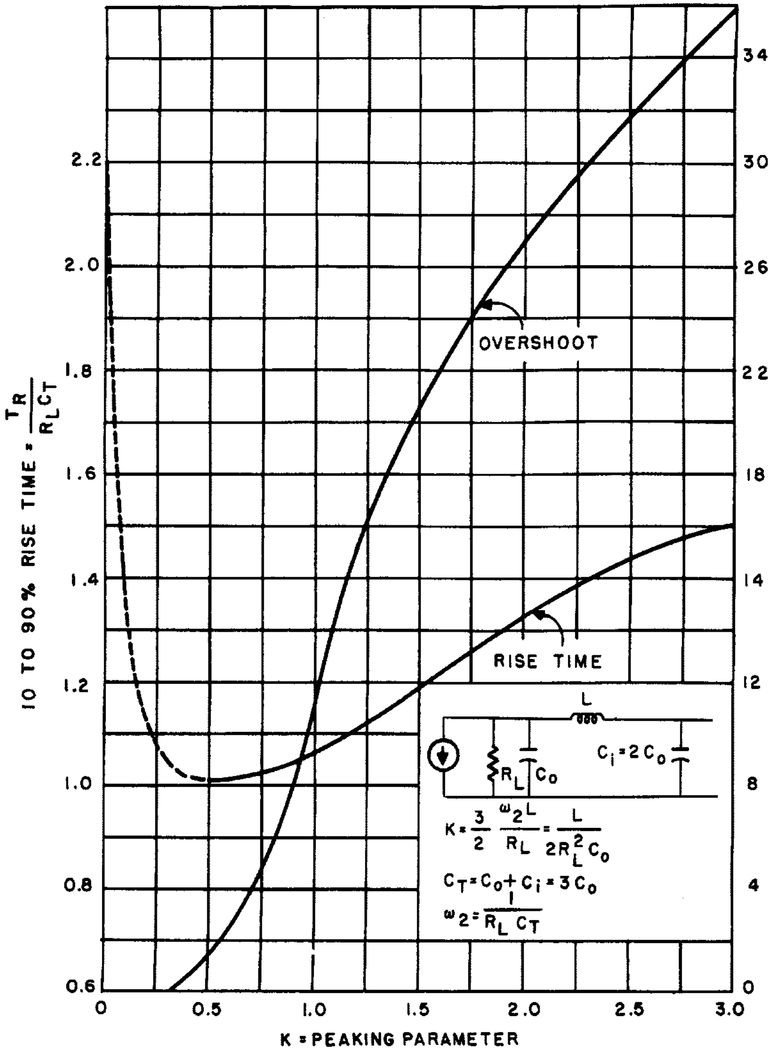


Fig. 4.29. Overshoot and 10%–90% rise time as a function of the peaking parameter in a series peaked amplifier. A 2 : 1 capacity ratio is assumed. Dotted section shows inferred values because points were not calculated.

amplifier. This makes it possible to compute and plot the upper cutoff frequency as a function of the peaking parameter K. The resulting curve is shown in figure (4.30). It is interesting to note that the upper cutoff frequency is three times the cutoff frequency of a resistance coupled amplifier if the peaking parameter is equal to 0.5. Any intermediate cutoff frequency can be obtained by selection of the value for K. However, the value of 0.5, or possibly a slightly larger value, appears to be the optimum choice.

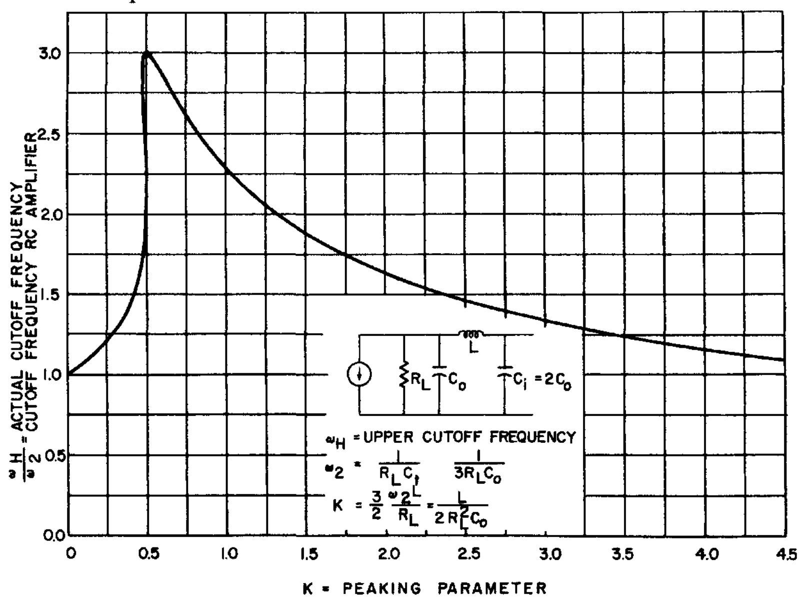


Fig. 4.30. Effect of the peaking parameter K on the upper cutoff frequency of a series peaked amplifier with a 2 : 1 capacity ratio. (Computed by L. C. Menasco.)

4.12. Low Frequency Compensation, No Degeneration

The discussion in the preceding articles was concerned with the high frequency and edge response characteristics of three different types of amplifiers. While these three circuits had different high frequency characteristics, their mid-frequency and low frequency characteristics were all exactly the same. Consequently, because the following discussion is concerned with the low frequency response, the remarks to be made will apply to all three circuits with equal validity.

For the resistance coupled amplifier it was shown that the sag and the lower cutoff frequency were controlled by the time constant in the low frequency equivalent plate circuit, neglecting degenerative effects. The time constant was found to be

$$(R_1 + R_g)C_c \quad \text{where} \quad R_1 = \frac{r_pR_L}{r_p + R_L} \tag{4.68}$$

In general, the tube chosen and the value of R_L are governed by the high frequency and gain requirements. Hence the value of R_1 is generally fixed in so far as the low frequency and sag response are concerned. Therefore the time constant is governed mainly by the values chosen for the gridleak resistance R_g and the coupling capacitance C_c. This is particularly true for a video amplifier because R_g is much larger than R_1 in such circuits.

If you want to reduce the sag and the lower cutoff frequency, you should make R_g and C_c as large as possible. However, as noted in section (4.7), there are definite maximum values that C_c and R_g can assume. If the largest possible values are used for these two circuit elements, and if the sag and lower cutoff frequency are still too large, a reduction in these two factors can be brought about only through some change in the circuitry.

An amplifier using one kind of *low frequency compensating circuit* is shown in figure (4.31), together with its class A equivalent plate circuit. Although a resistance coupled amplifier is shown in the figure, it could also be a shunt or series peaked amplifier. A pentode tube is shown, but it will frequently be a triode. It is assumed that *no* degeneration effects are involved and that the low frequency cutoff is caused solely by the effects of the coupling capacitor. The degenerative case will be treated later.

The general equation for the voltage amplification of the circuit is

$$A = -g_mZ_m = -g_m\frac{Z_1Z_2}{Z_1 + Z_2 + Z_3}$$

From the equivalent plate circuit in figure (4.31b) it is clear that

$$Z_1 = \frac{r_p\left(R_L + \dfrac{R_f}{sR_fC_f + 1}\right)}{r_p + R_L + \dfrac{R_f}{sR_fC_f + 1}} \tag{4.69}$$

$$Z_2 = R_g \tag{4.70}$$

$$Z_3 = \frac{1}{sC_c} \tag{4.71}$$

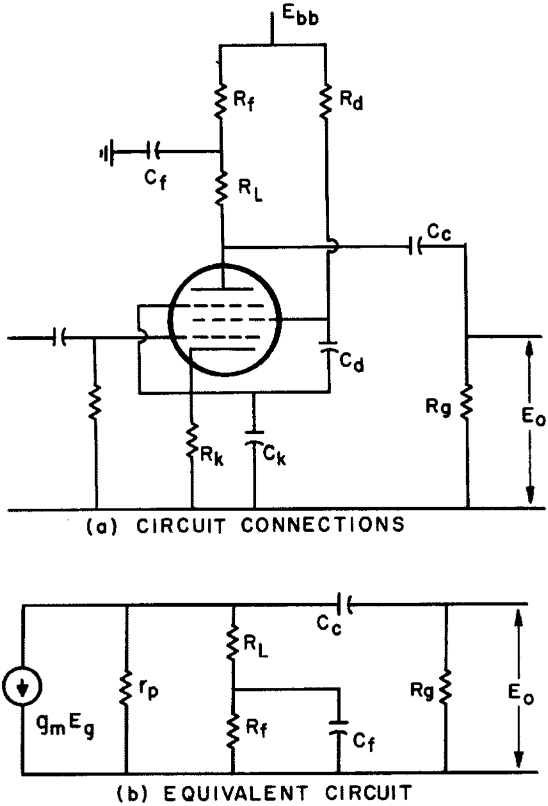


Fig. 4.31. Low frequency compensation of an amplifier.

These expressions for the three impedances should be substituted into the general voltage amplification equation. Although a considerable amount of tedious algebraic manipulation is involved, the end result for the voltage amplification can be written:

$$A = -A_r \frac{s(s + \omega_L + \omega_f)}{s^2 + s\left[\omega_1 + \omega_f\left(1 + \dfrac{R_f}{R_L + R_2}\right)\right] + \left(1 + \dfrac{R_f}{r_p + R_L}\right)\omega_1\omega_f} \tag{4.72}$$

where

$$\omega_1 = \frac{1}{(R_1 + R_g)C_c} \tag{4.73}$$

$$\omega_f = \frac{1}{R_f C_f} \tag{4.74}$$

$$\omega_L = \frac{1}{R_L C_f} \tag{4.75}$$

$$A_r = g_m R \tag{4.76}$$

$$R_1 = \frac{r_p R_L}{r_p + R_L} \tag{4.77}$$

$$R_2 = \frac{r_p R_g}{r_p + R_g} \tag{4.78}$$

$$R = \frac{R_1 R_g}{R_1 + R_g} \tag{4.79}$$

This function is usually too complex for general use: it can be simplified somewhat through the use of appropriate approximations. For example, if the amplifier tube is a triode, it is approximately true that

$$r_p \ll R_g \quad \text{and} \quad R_2 = r_p \tag{4.80}$$

Consequently, the term involving R_2 in equation (4.72) can be simplified as follows:

$$\omega_f\left(1 + \frac{R_f}{R_L + R_2}\right) \doteq \omega_f\left(1 + \frac{R_f}{r_p + R_L}\right) \tag{4.81}$$

This makes it possible to factor the amplifier gain function as given in (4.72) into the following form:

$$A \text{ (triode)} \doteq -A_r \frac{s(s + \omega_L + \omega_f)}{(s + \omega_1)\left[s + \omega_f\left(1 + \dfrac{R_f}{r_p + R_L}\right)\right]} \tag{4.82}$$

If the amplifier tube had been a pentode, a different result would be obtained. In this case it is generally true that

$$\left.\begin{aligned} &R_L \ll r_p; \quad R_L \ll R_g \quad \text{so that} \quad R_L \ll R_2 \\ &R_f \ll r_p; \quad R_f \ll R_g \quad \text{so that} \quad R_f \ll R_2 \end{aligned}\right. \tag{4.83}$$

Consequently, the two terms involving these various resistances in the general gain equation given in (4.72) can be simplified into the following forms:

$$\left.\begin{aligned} &\omega_f\left(1 + \frac{R_f}{R_L + R_2}\right) \doteq \omega_f\left(1 + \frac{R_f}{R_2}\right) \doteq \omega_f \\ \text{and} \qquad &\omega_f\left(1 + \frac{R_f}{r_p + R_L}\right) \doteq \omega_f\left(1 + \frac{R_f}{r_p}\right) \doteq \omega_f \end{aligned}\right\} \tag{4.84}$$

As a result, the amplifier gain function for a pentode can be factored and written

$$A\ (\text{pentode}) \doteq -A_r \frac{s(s+\omega_L+\omega_f)}{(s+\omega_1)\,(s+\omega_f)} \tag{4.85}$$

The equation for the voltage gain of the *triode* amplifier can be put into a form exactly like that for the pentode if a new parameter is defined as follows:

$$\omega_t = \omega_f\left(1+\frac{R_f}{r_p+R_L}\right) \tag{4.86}$$

Hence the gain equation for the triode amplifier is

$$A\ (\text{triode}) \doteq -A_r \frac{s(s+\omega_L+\omega_f)}{(s+\omega_1)\,(s+\omega_t)} \tag{4.87}$$

Note that this is exactly the same function as that obtained for the pentode amplifier except that ω_t in the triode case is ω_f for the pentode. Therefore, although the discussion that follows is based upon an analysis of the pentode function, an exactly parallel development can be easily worked out for the triode.

The purpose of the compensating circuit is to reduce the sag and lower cutoff frequency of the amplifier. Because both factors are affected by the same circuit constants, discussion of either applies to the other as well. The analysis of the sag response is somewhat easier than steady state analysis and that will be the viewpoint adopted here.

The output voltage from the amplifier is given by

$$E_o(s) = E_i(s)A(s)$$

If the input voltage as a function of time is a *negative* going unit step function, the transform of the input voltage is $E_i(s) = -1/s$. Therefore the transform of the output voltage from a pentode, low frequency compensated amplifier is approximately

$$E_o\ (\text{pentode}) \doteq \frac{A_r}{s}\cdot\frac{s(s+\omega_L+\omega_f)}{(s+\omega_1)\,(s+\omega_f)} \tag{4.88}$$

Cancel the common s in the numerator and denominator and expand the result into its partial fractions as

$$E_o\ (\text{pentode}) \doteq \frac{A}{s+\omega_1}+\frac{B}{s+\omega_f} \tag{4.89}$$

The coefficients A and B are readily computed by the method given in chapter 2. They are

$$A = A_r \frac{\omega_L + \omega_f - \omega_1}{\omega_f - \omega_1} \tag{4.90}$$

$$B = A_r \left(- \frac{\omega_L}{\omega_f - \omega_1} \right) \tag{4.91}$$

Therefore the inverse transform of $E_o(s)$ is

$$e_o(t) \doteq A_r \left(\frac{\omega_L + \omega_f - \omega_1}{\omega_f - \omega_1} \varepsilon^{-\omega_1 t} - \frac{\omega_L}{\omega_f - \omega_1} \varepsilon^{-\omega_f t} \right) \tag{4.92}$$

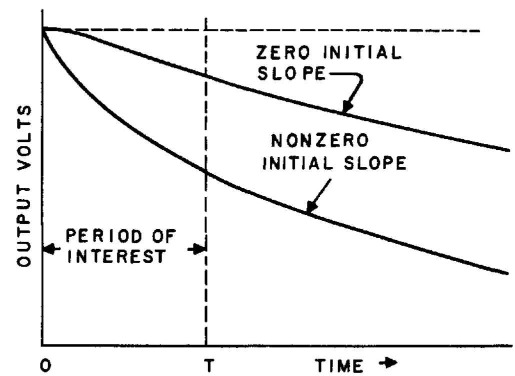


Fig. 4.32. Two possible flat top responses of a low frequency compensated amplifier.

This is the transient response of the pentode, low frequency compensated amplifier.

If the circuit constants in this amplifier can be adjusted so that the initial slope of the output voltage is zero, this will represent the *optimum* operating condition. In other words, this combination of circuit constants will cause the least sag and the lowest lower cutoff frequency possible. The reason for this should be apparent from figure (4.32).

To determine the condition for *optimum compensation* it is necessary to determine the requirements for zero initial slope. Therefore, differentiate equation (4.92) with respect to t, set the result equal to zero. This procedure yields the following expression:

$$\frac{(\omega_f - \omega_1)(\omega_1 - \omega_L)}{(\omega_f - \omega_1)} = 0 \tag{4.93}$$

After cancellation of the common factor in the numerator and denominator, the condition for optimum compensation is given by $(\omega_1 - \omega_L) = 0$, or

$$\omega_1 = \omega_L \tag{4.94}$$

Superficially, from equation (4.93) it might appear that the initial slope could also be made zero if $\omega_1 = \omega_f$. However, when this equality exists, the amplifier has a second-order pole at $-\omega_1$, and the partial fraction expansion used in the preceding development does not apply. If this case is actually worked out it will be found that optimum compensation (zero initial slope) is obtained if $\omega_1 = \omega_f = \omega_L$. This requires that $R_f = R_L$.

Another relationship is often used in the design of the compensating network. The following equality is established:

$$(\omega_L + \omega_f) = \omega_1 \tag{4.95}$$

As a result, the amplifier gain function becomes

$$A(s) = -A_r \frac{s}{s + \omega_f} \tag{4.96}$$

For the *uncompensated* amplifier the gain function is

$$A(s) = -A_r \frac{s}{s + \omega_1} \tag{4.97}$$

Therefore ω_1 = lower cutoff frequency, no compensation; ω_f = lower cutoff frequency with compensation. Clearly, if the constants of the compensating network are adjusted so that $\omega_f < \omega_1$, the low frequency and sag response characteristics of the compensated amplifier will be better than those of the uncompensated amplifier. In this book, compensation according to equation (4.95) is called *conventional* compensation. The amplifier characteristics with conventional compensation are inferior to those resulting from optimum compensation.

The procedure for designing a low frequency compensated amplifier can now be briefly summarized for the case of a pentode. A similar procedure can be developed for the triode case.

The suggested design procedure is outlined as follows:

(1) Values of R_g and C_c should be made as large as possible consistent with other requirements.

(2) Value of R_L is usually determined from the high frequency and edge response characteristics and may be assumed to be known at this point. Value of r_p is also known, because the tube should have been selected prior to this design.

(3) Calculate R_1 and ω_1 from equations (4.73) and (4.77).

(4) Specify the kind of compensation to be used.

(a) For optimum compensation:

(i) Set $\omega_1 = \omega_L$.

(ii) Thus C_f can be calculated from

$$C_f = C_c \frac{R_1 + R_g}{R_L}$$

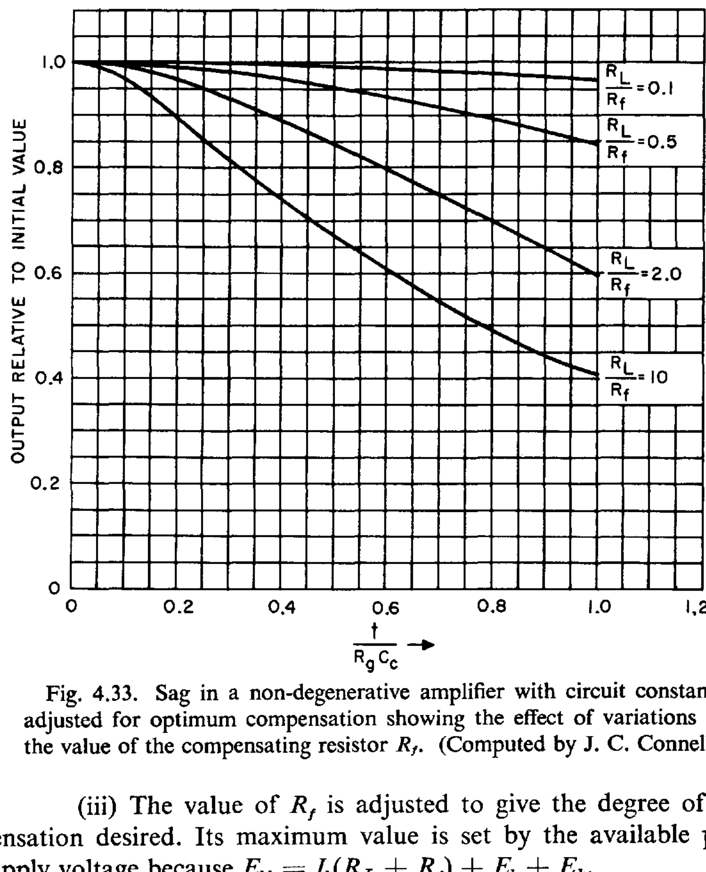


Fig. 4.33. Sag in a non-degenerative amplifier with circuit constants adjusted for optimum compensation showing the effect of variations in the value of the compensating resistor R_f. (Computed by J. C. Connell.)

(iii) The value of R_f is adjusted to give the degree of compensation desired. Its maximum value is set by the available power supply voltage because $E_{bb} = I_b(R_L + R_f) + E_b + E_k$.

(b) For conventional compensation:

(i) Set $(\omega_f + \omega_L) = \omega_1$.

(ii) In this case $\omega_f =$ lower cutoff frequency of the compensated amplifier and its value is presumably specified in the design requirements.

(iii) Hence both ω_1 and ω_f are known, and $\omega_L = \omega_1 - \omega_f$.

(iv) Therefore

$$C_f = \frac{1}{(\omega_1 - \omega_f)R_L}; \qquad R_f = \frac{1}{R_f C_f}$$

From step (iii) of the design for optimum compensation it is clear that there is an arbitrariness in the value of the compensating resistor R_f. The effect of variations in the size of R_f relative to R_L on the response under optimum conditions is shown by the curves of figure (4.33). Thus while zero initial slope is obtained in all cases shown in this figure, the least sag results when R_f is large compared with R_L. If long pulses are involved, R_f will have to be large to avoid excessive sag, and the power supply voltage will necessarily have to be fairly high.

4.13. Compensation for Degenerative Effects

The gain equation for the low frequency compensated amplifier was derived in the preceding section, and the result was found to have the following approximate form when a pentode tube is used:

$$A(s) \doteq -A_r \frac{s(s + \omega_L + \omega_f)}{(s + \omega_1)(s + \omega_f)}$$

If the cathode resistor of the pentode is imperfectly bypassed, the gain is altered because of the change in the effective transconductance. For a pentode, it was shown in equation (3.26) that this transconductance was

$$g_m' \doteq g_m \frac{s + \omega_k}{s + (1 + g_m R_k)\omega_k}$$

where $\omega_k = 1/R_k C_k$. Define another term:

$$\omega_3 = \omega_k(1 + g_m R_k) \tag{4.98}$$

Consequently, the gain equation for a pentode amplifier with low frequency compensation and imperfect cathode bypassing can be written

$$A(s) \doteq -A_r \frac{s(s + \omega_L + \omega_f)(s + \omega_k)}{(s + \omega_1)(s + \omega_3)(s + \omega_f)} \tag{4.99}$$

If the circuit is designed so that

$$\omega_k = \omega_1 \quad \text{and} \quad \omega_3 = \omega_L + \omega_f \tag{4.100}$$

then the resulting gain function for the ampifier is

$$A(s) \doteq -A_r \frac{1}{s + \omega_f} \tag{4.101}$$

The lower cutoff frequency of the amplifier is now ω_f. This is precisely the same result as that obtained in the preceding section for the case of conventional compensation.

The relationships specified by equations (4.100) are not necessarily the conditions for optimum compensation, but represent simply a convenient set of relationships between the circuit constants that will provide some measure of compensation for the effects of the coupling and cathode bias circuits.

The conditions in the circuit for zero initial slope and optimum compensation can be derived by the method outlined in the preceding section. Because of the increased number of variables, there is a greater degree of arbitrariness in the design procedure.

The design procedure for a pentode amplifier with conventional compensation might proceed as follows:

(1) Values of R_g, C_c, R_k, and R_L are determined by other design considerations or are known.

(2) According to equation (4.100), $\omega_k = \omega_1$, so that

$$C_k = C_c \frac{R_g}{R_k} \tag{4.102}$$

(3) Sufficient information is now available to compute ω_k and ω_3. Consequently, according to equation (4.100), $\omega_L = \omega_3 - \omega_f$, where ω_f will normally be specified as the desired lower cutoff frequency, so that it is assumed known. Thus

$$R_f = R_L \left(\frac{\omega_3}{\omega_f} - 1\right) \tag{4.103}$$

(4) As a result,

$$C_f = \frac{1}{\omega_f R_f}$$

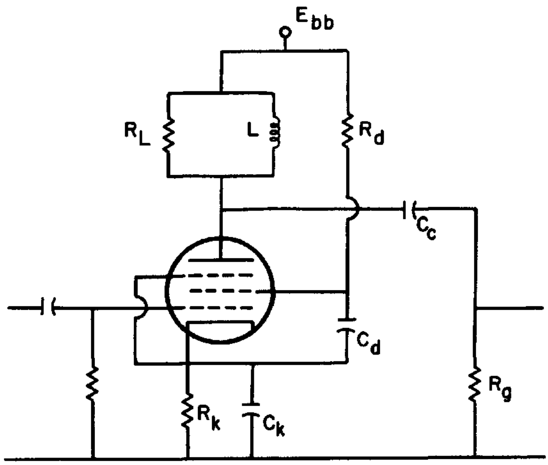
(5) The final value of ω_f, and therefore R_f and C_f, may be determined by the available power supply voltage E_{bb}.

If the amplifier is also subject to screen circuit degeneration, the effective transconductance of the tube is further altered and a new set of relationships must be derived. This is left as an exercise for the reader; it is not especially difficult.

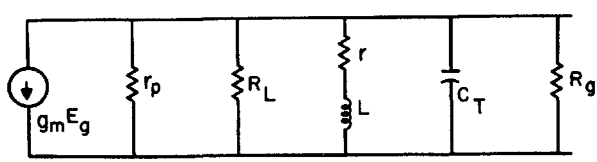
4.14. Single Tuned Amplifier

All the preceding material is concerned with high pass and low pass amplifiers; these are encountered in the amplification of audio and video signals. *Band pass* amplifiers, the subject to be discussed in the remainder of the chapter, are equally common, and find a wide range of use in the amplification of bands of frequencies occurring almost anywhere in the frequency spectrum up to a maximum of about 3000 mcps.

The simplest band pass amplifier in general use is the *single tuned circuit* shown in figure (4.34); the equivalent plate circuit is shown in the



(a) SINGLE TUNED AMPLIFIER CIRCUIT

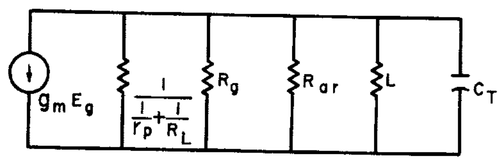


(b) EQUIVALENT CIRCUIT

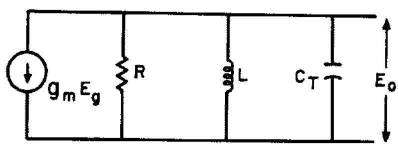
Fig. 4.34. Single tuned amplifier.

same figure. In the equivalent plate circuit the coupling capacitor has been replaced by a short circuit because the operating frequency is so high that the reactance of C_c is negligible compared with the resistance of R_g. The total shunt capacitance C_T is the sum of the output capacitance of the tube, the distributed wiring capacitance, and the input capacitance of the circuit connected to the output terminals of the amplifier.

The coil resistance is shown as r. A parallel tuned circuit consisting of a resistance and inductance in series, both in parallel with a capacitance, has an input impedance at resonance equal to $R_{ar} = L/rC_T$. If the coil Q is reasonably high, say 6 or 7, the original parallel tuned circuit can be closely approximated by a parallel circuit consisting of R_{ar}, L, and C_T all in parallel with one another. The equivalent plate circuit resulting from this approximation is shown in figure (4.35). Then, combining all shunt resistances into a single resistor R, the final equivalent circuit of figure (4.35b) results.



(a) ALTERNATIVE FORM



(b) FINAL FORM

Fig. 4.35. Equivalent circuits for a single tuned amplifier.

The foregoing approximation is *not* a serious limitation in practical circuits because the coil Q will usually be greater than 6 or 7.

Note that the inductance has been shown as the variable element in the parallel resonant circuit. The circuit could be tuned with equal ease by varying the capacitance. However, if C_T is adjustable, an extra physical capacitance must be added to the circuit. This would increase the total shunt capacitance and reduce the apparent resistance, $R_{ar} = L/rC_T$, of the circuit; this will lower the *circuit* Q. In some cases this might be unimportant, and the desirability of using a capacitor for tuning purposes might override other considerations. At high frequencies, the loss in Q *cannot* be tolerated, and it is more convenient to vary the inductance. This is usually done by varying the degree of insertion of a powdered iron or other special types of cores in the mutual flux path of the coil.

The amplification equation for the single tuned amplifier is determined by multiplying the mutual impedance of the circuit of figure (4.35b) by $-g_m$. The mutual impedance is equal to the input impedance of the circuit in this case and this is

$$Z_m(s) = \frac{1}{1/R + 1/sL + sC_T}$$

$$= \frac{1}{C_T} \cdot \frac{s}{s^2 + s/RC_T + 1/LC_T} \tag{4.104}$$

or

$$Z_m(s) = \frac{R}{1 + RC_T s + R/sL} \tag{4.105}$$

Consequently, using equation (4.105) and converting to the steady state case, the voltage amplification can be written

$$A(j\omega) = -g_m Z_m = -g_m R \frac{1}{1 + j(\omega RC_T - R/\omega L)} \tag{4.106}$$

At resonance, the Q of the parallel tuned circuit is

$$Q = \frac{R}{\omega_0 L} = \omega_0 RC_T \tag{4.107}$$

where $\omega_0 =$ parallel resonant frequency. Hence

$$\frac{R}{L} = \omega_0 Q \quad \text{and} \quad RC_T = \frac{Q}{\omega_0}$$

Substitute these relationships into the gain equation and the gain becomes

$$A(j\omega) = -g_m R \frac{1}{1 + jQ(\omega/\omega_0 - \omega_0/\omega)} \tag{4.108}$$

This is the general equation for the voltage amplification of a single tuned amplifier as a function of the frequency ω.

It is clear from equation (4.108) that the reference gain of the amplifier is $A_r = g_m R$ because the gain function is a real number equal to A_r when $\omega = \omega_0$.

The cutoff frequencies can be evaluated from equation (4.108) because the imaginary term in the denominator will be equal to ± 1 at the two cutoff frequencies. Hence, if $\omega_L =$ lower cutoff frequency and $\omega_H =$ upper cutoff frequency, then from equation (4.108),

$$Q\left(\frac{\omega_L}{\omega_0} - \frac{\omega_0}{\omega_L}\right) = -1 \quad \text{and} \quad Q\left(\frac{\omega_H}{\omega_0} - \frac{\omega_0}{\omega_H}\right) = +1$$

Rearrange terms as follows:

$$Q\left(\frac{\omega_L^2 - \omega_0^2}{\omega_0 \omega_L}\right) = -1 \quad \text{and} \quad Q\left(\frac{\omega_H^2 - \omega_0^2}{\omega_0 \omega_H}\right) = +1$$

Subtract the first equation from the second, so that

$$\frac{Q}{\omega_0}\left(\frac{\omega_H^2 - \omega_0^2}{\omega_H} - \frac{\omega_L^2 - \omega_0^2}{\omega_L}\right) = 2$$

or, in an alternative form,

$$\omega_H - \frac{\omega_0^2}{\omega_H} - \omega_L + \frac{\omega_0^2}{\omega_L} = \frac{2\omega_0}{Q}$$

The impedance characteristics of a parallel resonant circuit exhibit geometric symmetry about the resonant frequency, so that $\omega_0^2 = \omega_L\omega_H$. Consequently, the preceding equation reduces to

$$(\omega_H - \omega_L) - (\omega_L - \omega_H) = \frac{2\omega_0}{Q}$$

Combine terms and cancel the common factor of 2. Thus

$$\omega_H - \omega_L = \frac{\omega_0}{Q} = \text{amplifier bandwidth}$$

The difference between the two cutoff frequencies was defined at the beginning of the chapter as the amplifier bandwidth B. Hence

$$B = \omega_H - \omega_L = \text{amplifier bandwidth} = \frac{\omega_0}{Q} \qquad (4.109)$$

Substitute equation (4.107) for Q into equation (4.109) to obtain

$$B = \frac{1}{RC_T} = \omega_2 \qquad (4.110)$$

Note that the bandwidth of the single tuned amplifier is numerically equal to the upper cutoff frequency ω_2 of a resistance coupled amplifier. Therefore both amplifiers have the same figure of merit, namely,

$$F_a = A_rB = A_r\omega_2 = \frac{g_m}{C_T} \qquad (4.111)$$

Hence the factors that make a good resistance coupled video amplifier will also make a good single tuned amplifier.

The equations derived in the preceding paragraphs for the reference gain, bandwidth, and figure of merit are general for all single tuned amplifiers using *high Q coils*. The equations are valid for all values of the *circuit Q*.

The case of the high Q circuit is of interest. Because of the assumed high Q, the bandwidth of the amplifier is narrow, and interest centers upon the frequencies that are very close to the resonant frequency.

This allows the gain function to be expressed in an alternative form useful in chapter 5.

Refer to equation (4.108); this equation for the voltage amplification may be rewritten if an ω_0 is factored out of the imaginary term as follows:

$$A(j\omega) = \frac{-A_r}{1 + jQ(\omega^2 - \omega_0^2)/\omega_0\omega}$$

However, it has been shown that

$$A_r = g_m R; \qquad B = \frac{\omega_0}{Q}; \qquad A_r B = \frac{g_m}{C_T}$$

Hence the preceding equation for the amplification can be written

$$A(j\omega) = -\frac{g_m}{C_T} \cdot \frac{1}{B + j(\omega^2 - \omega_0^2)/\omega}$$

Rearrange the imaginary term into the more convenient form shown in the next equation.

$$A(j\omega) = -\frac{g_m}{C_T} \cdot \frac{1}{B + (j/\omega)(\omega - \omega_0)(\omega + \omega_0)}$$

If the high Q case is assumed, only those frequencies very close to resonance are of interest. Consequently, $\omega \doteq \omega_0$ and

$$\omega + \omega_0 \doteq 2\omega \tag{4.112}$$

$$\omega - \omega_0 = \pm\Delta\omega = \text{frequency deviation off resonance} \tag{4.113}$$

Therefore the approximate gain function for the case of the high Q circuit is

$$A(j\omega) \doteq -\frac{g_m}{C_T} \cdot \frac{1}{B \pm j2\Delta\omega} \tag{4.114}$$

This may be expressed in an alternative form as

$$A(j\omega) \doteq -A_r \frac{1}{1 \pm j2\Delta\omega/B} \tag{4.115}$$

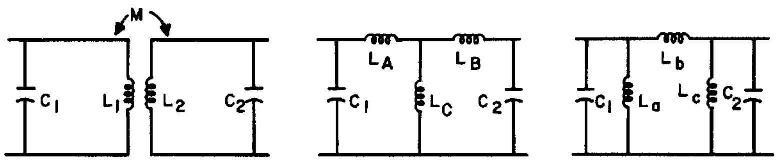
and in an even more compact form

$$A(j\omega) \doteq -A_r \frac{1}{1 \pm jX} \quad \text{where} \quad X = \frac{2\Delta\omega}{B} \tag{4.116}$$

The last two equations are particularly useful in the discussion of stagger tuned amplifiers in the next chapter.

4.15. Double Tuned Amplifiers

In discussing the series peaked amplifier we observed that as long as consideration was restricted to two-terminal networks, the performance of a shunt peaked amplifier was the best that could be expected. We then found that improved operating characteristics resulted from the use of *three*-terminal networks. The same argument holds true for bandpass amplifiers. As long as two-terminal networks are used, the



$$L_a = \frac{L_1 L_2 - M^2}{L_2 - M} \qquad L_A = L_1 - M$$

$$L_b = \frac{L_1 L_2 - M^2}{M} \qquad L_B = L_2 - M$$

$$L_c = \frac{L_1 L_2 - M^2}{L_1 - M} \qquad L_C = M$$

$$M = k\sqrt{L_1 L_2}$$

Fig. 4.36. Equivalent double tuned circuits.

single tuned amplifier provides reasonable operating characteristics. However, if we can forsake the simplicity of the circuit, improved amplifier performance can be obtained through the use of three- or four-terminal double tuned plate circuits.

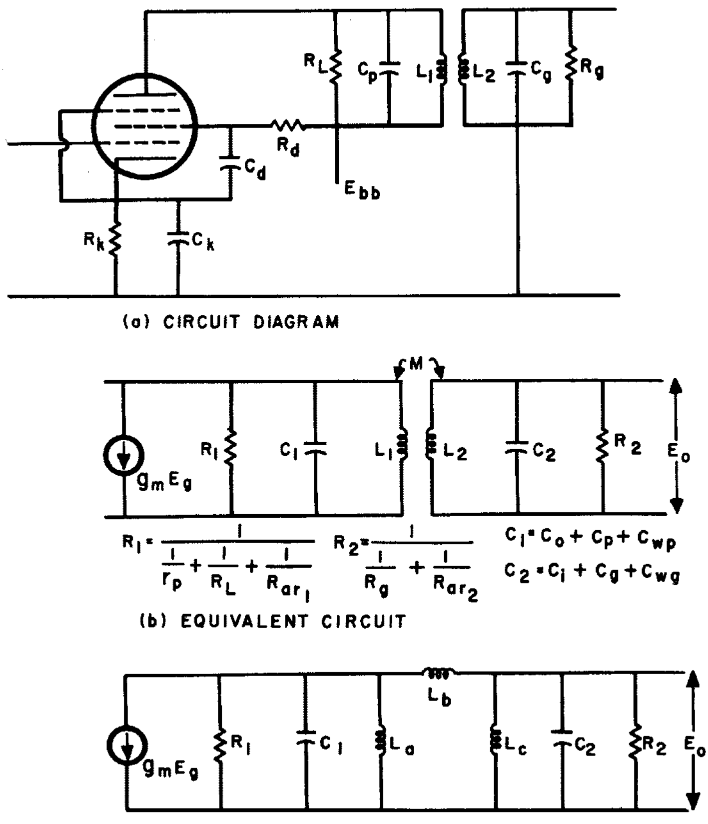
There are several different types of double tuned circuits, and the form used depends upon many factors such as ease of adjustment of the coupling coefficient, economy, operating frequency, and so on. Three possible forms of double tuned circuits are shown in figure (4.36). All three circuits are equivalent to one another according to the formulas given in the figure.

In the broadcast and short wave parts of the frequency spectrum, the double tuned transformer can be used effectively, and the coefficient of coupling can be adjusted to the desired value by varying the degree of insertion of a slug or core in the mutual flux path of the two coils. This adjustment is exceedingly difficult at higher frequencies and it may be more practical to use the tee or pi section.

Because all three of the circuits are equivalent, discussion of any one applies equally well to the other two.

The circuit diagram of a double tuned amplifier is shown in figure

(4.37a); the current source equivalent plate circuit is given in figure (4.37b), and the pi equivalent of the transformer is used in figure (4.37c). In these circuits the following terminology applies: $C_1 =$ total shunt



(a) CIRCUIT DIAGRAM

(b) EQUIVALENT CIRCUIT

(c) PI EQUIVALENT CIRCUIT

Fig. 4.37. Double tuned amplifier circuits.

capacitance on the primary side of the transformer; $C_2 =$ total shunt capacitance on the secondary side of the transformer; $R_1 =$ total effective shunt resistance on the primary side $=$ parallel combination of r_p, R_{ar}, and any connected load resistance R_L; $R_2 =$ total secondary shunt resistance; $L_1 =$ total primary inductance; $L_2 =$ total secondary inductance; $M =$ mutual inductance $= k\sqrt{L_1L_2}$; $k =$ coefficient of coupling.

You can see from figure (4.37c) that the amplifier equivalent plate circuit has the form of the general unloaded pi section discussed in section (4.3). Thus

$$Z_1(s) = \frac{1}{sC_1 + 1/R_1 + 1/sL_a} = \frac{1}{C_1} \cdot \frac{s}{s^2 + s/R_1C_1 + 1/L_aC_1}$$
$$= \frac{1}{C_1} \cdot \frac{s}{s^2 + B_1 s + \omega_a^2} \tag{4.117}$$

Similarly,

$$Z_2(s) = \frac{1}{C_2} \cdot \frac{s}{s^2 + B_2 s + \omega_c^2} \tag{4.118}$$

where
$$B_1 = \frac{1}{R_1C_1} \quad \text{and} \quad B_2 = \frac{1}{R_2C_2} \tag{4.119}$$

$$\omega_a^2 = \frac{1}{L_aC_1} \quad \text{and} \quad \omega_c^2 = \frac{1}{L_cC_2} \tag{4.120}$$

Also
$$Z_3(s) = sL_b \tag{4.121}$$

Therefore the voltage amplification of the double tuned pi section amplifier is

$$A = -g_m Z_m = -g_m \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3}$$

Substitution of the equations for Z_1, Z_2, and Z_3 into this will produce a rather complicated equation. However, if the common denominator is obtained and the ultimate denominator is written as a polynomial, after a great deal of algebraic manipulation the following result is obtained:

$$A(s) = -\frac{g_m k}{C_1C_2(1 - k^2)\sqrt{L_1L_2}} \cdot \frac{s}{s^4 + a_1 s^3 + a_2 s^2 + a_3 s + a_4} \tag{4.122}$$

where
$$a_1 = \omega_r \left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \tag{4.123}$$

$$a_2 = \frac{\omega_r^2}{Q_1Q_2} + \frac{1}{1 - k^2}(\omega_1^2 + \omega_2^2) \tag{4.124}$$

$$a_3 = \frac{\omega_r^2}{1 - k^2}\left(\frac{\omega_2^2}{Q_1} + \frac{\omega_1^2}{Q_2}\right) \tag{4.125}$$

$$a_4 = \frac{\omega_1^2\omega_2^2}{1 - k^2} \tag{4.126}$$

$$\omega_r = \text{resonant frequency of the amplifier} \tag{4.127}$$

$$\omega_1 = \sqrt{\frac{1}{L_1C_1}} \tag{4.128}$$

$$\omega_2 = \sqrt{\frac{1}{L_2C_2}} \tag{4.129}$$

$$Q_1 = (\text{primary } Q \quad \text{at} \quad \omega_r) = \omega_r C_1 R_1 \tag{4.130}$$

$$Q_2 = (\text{secondary } Q \quad \text{at} \quad \omega_r) = \omega_r C_2 R_2 \tag{4.131}$$

In nearly every practical case the primary and secondary are tuned to the same resonant frequency ω_0, so that

$$\omega_1 = \omega_2 = \omega_0 \tag{4.132}$$

and

$$a_1 = \omega_r\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \tag{4.133}$$

$$a_2 = \omega_0^2\left(\frac{\omega_r^2}{\omega_0^2 Q_1 Q_2} + \frac{2}{1-k^2}\right) \tag{4.134}$$

$$a_3 = \frac{\omega_0^2\omega_r}{1-k^2}\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \tag{4.135}$$

$$a_4 = \frac{\omega_0^4}{1-k^2} \tag{4.136}$$

Equation (4.122) is the general equation for the voltage amplification of a double tuned amplifier and it will provide the starting point for all further derivations.

For example, to determine the voltage amplification at the resonant frequency ω_r, let $s = j\omega_r$. This leads to:

$$A(j\omega_r) = -\frac{j\omega_r g_m k}{C_1C_2(1-k^2)\sqrt{L_1L_2}} \cdot \frac{1}{\omega_r^4 - a_2\omega_r^2 + a_4 - j(\omega_r^3 a_1 - \omega_r a_3)} \tag{4.137}$$

Assume that both the primary and secondary are tuned to the same frequency ω_0. At resonance the imaginary term in the denominator of the bracketed factor must be zero, or

$$\omega_r^3 a_1 = \omega_r a_3 \tag{4.138}$$

or

$$\omega_r^2 = \frac{a_3}{a_1} = \frac{\omega_0^2}{1-k^2} \tag{4.139}$$

Consequently, the resonant frequency of the amplifier is

$$\omega_r = \frac{\omega_0}{\sqrt{1 - k^2}} \tag{4.140}$$

Substitute this equation into equation (4.137) and the amplification of the amplifier at resonance becomes

$$A(j\omega_r) = +j\frac{kg_m\sqrt{R_1R_2}}{\sqrt{Q_1Q_2}(k^2 + 1/Q_1Q_2)} \tag{4.141}$$

Considerable algebraic manipulation is required to get the result into this form.

The value of k that will make the reference gain a maximum is easily computed by maximizing $A(j\omega_r)$ by the usual method. This yields

$$k_c = \sqrt{\frac{1}{Q_1Q_2}} \tag{4.142}$$

This is called the coefficient of *critical coupling.*

In practice it is frequently desirable to determine the coupling coefficient that produces the *flattest* response characteristic for the amplifier. This value of k is called the coupling coefficient for *transitional coupling*, and it is denoted by k_t. If k is increased beyond k_t, the response changes from a single peak to a double peaked form. The value for k_t was shown by Aiken[6] to be

$$k_t = \sqrt{\frac{1}{2}\left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2}\right)} \tag{4.143}$$

This is equal to k_c when the circuit Q's are equal.

For a transitionally coupled amplifier, the following formulas[7] can be used to compute the gain and bandwidth.

$$A_r = g_m\sqrt{R_1R_2}\frac{\sqrt{2p(1 + p^2)}}{(1 + p)^2} \tag{4.144}$$

$$B = \frac{1 + p}{\sqrt{2p}} \cdot \frac{\omega_r}{Q_2} \tag{4.145}$$

$$p = \frac{Q_1}{Q_2}$$

[6] C. B. Aiken, "Two Mesh Tuned Coupled Circuit Filters," *Proc. IRE.*, vol. 25, February, 1937.

[7] Formulas reprinted with permission from Valley and Wallman, *Vacuum Tube Amplifiers*, vol. 18, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, p. 218.

The gain-bandwidth product of a transitionally coupled double tuned amplifier is considerably larger than that of a single tuned amplifier.

The response characteristics of the amplifier are determined, as in any other circuit, by the poles of the gain function. Although this is a single stage amplifier, its response characteristics are most conveniently treated with the responses of multistage amplifiers. Further discussion of this circuit will be found in the next chapter.

PROBLEMS

4.1. A single stage resistance coupled amplifier is to be designed, using a type 6AC7 tube. The upper and lower cutoff frequencies are to be 5 mc and 20 c. The following data apply: $C_o = 5.0\ \mu\mu\text{f}$; $g_m = 9000\ \mu\text{mhos}$; $r_p = 1$ megohm; $C_w = 8.0\ \mu\mu\text{f}$; $R_g = 500{,}000$ ohms. Assume that the amplifier is unloaded and nondegenerative. Find the values of R_L, C_c, A_r, F_a, and f_a.

4.2. If a 6AK5 tube had been used in the preceding amplifier, would the reference gain be greater or less than that obtained with the 6AC7 if the upper cutoff frequency remains unchanged? Which *tube* has the larger figure of merit? For the 6AK5, $C_o = 2.8\ \mu\mu\text{f}$ and $g_m = 4300\ \mu\text{mhos}$.

4.3. A single stage shunt peaked amplifier is connected between the second detector and the intensity grid of a cathode ray tube in the receiver of a radar set. The input capacitance of the intensity grid circuit is $10\ \mu\mu\text{f}$. A 6AG7 amplifier tube is used, for which $C_{gp} = 0.06\ \mu\mu\text{f}$; $E_d = 300$ v; $R_g = 200{,}000$ ohms; $C_o = 7.5\ \mu\mu\text{f}$; $E_c = -2.0$ v; $I_b = 28$ ma; $C_i = 12.5\ \mu\mu\text{f}$; $r_p = 100{,}000$ ohms; $I_d = 7$ ma; $E_b = 300$ v; $g_m = 7700\ \mu\text{mhos}$. As a preliminary step in the design it is necessary to determine the distributed wiring capacitance. Thus the amplifier is constructed as a resistance coupled amplifier with an arbitrarily chosen plate load of 6800 ohms. The upper cut-off frequency f_2, with this load resistance is measured experimentally and found to be 1 megacycle. Calculate the wiring capacitance of the amplifier.

4.4. After the wiring capacitance has been determined for the preceding amplifier, the test value of the load resistance is removed. The design requirements of the amplifier are then specified as follows:

(a) Reference gain, not less than 10.
(b) 10–90% rise time, not more than 0.04 μsec.
(c) Overshoot not to exceed 8%.

Calculate the necessary values for R_L, m, L_b; also T_R, γ, and A_r. If the design requirements cannot be satisfied, discuss the problem and offer several alternative designs or possibilities.

4.5. For the same amplifier, calculate the values of R_k, R_d, and E_{bb}.

4.6. Assume that the coupling capacitor has a value of 0.1 μf and the cathode bypass capacitor is 25 μf. Calculate the total sag caused by these two circuits if the duration of the received radar pulse is 1 μsec. If the sag is not to exceed 0.1% over the 1 μsec interval, will these circuit elements yield satisfactory operation?

4.7. If the screen bypass capacitor is 8 μf, will the sag in the amplifier exceed 0.1%? (Use $r_s = 1800$ ohms.)

4.8. An amplifier to be used in a specified industrial application is to have a lower cutoff frequency of 8.7 c. A tube is to be used for which it has already been determined that $R_L = 2000$ ohms; $R_g = 500{,}000$ ohms; $I_b = 10$ ma; $C_c = 0.015$ μf; $g_m = 7700$ μmhos; $r_p = 10{,}000$ ohms; $E_c = -2.0$ v; $E_b = 250$ v. Here C_c and R_g are in the plate circuit of the tube. Design a low frequency compensating circuit, assuming the tube to be a triode, to meet the stipulated design requirements. Check to make sure that the power supply voltage is within reason.

4.9. A single tuned amplifier is needed for the intermediate frequency amplifier in a radio receiver. The amplifier is to have a gain of 100 and a bandwidth of 10 kc. The total interstage shunt capacitance, including the tuning capacitance is 318 $\mu\mu$f. What must be the transconductance of the tube to meet these requirements? Is this within practical limits? Compute the value of R required. If $r_p = 1$ megohm and $R_g = 500{,}000$ ohms, calculate the Q of the coil at reasonance assuming that there is no R_L in the circuit. Assume that the frequency is 455 kc. What is the permissible resistance of the coil?

4.10. A high Q single tuned amplifier using a 6SK7 tube is the last stage in an amplifier chain, and it feeds a load having an input impedance of 250,000 ohms, pure resistance. The interstage wiring capacitance is 10 $\mu\mu$f. The circuit is to be reasonant at 600 kc and it must have a bandwidth of 10 kc. The desired reference gain is 65. Here $g_m = 2000$ μmhos; $C_o = 7$ $\mu\mu$f; $r_p = 1$ megohm; $R_g = 100{,}000$ ohms (in the plate circuit). Find the required coil Q, value for R, the coil resistance and inductance, and the total shunt capacitance required to resonate the circuit.

4.11. Redesign the amplifier of problem (4.10) with a double tuned interstage network with transitional coupling. Assume that the wiring capacitance is equally split between the primary and secondary, and that both primary and secondary are tuned to the same frequency. Calculate the values for the various inductances and the coefficient of coupling. How does the performance of this amplifier compare with that of the single tuned amplifier?

4.12. Work out a design procedure for a triode, low frequency compensated amplifier, having cathode degeneration. Work out both the conventional and optimum cases.

4.13. For the low frequency pentode amplifier with optimum compensation, no degeneration, derive an equation for the lower cutoff frequency.

4.14. Repeat (4.13) for a triode amplifier and compare the results.

Chapter 5

MULTISTAGE AMPLIFIERS IN THE STEADY STATE

In most practical electronic systems, more than one stage of amplification is required to obtain the necessary over-all gain. For example, in the field of long distance telephony, the signal attenuation introduced by the transmission lines is such that large numbers of amplifiers are

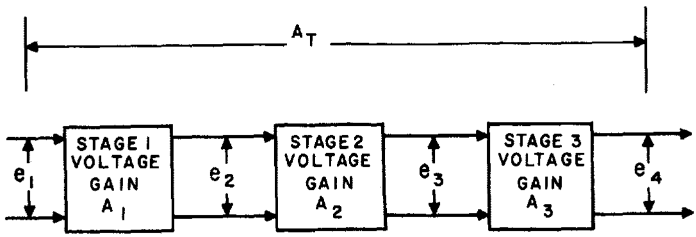


Fig. 5.1. Amplifiers in cascade.

required to maintain adequate signal strength. Radar receivers may use 6 or more stages of bandpass amplifiers to amplify the intermediate frequency signal. Twenty or more stages of video amplification may be required in television studio installations. Even the ordinary broadcast frequency superheterodyne receiver uses several stages of amplification. Consequently, it should be clear that the use of multiple amplifier stages is common.

If the output of one amplifier is connected to the input of another, the two amplifiers are said to be connected in *cascade*. A representation of the cascade connection is shown in figure (5.1). This is the most common method of using multiple stages of voltage amplification, so that most multistage amplifiers are of this type. A significant exception is the *distributed amplifier*, which is discussed toward the end of the chapter.

This chapter is concerned with the steady state response of multistage amplifiers.[1] The response to pulse inputs is discussed in the next chapter.

5.1. The Cascade Connection

One reason for using the cascade connection is to increase the over-all voltage gain. That this does result is clear from figure (5.1) because if A_1 = amplification of stage $1 = e_2/e_1$; A_2 = amplification of stage $2 = e_3/e_2$; A_3 = amplification of stage $3 = e_4/e_3$; then the over-all amplification of the cascade is $A_T = e_4/e_1$. It is apparent that A_T is the product of the individual stage gains; that is,

$$A_T = A_1A_2A_3 = (e_2/e_1)(e_3/e_2)(e_4/e_3) = e_4/e_1$$

The form of this equation is unchanged regardless of the number of stages in the cascade. Hence for a cascade of n stages,

$$A_T = A_1A_2A_3 \dots A_n \tag{5.1}$$

You can see from equation (5.1) that the over-all gain is increased over that of any one stage as long as the individual stage gains exceed unity. Hence the suggested primary purpose of cascading has been successfully accomplished.

There is another possible application of the cascade connection implicit in the form of equation (5.1). Several different kinds of amplifiers were analyzed in the preceding chapter and it was shown that their gain functions could be written in the form of a rational fraction of polynomials, with both the polynomials in the numerator and denominator being factorable. Therefore if a given function of the variable s or $j\omega$ can be factored into a series of functions containing terms of the forms obtained for the gain functions of these amplifiers, equation (5.1) states that the given function can be synthesized by a cascade connection of amplifiers. This is a statement of tremendous practical importance and the design process resulting from this use of equation (5.1) is called *synthesis by factoring.*

The general idea involved can be illustrated by a simple example. Suppose that a given electronic system requires the use of an amplifier having an amplitude response of the form

$$|A(X)| = \frac{K}{\sqrt{1 + X_1^2}\sqrt{1 + X_2^2}}$$

[1] A large fraction of this chapter is based, with permission, upon a course given by Professor J. M. Pettit of the Electrical Engineering Department of Stanford University.

where X_1 and X_2 are governed by a frequency deviation from some reference frequency. According to equation (4.116) in the preceding chapter, the gain function of a high Q, single tuned amplifier was shown to be

$$A(j\omega) = -A_r \frac{1}{1 + jX} = \frac{A_r \underline{/\theta}}{\sqrt{1 + X^2}}$$

where the X represents the frequency deviation off resonance. It is clear that the desired gain function can be constructed by connecting two single tuned amplifiers in cascade with the bandwidths, Q's, and resonant frequencies adjusted to give the desired values for X_1 and X_2. This example is indicative in a rather oversimplified way of one of the most practical applications of the cascade connection.

There are limitations on the extent to which the cascading process can be usefully employed. It turns out that a simple cascade of identical amplifiers has an over-all bandwidth that is less than the bandwidth of any one stage in the cascade. Hence, even though the increase in over-all gain is accomplished, the reduction in bandwidth also reduces some of the advantage derived. As a matter of fact, this bandwidth reduction eventually imposes an upper limit on the number of stages that can be cascaded. This problem is discussed at some length in subsequent sections.

5.2. Identical Resistance Coupled Amplifiers in Cascade

Suppose that a given application requires the use of resistance coupled amplifiers and that the gain and upper cutoff frequency requirements are such that one stage is insufficient. It is decided to try a cascade connection of identical resistance coupled amplifiers. The immediate point of interest is the determination of the effects of this multistaging on the over-all reference gain and upper cutoff frequency.

For the purposes of the following derivation, define terms as follows:

$\omega_2 =$ upper cutoff frequency of each individual stage.
$\omega_1 =$ lower cutoff frequency of each individual stage.
$\omega_H =$ upper cutoff frequency of the cascade.
$\omega_L =$ lower cutoff frequency of the cascade.
$A(j\omega) =$ gain function of a single amplifier stage.
$A_T(j\omega) =$ gain function for the cascade.

It was shown in section (4.5) that the high frequency gain function for a resistance coupled amplifier was

$$A(j\omega) = -A_r \frac{1}{1 + j\omega/\omega_2} \tag{5.2}$$

Hence the magnitude of this complex number is

$$|A(j\omega)| = A_r \frac{1}{\sqrt{1 + (\omega/\omega_2)^2}} \tag{5.3}$$

If n identical stages are connected in cascade, the over-all gain of the system is given by equation (5.1). Because all amplifier stages are assumed identical in this case, the over-all gain of the cascade is equal to the gain of a single stage raised to the nth power. That is

$$|A_T(j\omega)| = |A(j\omega)|^n = A_r^n \left(\frac{1}{\sqrt{1 + (\omega/\omega_2)^2}}\right)^n \tag{5.4}$$

or

$$|A_T(j\omega)| = A_r^n \frac{1}{[1 + (\omega/\omega_2)^2]^{n/2}} \tag{5.5}$$

It is clear from equation (5.5) that the reference gain of the amplifier cascade is A_r^n. Therefore, when the signal frequency ω is equal to the upper cutoff frequency ω_H of the cascade, the radical in the denominator of equation (5.5) must be equal to $\sqrt{2}$. Therefore

$$[1 + (\omega_H/\omega_2)^2]^{n/2} = \sqrt{2} \tag{5.6}$$

Square both sides of this equation and then take the nth root so that

$$1 + (\omega_H/\omega_2)^2 = 2^{1/n} \tag{5.7}$$

Then solve for the upper cutoff frequency of the cascade.

$$\omega_H = \omega_2\sqrt{2^{1/n} - 1} \tag{5.8}$$

If this same procedure is followed for the low frequency response it will be found that the over-all lower cutoff frequency of the cascade is

$$\omega_L = \frac{\omega_1}{\sqrt{2^{1/n} - 1}} \tag{5.9}$$

By evaluating the radical in each of the foregoing equations for the over-all cutoff frequencies of the cascade, the curves of figure (5.2) are obtained. It should be clear from this figure that the over-all cutoff frequencies of the cascade are much closer together than are the cutoff frequencies of the individual stages. Cascading causes a reduction in the upper cutoff frequency and an increase in the lower cutoff frequency.

The effects of cascading may be illustrated in another way as shown for the high frequency case in figure (5.3). This curve shows how the upper cutoff frequency of the individual stages must be increased to keep the over-all cutoff frequency constant as more stages are added to the cascade. The gain-bandwidth product of the amplifier stages is a constant. Therefore as the stage bandwidth is increased to keep the over-all bandwidth constant, this causes a corresponding reduction in the stage gain. From this it is reasonable to conclude that a point is eventually reached where the decrease in stage gain required to keep the

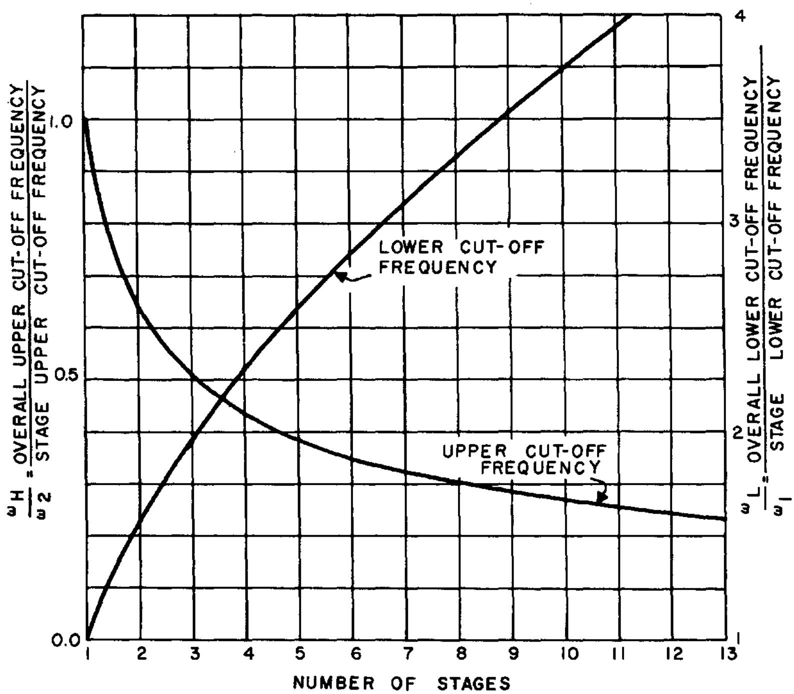


Fig. 5.2. Effect of cascading on the over-all cutoff frequencies.

over-all upper cutoff frequency constant causes the over-all gain to decrease. At this point, further cascading is useless.

The effect just described can be verified mathematically by deriving an equation for the reference gain of the cascade as a function of the over-all upper cutoff frequency ω_H and the number of stages n. The reference gain for a single stage of the cascade is

$$A_r = g_m R = \frac{g_m}{C_T} \cdot \frac{1}{\omega_2} = \frac{F_a}{\omega_2} \tag{5.10}$$

Hence the reference gain of the n stage cascade will be

$$A_r^n = F_a^n \left(\frac{1}{\omega_2}\right)^n \tag{5.11}$$

However, from equation (5.8), the upper cutoff frequency of each stage can be expressed as

$$\omega_2 = \frac{\omega_H}{\sqrt{2^{1/n} - 1}} \tag{5.12}$$

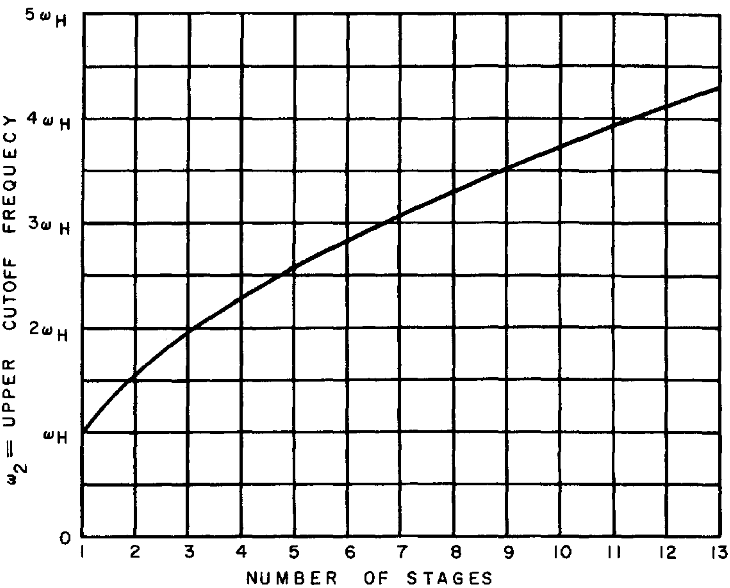


Fig. 5.3. Effect of cascading on the upper cutoff frequency required for each stage to keep the over-all upper cutoff frequency, ω_H, constant.

Therefore the reference gain of the n stage cascade becomes

$$A_r^n = F_a^n \left(\frac{\sqrt{2^{1/n} - 1}}{\omega_H}\right)^n \tag{5.13}$$

Rewrite this as

$$A_r^n = \left(\frac{F_a}{\omega_H}\right)^n (2^{1/n} - 1)^{n/2} = A_T \tag{5.14}$$

In terms of volt logits, the over-all gain is

$$A_T(\text{volt logits}) = 10n \log_{10} \frac{F_a}{\omega_H} + 5n \log_{10} (2^{1/n} - 1) \tag{5.15}$$

If a given amplifier is considered, the amplifier figure of merit F_a

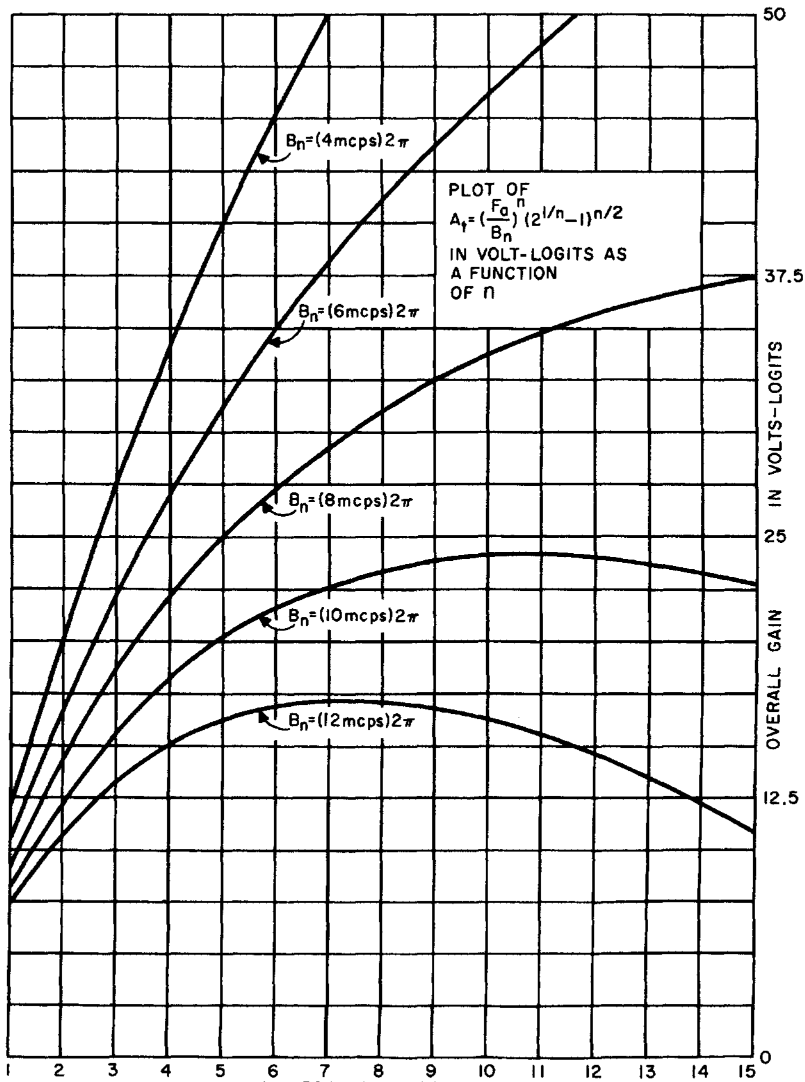


Fig. 5.4. Effect of cascading on the over-all gain of synchronously connected stages. Over-all bandwidth or upper cutoff frequency held constant. Assumed that $F_a = 409 \times 10^6$ rps.

may be presumed known. Therefore, by treating the over-all cutoff frequency ω_H of the cascade as a parameter, the gain of the cascade in volt logits can be computed and plotted as a function of the number of stages n, in the cascade. The results of such a computation are shown in figure (5.4).

Figure (5.4) illustrates two points of considerable practical importance, as follows:

(1) For a given over-all upper cutoff frequency ω_H, the process of continuously adding stages to the cascade causes the over-all gain to increase at a continually diminishing rate up to a certain point. Thereafter, further increases in the number of stages cause a reduction in the over-all reference gain.

(2) For a given over-all upper cutoff frequency, there is a particular value of n that will produce the maximum gain.

It should be clear that it may or may not be possible to obtain specified over-all gain and upper cutoff frequency requirements by the simple process of cascading.

5.3. Cascading Compensated Amplifiers

The gain function of a single stage shunt peaked amplifier was derived in chapter 4 and shown to be

$$A(s) = -A_r \frac{1 + ms/\omega_2}{1 + s/\omega_2 + ms^2/\omega_2^2}$$

If n identical stages are cascaded, the over-all gain of the cascade is

$$A(s) = (-A_r)^n \left(\frac{1 + ms/\omega_2}{1 + s/\omega_2 + ms^2/\omega_2^2}\right)^n$$

From this equation you can compute the over-all upper cutoff frequency of the cascade, using the usual method, to obtain

$$\omega_H = 0.707 \frac{\omega_2}{m} \sqrt{(m^2 2^{1/n} + 2m - 1) + \sqrt{(m^2 2^{1/n} + 2m - 1)^2 + 4m^2(2^{1/n} - 1)}} \tag{5.16}$$

This looks pretty terrifying, but it is fairly easy to compute ω_H as a function of the number of stages in the cascade for various values of the peaking parameter. Results are shown in figure (5.5). These curves show that cascading reduces the over-all upper cutoff frequency regardless of the value of the peaking parameter m. However, the rate of decrease is less for the larger values of m. This is more apparent from figure (5.6). Here the same data has been replotted to show the decrease from the single stage cutoff frequency, which is assumed to be the same in all cases.

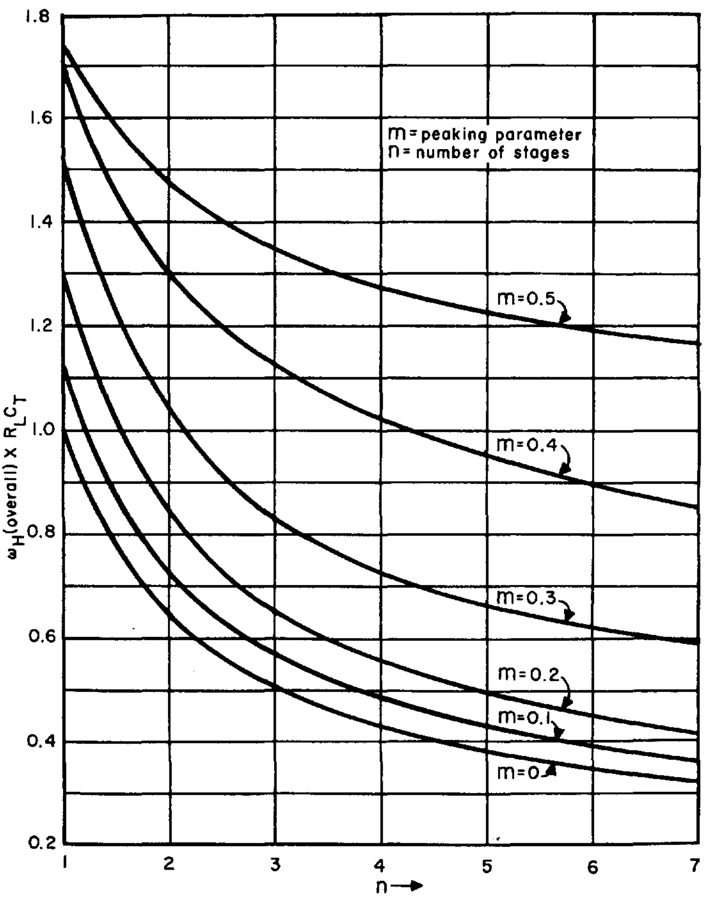


Fig. 5.5. Identical shunt peaked amplifiers in cascade. Effect on upper cutoff frequency.

Corresponding results should be expected for other compensated amplifiers of any type. The elementary low frequency compensated amplifier is an interesting exercise.

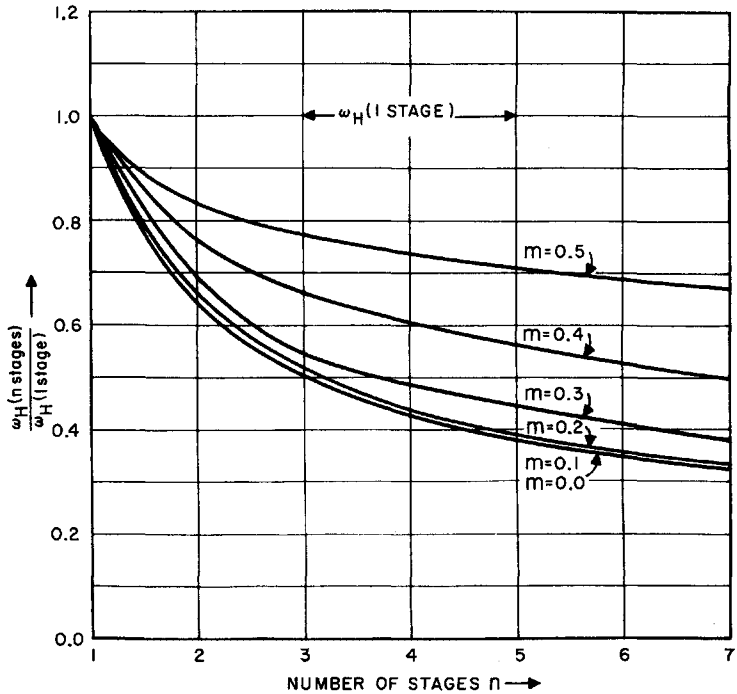


Fig. 5.6. Identical shunt peaked amplifiers in cascade. Effect on upper cutoff frequency.

5.4. Synchronous Single Tuning

Bandpass amplifiers are cascaded in the same way as low pass resistance coupled amplifiers. There is one important change. For the tuned bandpass amplifier, both the gain and bandwidth can be adjusted over a wide range just as the gain and upper cutoff frequency of the resistance coupled amplifier were under control. However, in the bandpass case, the center frequency can also be varied. Because of this possibility of center frequency adjustment, there are two ways of cascading bandpass amplifiers: (1) all center frequencies can be the same, or (2) the center frequencies can be different.

The first kind of cascading is called the *synchronous* connection because all stages of the cascade are synchronously tuned to the same band center. The second possibility is called *stagger tuning* when the center frequencies of the successive stages in the cascade bear particular relationships with respect to one another. The synchronous connection is discussed in this section.

It was shown in the preceding chapter that the reference gain and the bandwidth of a single tuned amplifier are

$$A_r = g_m R; \qquad B = \frac{1}{RC_T} = \frac{\omega_0}{Q}$$

so that the gain-bandwidth product is

$$F_a = A_r B = \frac{g_m}{C_T}$$

The equation for the voltage amplification in the steady state is

$$A(j\omega) = -A_r \frac{1}{1 + jQ(\omega/\omega_0 - \omega_0/\omega)} \tag{5.17}$$

or, alternatively,

$$A(j\omega) = -A_r \frac{1}{1 + jQ(\omega^2 - \omega_0^2)/\omega_0\omega} \tag{5.18}$$

Now suppose that we let $B = \omega_0/Q$ and then evaluate the *magnitude* of the gain function. This gives us

$$|A(j\omega)| = A_r \left[1 + \left(\frac{\omega^2 - \omega_0^2}{B\omega}\right)^2\right]^{-1/2} \tag{5.19}$$

If n identical stages are synchronously cascaded, the amplitude of the over-all gain function is

$$|A_T(j\omega)| = |A(j\omega)|^n = A_r^n \left[1 + \left(\frac{\omega^2 - \omega_0^2}{B\omega}\right)^2\right]^{-n/2} \tag{5.20}$$

Solve this equation for $(\omega^2 - \omega_0^2)$.

$$(\omega^2 - \omega_0^2) = \pm \omega B \sqrt{\left(\frac{A_r}{A_T}\right)^{2/n} - 1} \tag{5.21}$$

At the upper and lower cutoff frequencies of the cascade, we know that

$$\frac{A_r^n}{A_T} = \sqrt{2} = 2^{1/2} \tag{5.22}$$

Hence at the cutoff frequencies, equation (5.21) reduces to

$$\omega_H^2 - \omega_0^2 = \omega_H B\sqrt{2^{1/n} - 1} \tag{5.23}$$

$$\omega_L^2 - \omega_0^2 = -\omega_L B\sqrt{2^{1/n} - 1} \tag{5.24}$$

Subtract equation (5.24) from equation (5.23), so that

$$\begin{aligned} \omega_H^2 - \omega_L^2 &= (\omega_H + \omega_L)(\omega_H - \omega_L) \\ &= (\omega_H + \omega_L)B\sqrt{2^{1/n} - 1} \end{aligned} \tag{5.25}$$

Cancel the common terms on each side of this equation and define the over-all bandwidth of the n stage cascade as

$$B_n = \omega_H - \omega_L = \text{over-all bandwidth}$$

This reduces equation (5.25) to

$$B_n = B\sqrt{2^{1/n} - 1} \tag{5.26}$$

The over-all reference gain of the cascade is clearly

$$A_T = A_r^n = (g_m R)^n \tag{5.27}$$

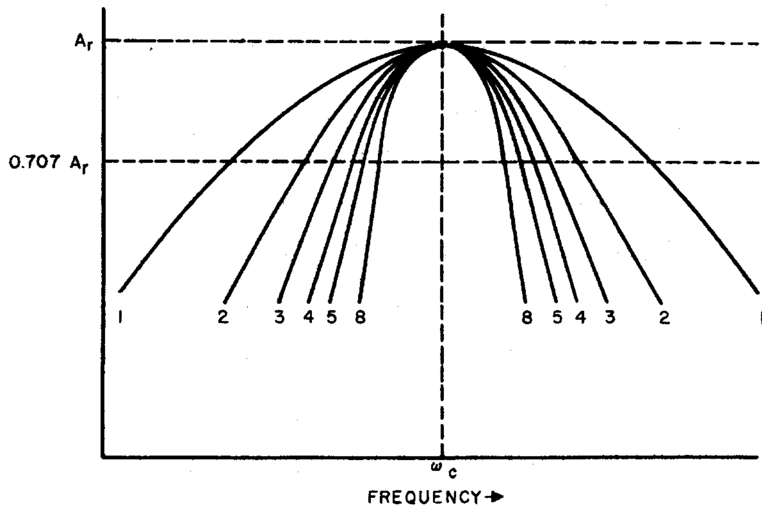


Fig. 5.7. Effect of cascading on the rate at which the voltage gain drops off near the cutoff frequencies.

We have shown that the equations for the over-all reference gain and bandwidth of the synchronously cascaded single tuned amplifier are precisely the same as those obtained for the over-all gain and upper cutoff frequency of the cascaded RC coupled amplifier. Consequently, all the derivations carried through in section (5.2) for the low pass characteristics of the cascaded resistance coupled amplifier apply directly to the synchronous single tuned amplifier. It is necessary only to change all upper cutoff frequency terms, ω_2 and ω_H, to the corresponding bandwidth terms, B and B_n. Therefore in figure (5.2), the over-all bandwidth is shown as a function of the number of stages n when the bandwidth of the individual stages is held constant. Similarly, figure (5.3) shows how the stage bandwidth must be increased, as n is increased, to keep the over-all bandwidth B_n constant. Figure (5.4) also shows how the over-all gain varies as the over-all bandwidth is held constant and the cascading process is continued.

Finally, figure (5.7), which applies to cascaded resistance coupled or single tuned amplifiers, shows that a sharper cutoff characteristic results from multistaging identical stages.

Although the double tuned amplifier has not been discussed at length

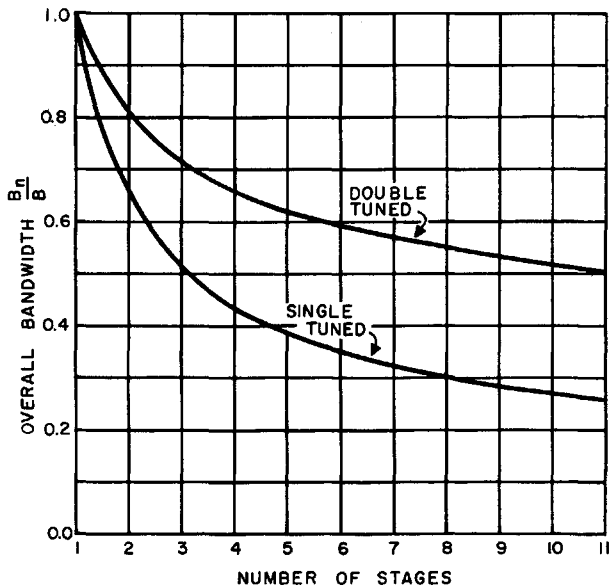


Fig. 5.8. Comparison of synchronously connected single and double tuned amplifiers; double tuned stages are assumed to be transitionally coupled.

yet, the methods just outlined for the single tuned synchronous cascade can be used to show that the over-all bandwidth of synchronous connected, double tuned, transitionally coupled amplifiers is

$$B_n = B(2^{1/n} - 1)^{1/4} \tag{5.28}$$

This equation is derived in section (5.16).

If you compare equations (5.26) and (5.28) you can see that the reduction in over-all bandwidth caused by cascading proceeds at a much slower rate for the double tuned amplifier. In other words, for any given number of stages, the over-all gain, bandwidth, and figure of merit of the double tuned cascade will be larger than that of the single tuned cascade. This is illustrated in figure (5.8).

5.5. Number of Stages Required for Maximum Gain

You can see from figure (5.4) that there is a maximum value of gain that cannot be exceeded for a given over-all bandwidth of a cascade of resistance coupled or single tuned amplifiers. That is, if the over-all bandwidth is specified, a particular number of stages are required to produce the maximum possible gain.

The derivation of the value for n that will make the gain a maximum for a specified over-all bandwidth is simplified through the use of an approximation. You will recall that the equations for the over-all bandwidth and upper cutoff frequency of the amplifier cascades contained terms of the form $(2^{1/n} - 1)^{1/2}$. The form of this factor can be altered by expanding the $2^{1/n}$ term in a power series as follows:

$$2^{1/n} = 1 + \frac{1}{n} \ln 2 + \frac{1}{2!}\left(\frac{1}{n} \ln 2\right)^2 + \frac{1}{3!}\left(\frac{1}{n} \ln 2\right)^3 + \ldots \quad (5.29)$$

Although there is no basis for the action at this time, it is assumed that all but the first two terms in the series can be neglected as long as n is greater than 1. The accuracy of this assumption will be checked later. Therefore

$$2^{1/n} \doteq 1 + \frac{1}{n} \ln 2 \quad (5.30)$$

Therefore the original factor can be approximated by

$$(2^{1/n} - 1)^{1/2} \doteq \left(1 + \frac{1}{n} \ln 2 - 1\right)^{1/2}$$

$$\doteq \left(\frac{1}{n} \ln 2\right)^{1/2}$$

$$\doteq \frac{1}{1.2\sqrt{n}} \quad (5.31)$$

The validity of the approximation previously made is clear from table 7; this shows the exact and approximate values of the factor for various values of n. As long as n exceeds 3, the error in the approximation will be of the order of 5% or less.

As a result of the preceding approximation, the over-all bandwidth or upper cutoff frequency, as the case may be, can be expressed as

$$\omega_H \doteq \frac{\omega_2}{1.2\sqrt{n}} \quad (5.32)$$

$$B_n \doteq \frac{B}{1.2\sqrt{n}} \quad (5.33)$$

TABLE 7

EXACT AND APPROXIMATE VALUES OF THE BANDWIDTH REDUCTION FACTOR, SINGLE TUNED AMPLIFIER

n	$\sqrt{2^{1/n} - 1}$	$\dfrac{1}{1.2\sqrt{n}}$	% error
1	1.000	0.833	16.7
2	0.643	0.589	8.4
3	0.510	0.481	5.7
4	0.435	0.416	4.4
5	0.387	0.372	3.9
6	0.350	0.340	2.9
7	0.323	0.315	2.5
8	0.301	0.294	2.3
9	0.283	0.278	1.8
10	0.268	0.264	1.5

The over-all reference gain of the amplifier cascade is

$$A_T = A_r^n = (g_m R)^n$$

However, it has been shown that

$$B = \text{stage bandwidth} = \frac{1}{RC_T}$$

Consequently, the over-all gain can be written

$$A_T = \left(\frac{g_m}{C_T} \cdot \frac{1}{B}\right)^n = (F_a/B)^n \tag{5.34}$$

Solve equation (5.33) for the stage bandwidth.

$$B = B_n(1.2)\sqrt{n} \tag{5.35}$$

Substitute this result into equation (5.34), so that the over-all gain of the cascade is expressed by

$$A_T = \left(\frac{F_a}{1.2} \cdot \frac{1}{B_n\sqrt{n}}\right)^n \tag{5.36}$$

Equation (5.36) for the over-all gain of the cascade can be used to determine the value of n that will make A_T a maximum when the over-all bandwidth B_n and stage figure of merit are constant. All the terms in equation (5.36), except for the radical, are independent of n. Therefore it is convenient to rewrite the equation as

$$A_T = K^n(n)^{-n/2} \tag{5.37}$$

where

$$K = \frac{F_a}{1.2B_n} \tag{5.38}$$

The existence of a maximum value for the over-all gain was shown in figure (5.4). The gain was expressed in volt logits in that case, so the same evaluation should be used in the present calculation. The same kind of curve is obtained if natural logarithms are used instead of logarithms to the base 10, and this simplifies the calculus involved in the derivation. The natural logarithm of equation (5.37) is

$$\ln A_T = n \ln K - \frac{n}{2} \ln n$$

Maximize this equation; differentiate with respect to n and set the result equal to zero.

$$0 = \ln K - \frac{1}{2} - \frac{1}{2} \ln n$$

Therefore

$$\ln \frac{K^2}{n} = 1$$

and so

$$n = \frac{K^2}{\varepsilon} \tag{5.39}$$

The equation for the value of n required to make the over-all gain a maximum is obtained by inserting the value for K in equation (5.38) into the foregoing equation for n. The result is

$$n = \frac{1}{\varepsilon}\left(\frac{F_a}{1.2B_n}\right)^2 = \frac{1}{\varepsilon}\left(\frac{g_m}{1.2B_nC_T}\right)^2 \tag{5.40}$$

The maximum possible gain corresponding to a specified over-all bandwidth is evaluated by substituting the value for n computed from equation (5.40) into equation (5.36) for the over-all gain.

5.6. Stage Gain for Maximum Bandwidth, Synchronous Connection

It is often important to know the number of stages and the gain per stage necessary to produce the *maximum over-all bandwidth* for a specified over-all gain. As before, the discussion here applies only to the cascading of identical resistance coupled or single tuned amplifiers. However, the general technique is applicable to any cascade.

The derivation of the necessary equations is similar to that followed in the preceding article. The first step is to define the figure of merit of one stage of the amplifier as

$$F_a = A_r B$$

The over-all gain and bandwidth (or upper cutoff frequency) for a cascade of identical stages were shown to be

$$A_T = A_r^n; \qquad B_n = \frac{B}{1.2\sqrt{n}}$$

Therefore the *stage* gain and bandwidth can be expressed in terms of the *over-all* gain and bandwidth as

$$A_r = A_T^{1/n}; \qquad B = 1.2\sqrt{n}\, B_n$$

and the amplifier figure of merit is

$$F_a = A_T^{1/n}(1.2B_n)\sqrt{n} \tag{5.41}$$

Solve this equation for the over-all bandwidth to obtain

$$B_n = \frac{F_a}{A_T^{1/n}(1.2)\sqrt{n}} = \frac{F_a}{1.2}(A_T^{-1/n} n^{-1/2}) \tag{5.42}$$

To determine the number of stages n that will make the over-all bandwidth B_n a maximum when the over-all gain A_T is constant, differentiate B_n with respect to n and set the result equal to zero. After a slight rearrangement of terms,

$$A_T^{1/n} \frac{\partial}{\partial n}(n^{-1/2}) = -n^{-1/2} \frac{\partial}{\partial n}(A_T^{1/n}) \tag{5.43}$$

Carry out the mathematical operations that are indicated and write

$$\frac{1}{2} = \frac{1}{n} \ln A_T \tag{5.44}$$

so that $n = 2 \ln A_T$ = number of stages that will make the over-all bandwidth a maximum when the over-all gain is specified. (5.45)

Take the antilogarithm of both sides of equation (5.45).

$$A_T = \varepsilon^{n/2} \tag{5.46}$$

which is the over-all gain when the circuit is designed for maximum over-all bandwidth. In terms of the stage gain, this is

$$A_r^n = \varepsilon^{n/2} \tag{5.47}$$

or

$$A_r = \sqrt{\varepsilon} = 1.65 = 2.17 \quad \text{volt logits} \tag{5.48}$$

Equation (5.48) is the *stage* gain required to make the over-all bandwidth a maximum. From this result you can see that the maximum over-all bandwidth always results if the gain of each stage in the cascade is set to 1.65. This is true regardless of the tube used or the number of tubes in the cascade.

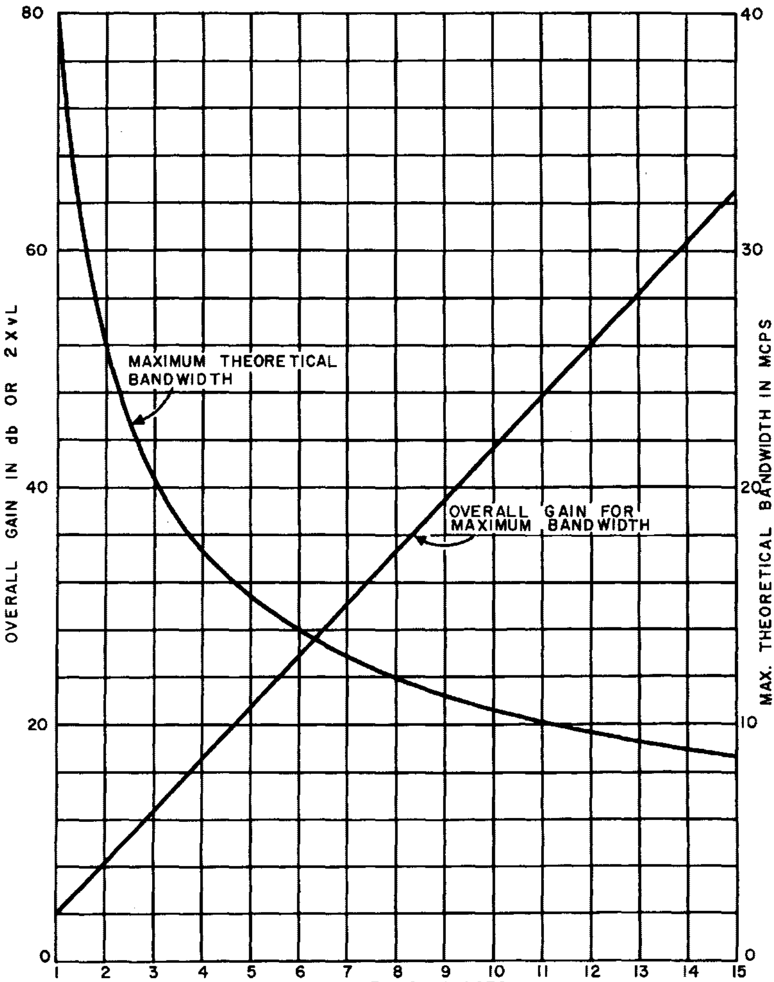


Fig. 5.9. Maximum theoretical bandwidth as a function of the number of stages. Data computed assuming synchronous connected single tuned amplifiers using 6AK5 tubes for which $F_a = 409 \times 10^6$ rps.

The actual maximum value of the over-all bandwidth obtainable under this condition is determined by substituting the value of n given by equation (5.45) into equation (5.42) for the over-all bandwidth. The problem is more apparent when these two equations are plotted together on the same graph. The results are shown in figure (5.9). The straight line represents the over-all gain A_T as a function of n when the stage gain is adjusted to 1.65, thereby producing the maximum over-all bandwidth.

5.7. Symmetry in the Response of Tuned Amplifiers

The equivalent plate circuit of a single tuned amplifier is given in figure (5.10). The voltage amplification of the circuit is

$$A(s) = -g_m Z_m(s)$$

or

$$A(s) = -g_m R \frac{1}{1 + \omega_0 R C_T(s/\omega_0 + \omega_0/s)} \tag{5.49}$$

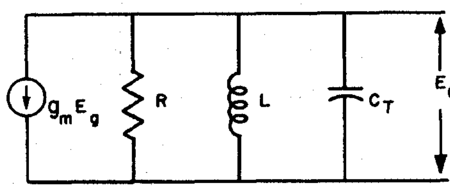


Fig. 5.10. Equivalent plate circuit of a single tuned amplifier.

where

$$\omega_0 = \sqrt{\frac{1}{LC_T}} = \text{bandcenter of the frequency response of the stage} \tag{5.50}$$

$$B = \frac{1}{RC_T} = \text{bandwidth} \tag{5.51}$$

$$A_r = g_m R = \text{reference gain} \tag{5.52}$$

Therefore the voltage amplification function can be written

$$A(s) = -A_r \frac{1}{1 + \dfrac{\omega_0}{B}\left(\dfrac{s}{\omega_0} + \dfrac{\omega_0}{s}\right)} \tag{5.53}$$

The quantity in front of the term in parentheses is Q. That is,

$$Q = \frac{\omega_0}{B} \tag{5.54}$$

so that the preceding gain equation becomes

$$A(s) = -A_r \frac{1}{1 + Q(s/\omega_0 + \omega_0/s)} \tag{5.55}$$

Now write the steady state response of the amplifier by replacing s with $j\omega$ in the preceding equation. Thus

$$A(j\omega) = -A_r \frac{1}{1 + jQ(\omega/\omega_0 - \omega_0/\omega)}$$

$$= \frac{A_r \underline{/\theta(\omega)}}{\sqrt{1 + Q^2(\omega/\omega_0 - \omega_0/\omega)^2}} \tag{5.56}$$

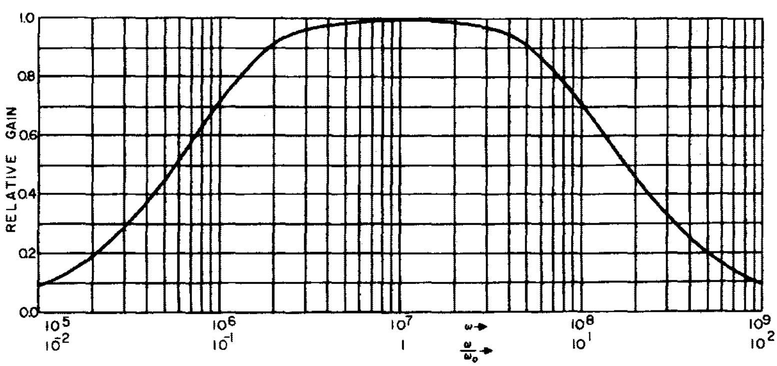


Fig. 5.11. Geometric symmetry in the amplitude response of a single tuned amplifier.

If the magnitude of this function is plotted against ω or ω/ω_0, the results will appear as shown in figure (5.11).

The curve of figure (5.11) is valid for any parallel tuned circuit of any Q. Note that the curve is symmetrical about the center frequency ω_0 when frequency is plotted on a logarithmic scale. This is *geometric symmetry*, and for such a function, bandcenter is related to the two cutoff frequencies by $\omega_0^2 = \omega_H\omega_L$.

This appearance of symmetry disappears when the response is plotted on a *linear* frequency scale as shown in figure (5.12). However, as the Q of the circuit is made progressively larger, the curve becomes more nearly symmetrical about the center frequency. Finally, when

Q is 20 or more, the response is virtually *arithmetically* symmetrical about the center frequency. In this case, the center frequency and cutoff frequencies are related to one another as follows:

$$\omega_0 = \frac{\omega_H + \omega_L}{2}$$

We can now make the following statements:

(1) For any Q the frequency response of a single tuned amplifier is *geometrically* symmetrical about the center frequency ω_0.

(2) When the circuit Q is 20 or more, the response is *arithmetically* symmetrical about the center frequency.

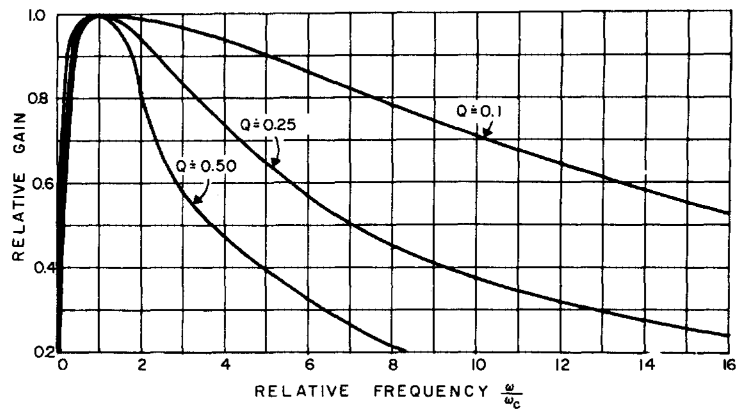


Fig. 5.12. Steady state frequency response characteristics of a single tuned amplifier plotted on a linear frequency scale.

Functions exhibiting arithmetic symmetry are easier to use than those having geometric symmetry. Thus the work that follows will be divided into two categories:

(1) High Q cases where the response is arithmetically symmetrical in its own right.

(2) Low Q cases where the response exhibits geometric symmetry. Here it is desirable to make a transformation of variables in the response function so that the transformed response will exhibit arithmetic symmetry as a function of the new variable.

The change of variables required to transform geometrically symmetrical functions of one variable into arithmetically symmetrical functions of a new variable is not hard to figure out. First, in equation (5.55) define the following term:

$$S = \text{normalized } s = \frac{s}{\omega_0} \tag{5.57}$$

$$= \frac{\text{complex frequency } s}{\text{bandcenter frequency } \omega_0}$$

Thus the new gain function for the single tuned amplifier is

$$A(s) = -A_r \frac{1}{1 + Q(S + 1/S)} \tag{5.58}$$

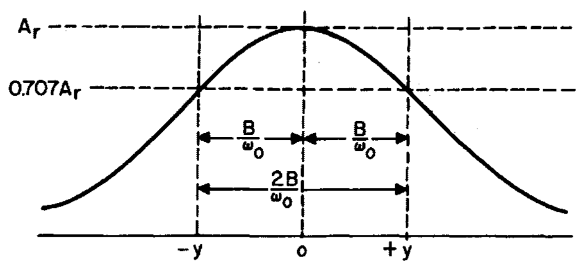


Fig. 5.13. Arithmetic symmetry as a function of y.

Now transform the variable. Let

$$p = S + 1/S \tag{5.59}$$

Because S is a complex number, p will also be complex and of the form $p = x + jy$. Substitute the transformation equation (5.59) into the gain function so that

$$A(p) = -A_r \frac{1}{1 + Qp} \tag{5.60}$$

The *steady state* response of the amplifier as a function of the new variable is obtained if we replace p with jy. Then, because y can be positive or negative, the steady state gain function is

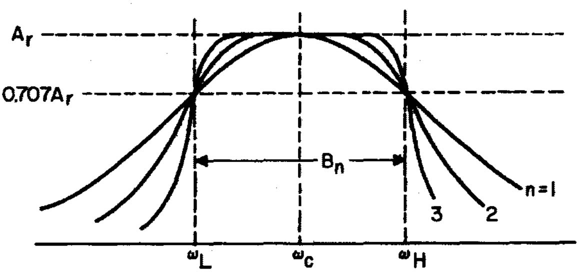
$$A(jy) = -A_r \frac{1}{1 \pm jQy} = \frac{A_r \underline{/\theta}}{\sqrt{1 + Q^2y^2}}$$

This response, as a function of y, is arithmetically symmetrical about $y = 0$, regardless of the value of Q. This is shown in figure (5.13).

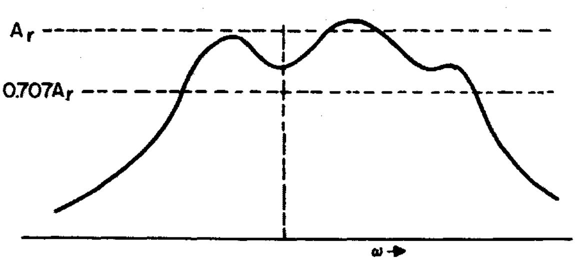
To generalize from this specific case, the following statement is made:

A function that exhibits geometric symmetry as a function of ω or ω/ω_0 can be transformed into a function that is arithmetically symmetrical about $y = 0$ if

$$p = x + jy = S + 1/S$$



(a) MAXIMALLY FLAT FUNCTIONS; n=1,2, AND 3

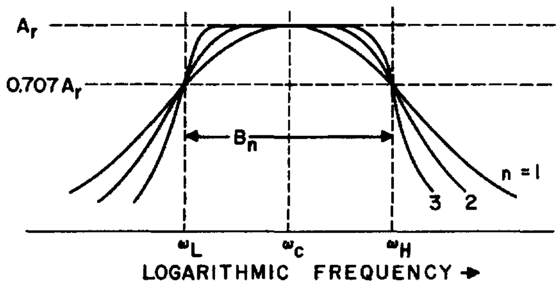


(b) INCORRECT ADJUSTMENT OF STAGES

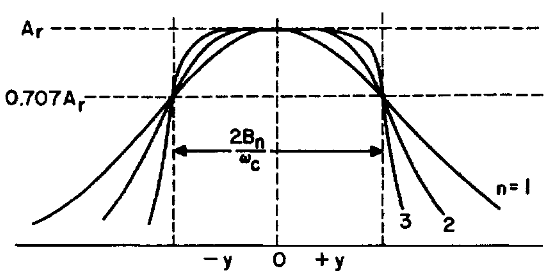
Fig. 5.14. Some responses when ω_c/B_n exceeds 20.

5.8. Flat Functions

The discussion in the preceding section showed that the responses of individual single tuned amplifier stages could exhibit either of two general forms: (1) arithmetic symmetry when ω_0/B is greater than 20; (2) geometric symmetry when ω_0/B is less than 20. The same general remarks apply when n such stages are cascaded though no proof is given for this statement. In the cascaded case, we let ω_c designate the bandcenter frequency and B_n represent the bandwidth of the cascade. Then, in this case when ω_c/B_n is 20 or more, we will find that there is *one* unique system of tuning and adjusting the individual stages so that the response of the cascade appears as shown in figure (5.14a). Note that it is arithmetically symmetrical about the center frequency ω_c. Any alteration in the adjustments of the individual stages will cause the response to become *lumpy* as shown in figure (5.14b). Therefore the



(a) EXACT FLAT FUNCTION OF ω



(b) BECOMES MAXIMALLY FLAT AS A FUNCTION OF y IF

$$p = x + jy, \quad p = S + \frac{1}{S}$$

Fig. 5.15. Response functions when ω_c/B_n is less than 20.

original correctly adjusted response is said to be *maximally flat*. This terminology will be used to mean the following:

(1) The value of ω_c/B_n is 20 or more and the response is arithmetically symmetrical about ω_c.

(2) The response is the flattest obtainable by adjustment of the individual stages.

When ω_c/B_n is less than 20, another situation prevails. For any value of n, ω_c, and B_n we will find that there is one way of adjusting the individual stages so that the response is geometrically symmetrical about ω_c and flat as shown in figure (5.15a). Any change in the adjustment of the individual stages will cause the response to become lumpy or to tilt. This is also a *flat* function, but since the high Q approximation was not involved, it is called the *exact flat* function. You will observe from figure (5.15b) that the function can be made maximally flat as a function of y by using the transformation of variables discussed in the preceding section. This is an important aspect of a design procedure developed in a later section.

5.9. The Maximally Flat Function

It was shown in section (5.4) that the simple process of cascading identical single tuned amplifiers will not necessarily produce the desired combination of gain and bandwidth. In many cases the over-all bandwidth and gain requirements are so extreme that the synchronous connection will not yield the specified gain-bandwidth product, even if double tuned stages are used. When this occurs, resort is made to *stagger tuning*.

In this section we will specify the character of an over-all gain function that will have a better gain-bandwidth product than synchronous tuning can produce. It will be shown that this function can be synthesized by a cascade of stagger tuned stages.

It turns out that one of the best gain-bandwidth products is obtained by using the equation for *maximal flatness*. The maximally flat function is defined in terms of the complex frequency p, where $p = x + jy$. Therefore, in the steady state, the maximally flat function is

$$|g(jy)| = \frac{1}{\sqrt{1 + y^{2n}}} \tag{5.61}$$

where n = any integer = 1, 2, 3, and so on.

There is nothing particularly advantageous about the form of this function except that it has a good gain-bandwidth product. Maximal flatness is *not* necessarily an ideal response characteristic, because it invariably leads to overshoot in the transient response (see chapter 6). Maximal flatness is considered to be the penalty paid for achieving a good gain-bandwidth product.

The characteristics of the maximally flat function are shown in figure (5.16) in terms of the transformed frequency y. Note that the bandwidth is always 2, in terms of the frequency y, regardless of the value of n. Although the bandwidth is constant, the sharpness of the cutoff characteristic increases with larger values of n.

We want to synthesize the maximally flat function, using ordinary amplifiers. To do this, we must factor the maximally flat function by computing the poles; each pole represents the contribution of a single amplifier stage. Then the design of the individual amplifier stages is

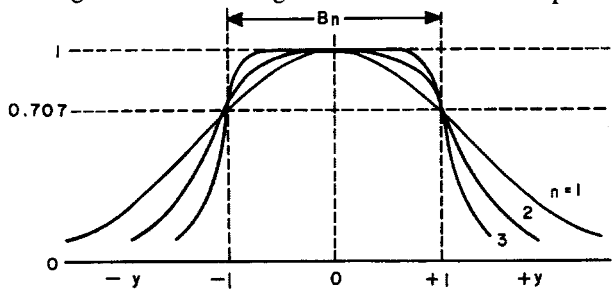


Fig. 5.16. Characteristics of the maximally flat function.

performed by establishing an equivalence between the poles of the maximally flat function and the poles of the amplifier cascade. So you can see that we must do two things:

(1) Factor the maximally flat function and determine the poles in the complex p plane.

(2) Compute the poles of a single tuned amplifier and express them so they can be located in the complex p plane.

These two problems are treated separately in the next two sections.

5.10. Poles of the Maximally Flat Function[2]

The maximally flat function was defined in the preceding section as

$$|g(jy)| = \frac{1}{\sqrt{1 + y^{2n}}}$$

where y is the imaginary part of the complex frequency p. That is, $p = x + jy$. We want to determine the p plane poles of the maximally function that fall in the *left half plane*, thereby insuring that the poles are physically realizable.

You can see from the form of the function that there must be n poles and these poles may be either real or complex. So let p_0 = real poles; p_m = complex poles. Of course, complex poles always occur in conjugate pairs.

[2] The method of factoring used here follows that given by Valley and Wallman, *Vacuum Tube Amplifiers*, vol. 18, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, pp. 176–180. Also see V. D. Landon, "Cascade Amplifiers With Maximal Flatness," *RCA Rev.*, 1941, pp. 347–362.

It is not easily and directly provable, but you can show by trial and error that *all* the poles are complex when n is even. Therefore, when n is even, the poles of the maximally flat function will have the form

$$p_1, p_1', p_2, p_2', \ldots, p_m, p_m', \ldots, p_{n/2}, p_{n/2}'$$

where the primes indicate conjugates.

In a similar way you can also show that there is only *one real pole* when n is odd. So, for odd values of n, the poles are

$$p_0, p_1, p_1', p_2, p_2', \ldots, p_m, p_m', \ldots, p_{(n-1)/2}, p_{(n-1)/2}'$$

We will discuss only the *even* case here, but the same general procedure is followed for odd values of n.

We have specified that the maximally flat function is to have complex conjugate poles located in the left half plane when n is even. Therefore the function can be factored in terms of these poles as

$$|g(jy)| = \frac{1}{|(jy - p_1)(jy - p_1')| \ldots |(jy - p_n/2)(jy - p_n'/2)|}$$

or, in a more symbolic form as

$$|g(jy)| = \prod_{m=1}^{m=n/2} \frac{1}{|(jy - p_m)(jy - p_m')|} \tag{5.62}$$

The large π symbol signifies the *product* of the series of factors obtained by successively inserting values of m from 1 to $n/2$.

It really does not look as though we have done much except confuse things because we still do not know the values for the poles p_m. Unfortunately, there is no simple method of finding the poles directly. However, there is a fairly simple mathematical trick that helps a great deal.

First square the maximally flat function so that we get

$$|g(jy)|^2 = \frac{1}{1 + y^{2n}}$$

The roots of the denominator of this function are easily computed by setting the denominator equal to zero. This gives

$$y^{2n} + 1 = 0 \quad \text{or} \quad y^2 = (-1)^{1/n}$$

Therefore there are n roots of (-1) and we shall designate them $r_1, r_1', r_2, r_2', \ldots, r_m, r_m', \ldots, r_{n/2}, r_{n/2}'$. Because n is even in this case, the n roots of (-1) are all complex and occur in conjugate pairs. So

$$r_m = (-1)^{1/n}$$

Therefore the square of the maximally flat function can be expressed in factored form in terms of these roots of (—1) as

$$|g(jy)|^2 = \frac{1}{(y^2 - r_1)(y^2 - r_1')\ldots(y^2 - r_{n/2})(y^2 - r_{n/2}')}$$

or, in an alternative form as

$$|g(jy)|^2 = \prod_{m=1}^{n/2} \frac{1}{(y^2 - r_m)(y^2 - r_m')}$$

Now return to the original factored form of the maximally flat function given in product notation in equation (5.62). Square it and write

$$|g(jy)|^2 = \prod_{m=1}^{n/2} \frac{1}{|(jy - p_m)(jy - p_m')|^2}$$

If the maximally flat function can actually be factored as we have proposed, the two equations for $|g(jy)|^2$ must be equal. They can be equal only if

$$|(jy - p_m)(jy - p_m')|^2 = (y^2 - r_m)(y^2 - r_m')$$

or, expanding these terms, if

$$y^4 + (p_m^2 + p_m'^2)y^2 + p_m^2 p_m'^2 = y^4 - (r_m + r_m')y^2 + r_m r_m'$$

This equation will be true only if $r_m = -p_m^2$. But

$$r_m = (-1)^{1/n}$$

so that

$$p_m^2 = -(-1)^{1/n} = (-1)(-1)^{1/n}$$

or

$$p_m = [(-1)^{n+1}]^{1/2n} \tag{5.63}$$

Exactly the same result is obtained when n is odd.

Therefore equation (5.63) gives the poles p_m of the maximally flat function as the $2n$ roots of $(-1)^{n+1}$ that fall in the left of the complex p plane. From this you can see that when n is even, the poles are the $2n$ roots of (—1) in the left half of the p plane. When n is odd, the poles are the $2n$ roots of (+1) in the left of the p plane.

The roots of +1 and —1 all have a magnitude of 1 and can be expressed in rectangular form as

$$p_m = -\cos\theta_m \pm j\sin\theta_m \tag{5.64}$$

where θ_m = pole angle measured from the negative real axis in the p plane.

The proper values for θ_m are easily computed by the well-known methods of complex algebra. The results are:

(1) when n is even

$$\theta_m = \pm\left(\frac{2m-1}{2n}\right)\pi; \qquad m = 1, 2, 3, \ldots, (n-2) \qquad (5.65)$$

(2) when n is odd

$$\theta_m = \pm\left(\frac{m}{n}\right)\pi;$$

$$m = 0, 1, 2, 3, \ldots, (n-2) \qquad (5.66)$$

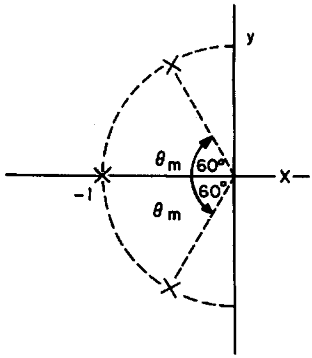


Fig. 5.17. Poles of the maximally flat function when $n=3$.

For example, when n is 3, equation (5.66) tells us that the three poles in the left half plane have angles of 0° and ±60°, all measured from the negative real axis, and a unit distance out from the origin. This is illustrated in figure (5.17). You can easily work out pole plots for any value of n simply by using equations (5.65) and (5.66).

We can summarize our accomplishments by observing that we have successfully factored the maximally flat function into the form

$$|g(jy)| = \frac{1}{\sqrt{1+y^{2n}}} = \prod_{m=1}^{n/2} \frac{1}{|(jy-p_m)(jy-p_m')|} \qquad (5.67)$$

or, because $p = jy$ in the steady state, in the general case

$$|g(p)| = \frac{1}{\sqrt{1+y^{2n}}} = \prod_{m=1}^{n/2} \frac{1}{|(p-p_m)(p-p_m')|}$$

The values of the poles are computed from equations (5.63) through (5.66).

It was shown in chapter 4 that every amplifier gain function is characterized by a particular configuration of poles in the complex plane. Then, it was shown at the beginning of this chapter that an array of poles could be synthesized by cascading amplifier stages so that each amplifier stage contributes one or more poles to the over-all array of poles of the system gain function. This was called the technique of synthesis by factoring.

Evidently then, if a maximally flat function is to be synthesized by an amplifier cascade, it is necessary to establish a method of specifying the locations of the poles of the individual amplifier stages so that they combine to produce an over-all pole configuration identical to that of the maximally flat function. Once the amplifier poles are specifically located, the individual stages are easily designed, because the pole location is governed by the constants of the amplifier circuit.

The necessary equivalence between amplifier poles and the poles of the maximally flat function is established in the next section.

5.11. Maximally Flat Staggered *n*-uples

It was shown in the preceding section that the equation for maximal flatness

$$|g(jy)| = \frac{1}{\sqrt{1 + y^{2n}}} \tag{5.68}$$

could be synthesized by using a cascade connection of n amplifier stages if an equivalence could be established between the poles of this function and the poles of the amplifier cascade.

The poles of the maximally flat function were shown to be of the form

$$p_m = -\cos\theta_m \pm j\sin\theta_m \tag{5.69}$$

The *factor* of the function from which this pole originated is easily written as

$$(p) = K\left(\frac{1}{p - p_m}\right) \quad \text{where} \quad p = x + jy \tag{5.70}$$

Thus, in the steady state case when $p = jy$, the factor becomes

$$f(jy) = K\left(\frac{1}{jy - p_m}\right) = K\frac{1}{\cos\theta_m + j(y \pm \sin\theta_m)} \tag{5.71}$$

It was shown in chapter 4 that the gain function of a high Q single tuned amplifier stage could be expressed as

$$A(j\omega) = -\frac{g_m}{C_T}\left(\frac{1}{B \pm j2\Delta\omega}\right) \tag{5.72}$$

where $B = 1/RC_T =$ stage bandwidth; $\Delta\omega = \omega - \omega_0 =$ frequency deviation from ω_0; $\omega =$ signal frequency; $\omega_0 = 1/\sqrt{LC_T} =$ bandcenter of the stage.

It is evident that the form of this factor is similar to that of the factor of the maximally flat function. However, the factor of the flat function is dimensionless, while the factor in the amplifier gain function has the dimensions of frequency. However, it can be made dimensionless through the use of a *normalizing* procedure. This is explained in the next two paragraphs.

It is assumed that n amplifier stages are cascaded and adjusted so that the over-all gain function is maximally flat as shown in figure (5.18). The over-all bandwidth of the cascade is B_n and the center frequency is ω_c. Now suppose that this function is to be constructed using high Q, single tuned amplifier stages.

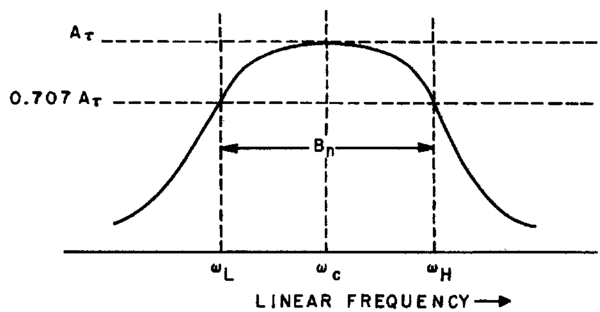


Fig. 5.18. Terminology for the maximally flat amplifier cascade.

In the single stage gain function given in equation (5.72), divide numerator and denominator through by the over-all bandwidth B_n of the cascade. This gives

$$A(j\omega) = -\frac{g_m}{B_n C_T} \cdot \frac{1}{B/B_n \pm j2\Delta\omega/B_n} \tag{5.73}$$

Now assume that the signal frequency is set equal to the bandcenter frequency ω_c of the cascade. Thus the gain of each amplifier stage becomes

$$A(j\omega_c) = -\frac{g_m}{B_n C_T} \cdot \frac{1}{B/B_n \pm j2\Delta\omega_c/B_n} \tag{5.74}$$

where $\Delta\omega_c = \omega_c - \omega_0 =$ frequency difference between the bandcenter of the cascade and bandcenter of the amplifier stage.

At the bandcenter for the maximally flat function, $y = 0$, and each factor of the function becomes

$$f(0) = K\frac{1}{\cos\theta_m \pm j\sin\theta_m} \tag{5.75}$$

From a term-by-term comparison of equations (5.74) and (5.75) it is evident that there is one-to-one correspondence. That is

$$\frac{B}{B_n} = \cos\theta_m \tag{5.76}$$

which is the real part of the pole of the maximally flat function.

$$\frac{2\Delta\omega_c}{B_n} = \sin\theta_m \tag{5.77}$$

which is the imaginary part of the pole of the maximally flat function.

$$K = \frac{g_m}{B_n C_T} = \text{scale factor}$$

Therefore, by proper adjustment of B and $\Delta\omega_c$, it is possible to make the amplifier poles exactly equal to the poles of the maximally flat function. For any value of n, B_n, and ω_c it would be possible to synthesize n single tuned amplifier stages to produce the same pole configuration as that of the maximally flat function. The relationships required to compute the angles θ_m were given previously in equations (5.66) and (5.67).

Generally speaking, the center frequencies of the individual stages are not equal to the bandcenter frequency of the cascade. Hence, the stages are said to be *staggered*. Because the staggering produces a maximally flat function, the cascade is called a *maximally flat n-uple*. The use of the term *n-uple* should be clear from table 8.

TABLE 8

MAXIMALLY FLAT STAGGERED n-UPLES

n = *Number of stages*	*Name of circuit*	*Gain function*
2	staggered pair	$K_2/(1 + y^4)^{1/2}$
3	staggered triple	$K_3/(1 + y^6)^{1/2}$
4	staggered quadruple	$K_4/(1 + y^8)^{1/2}$
5	staggered quintuple	$K_5/(1 + y^{10})^{1/2}$
n	staggered n-uple	$K_n/(1 + y^{2n})^{1/2}$

It is a relatively simple matter now to evaluate and tabulate the data necessary to design any maximally flat staggered n-uple by locating the poles as specified by equations (5.64) through (5.66). Then by evaluating the real and imaginary parts of the poles, the stage bandwidths

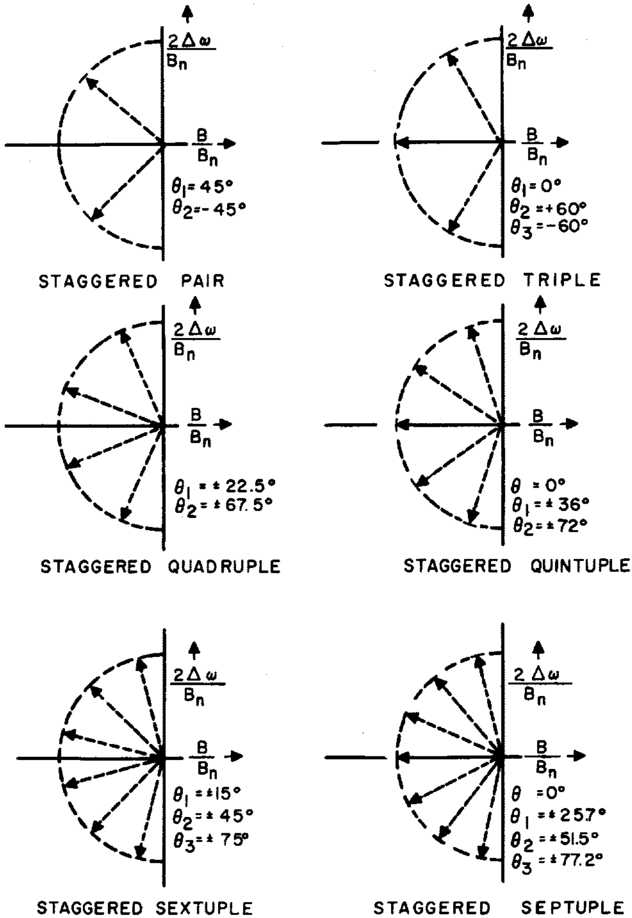


Fig. 5.19. Pole locations for maximally flat staggered n-uples.

and resonant frequencies can be found; this procedure is illustrated for several different values of n in figure (5.19) and the resulting data are summarized in table 9.

It is particularly interesting to note that the over-all bandwidth of the cascade is always equal to or greater than the bandwidth of any given single stage of the cascade. This is in sharp contrast with the synchronous connection.

The material presented in this article is valid as long as the Q's of the amplifier tuned circuits are about 20 or larger.

TABLE 9

DESIGN DATA FOR MAXIMALLY FLAT STAGGERED n-UPLES
[Computed from figure (5.19)]

n	Name of circuit	Number of stages	Center frequency of the stage	Stage bandwidth
2	Staggered pair	2	$\omega_c \pm 0.35B_n$	$0.71B_n$
3	Staggered triple	2	$\omega_c \pm 0.43B_n$	$0.50B_n$
		1	ω_c	$1.00B_n$
4	Staggered quadruple	2	$\omega_c \pm 0.46B_n$	$0.38B_n$
		2	$\omega_c \pm 0.19B_n$	$0.92B_n$
5	Staggered quintuple	2	$\omega_c \pm 0.29B_n$	$0.81B_n$
		2	$\omega_c \pm 0.48B_n$	$0.26B_n$
		1	ω_c	$1.00B_n$
6	Staggered sextuple	2	$\omega_c \pm 0.48B_n$	$0.26B_n$
		2	$\omega_c \pm 0.35B_n$	$0.71B_n$
		2	$\omega_c \pm 0.13B_n$	$0.97B_n$
7	Staggered septuple	2	$\omega_c \pm 0.49B_n$	$0.22B_n$
		2	$\omega_c \pm 0.39B_n$	$0.62B_n$
		2	$\omega_c \pm 0.22B_n$	$0.90B_n$
		1	ω_c	$1.00B_n$

The one remaining factor of interest is the over-all gain of the staggered cascade. According to equation (5.74), the function for maximal flatness can be synthesized by using a staggered cascade of amplifier stages having gain functions of the form

$$A(j\omega) = -\frac{g_m}{B_nC_T}\cdot\frac{1}{B/B_n \pm j2\Delta\omega_c/B_n} \tag{5.78}$$

or, in an alternative form,

$$A(j\omega) = -\frac{F_a}{B_n} \cdot \frac{1}{B/B_n \pm j2\Delta\omega_c/B_n} \tag{5.79}$$

Moreover, it was shown that

$$\frac{B}{B_n} \pm j\frac{2\Delta\omega_c}{B_n} = \cos\theta_m \pm j\sin\theta_m \tag{5.80}$$

so that the gain function for the amplifier stage can be written

$$A(j\omega) = -\frac{F_a}{B_n} \cdot \frac{1}{\cos\theta_m \pm j\sin\theta_m} \tag{5.81}$$

Consequently, if n stages are cascaded, the magnitude of the over-all gain equation is

$$A(j\omega) = \left(\frac{F_a}{B_n}\right)^n \frac{1}{\sqrt{1 + y^{2n}}} \tag{5.82}$$

Therefore the over-all reference gain at the bandcenter frequency, $y = 0$, is obviously

$$A_T = \left(\frac{F_a}{B_n}\right)^n \tag{5.83}$$

Because the bandwidth of the maximally flat function is independent of the value of n, the over-all bandwidth remains constant as additional stages are added to the cascade. However, the over-all gain increases with the addition of more stages, so that the over-all gain-bandwidth product of the amplifier increases continuously as n is increased.

5.12. Exact Flat Staggering, Discussion

Section (5.10) dealt with the synthesis of the maximally flat function from a cascade of staggered, high Q, single tuned amplifiers. The design method developed is generally valid only for circuit Q's of 20 or more. It may not be possible in every case to achieve the specifications of the design by using such high Q circuits, so that we should investigate the exact case in which the high Q approximations are not involved.

At the outset it is helpful to consider the gain function of a single tuned amplifier as a function of complex frequency. This equation was derived in chapter 4 and can be written

$$A(s) = -\frac{g_m}{C_T} \frac{s}{s^2 + s/RC_T + 1/LC_T} \tag{5.84}$$

It was also shown that

$$B = \frac{1}{RC_T} = \text{stage bandwidth} \tag{5.85}$$

$$\omega_0 = \sqrt{\frac{1}{LC_T}} = \text{resonant frequency of stage} \tag{5.86}$$

Hence the general gain equation can be expressed as

$$A(s) = -\frac{g_m}{C_T}\left(\frac{s}{s^2 + Bs + \omega_0^2}\right) \tag{5.87}$$

It is clear that this function has a zero at the origin and two conjugate poles given by

$$s_{1,2} = -\frac{B}{2} \pm j\sqrt{\omega_0^2 - \frac{B^2}{4}} \tag{5.88}$$

or

$$s_{1,2} = -\frac{B}{2} \pm j\omega_0\sqrt{1 - \left(\frac{B}{2\omega_0}\right)^2} \tag{5.89}$$

And finally

$$s_{1,2} = -\frac{B}{2} \pm j\omega_n \tag{5.90}$$

where ω_n is the imaginary component of equation (5.89).

The magnitude of the pole is equal to the radial distance from the origin of coordinates in the s plane to the pole. It is easily shown that this is

$$|s_{1,2}| = \omega_0 \tag{5.91}$$

Consequently, the poles are located in the complex s plane as shown in figure (5.20). Clearly, if the pole locations are known, ω_0 and B can be quickly computed. Then, because

$$B = \frac{1}{RC_T} \quad \text{or} \quad R = \frac{1}{BC_T} \tag{5.92}$$

$$\omega_0^2 = \frac{1}{LC_T}; \qquad L = \frac{1}{\omega_0^2 C_T} \tag{5.93}$$

a knowledge of the pole location permits the stage to be designed. Hence the design of a single tuned stage merely involves the specification of particular pole locations. Once the poles are known, the design of the amplifier is clear cut.

For the high Q circuit, the half bandwidth $B/2$ is so small compared with ω_n that $\omega_n \doteq \omega_0$. Because of this approximation it was possible to specify the pole locations in such a way that each single tuned stage provided one factor in the synthesis of the maximally flat function.

It would be extremely desirable to establish some similar method of specifying the pole locations in the exact case in which the individual amplifier stages exhibit geometric symmetry. Because of the simplicity and directness of the procedure developed for the case of arithmetic symmetry, it would be helpful if the results of that process could be applied to the solution of the exact case.

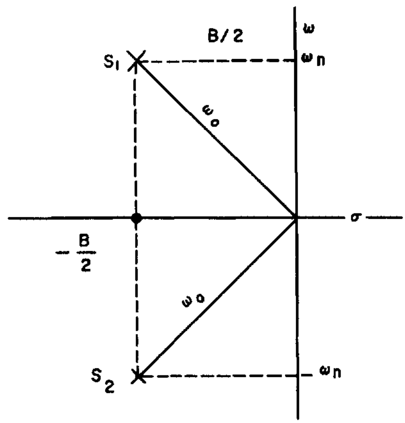


Fig. 5.20. Poles of a single tuned amplifier in the s plane.

It was shown in section (5.7) that a geometrically symmetrical function of the variable $S = s/\omega_c$ can be transformed into a function that exhibits maximal flatness with arithmetic symmetry if the variable is changed to p, where

$$p = S + 1/S = x + jy \tag{5.94}$$

Thus
$$S = \frac{s}{\omega_c} = \frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - 1} \tag{5.95}$$

Hence if an exact flat staggered n-uple is to be designed, the first step is to design it as a *maximally* flat amplifier as a function of the variable p. Then the variable is transformed into the s plane, using equation (5.95). Once the pole locations in the s plane are known, the amplifier is readily designed from a knowledge of figure (5.20).

The formal design procedure for the exact flat case is outlined in the next section.

5.13. Exact Flat Staggering, Design Procedure

Three factors must be specified before an exact flat staggered amplifier can be designed. They are:

(1) B_n = over-all bandwidth.
(2) ω_c = bandcenter frequency.
(3) n = number of stages.

The design procedure then follows the outline given below.

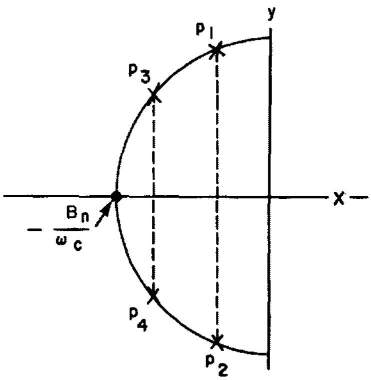


Fig. 5.21. Poles of the maximally flat function in the p plane.

(1) Construct the complex P plane.
   (a) Using the origin of coordinates of the plane as the center, draw a half circle of radius B_n/ω_c.
   (b) For the value of n specified, locate the poles in the P plane as shown in figure (5.19).
   (c) Record the value of each pole as $p_m = (-x_m \pm jy_m)B_n/\omega_c$.

(2) Construct the complex s plane.
   (a) Compute the poles of the exact flat stagger from the P plane poles by using the transformation given in equation (5.95).
   (b) Each P plane pole will produce two s plane poles.
   (c) Locate these poles in the s plane.
      (i) Pole magnitude $= \omega_0 = \sqrt{1/LC_T}$.
      (ii) Real component $= B/2 = 1/2RC_T$.

All data necessary to design the amplifier are now available.

In the complex P plane, the poles lie on the periphery of a half circle in the left half plane as shown in figure (5.21). When the poles are transformed to the s plane they fall on warped circles as shown in figure (5.22), where these circles are located in the second and third quadrants. Conjugate poles in the P plane produce s plane poles that fall on the same radial lines as shown in figure (5.22).

A special case develops whenever the stagger contains an odd number of stages. In these cases, one pole in the p plane falls on the negative real axis, at -1, and this often leads to two real poles in the s plane as shown in figure (5.23). The immediate question is the determination of the center frequency and bandwidth for this stage.

In such cases, the pole in the P plane is $P_m = -B_n/\omega_c$. Consequently, the corresponding s plane poles are

$$s_m = -\frac{B_n}{2} \pm j\omega_c\sqrt{1 - \left(\frac{B_n}{2\omega_c}\right)^2}$$

The general expression for the poles of a single tuned amplifier was derived in equation (5.89) as

$$s_{1,2} = -\frac{B}{2} \pm j\omega_0\sqrt{1 - \left(\frac{B}{2\omega_0}\right)^2}$$

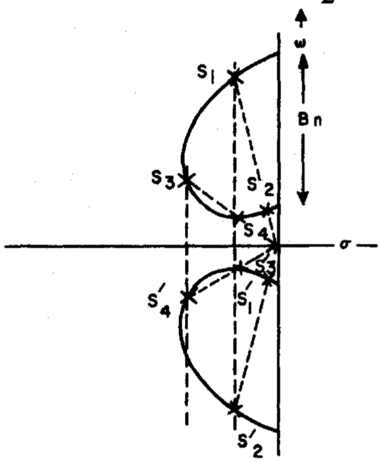


Fig. 5.22. Poles in the s plane for exact flat staggering.

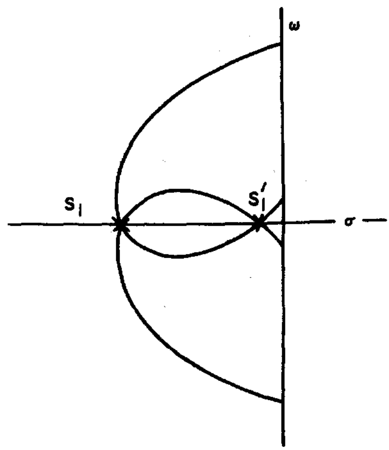


Fig. 5.23. Effect of a real pole in the P plane on poles in the s plane.

From a term-by-term comparison of these two equations it is clear that whenever n is odd there is one stage for which

$$\text{natural frequency} = \omega_0 = \omega_c = \text{bandcenter} \tag{5.96}$$

$$\text{stage bandwidth} = B = B_n = \text{bandwidth of stagger} \tag{5.97}$$

5.14. Overstaggering

The use of flat staggering as discussed in the preceding sections produces a considerable increase in the gain-bandwidth product over that possible with synchronous tuning. Now suppose that a maximally flat-staggered n-uple is designed, using high Q single tuned amplifiers. Further suppose that the resulting over-all gain is insufficient. In an effort to increase the gain, the circuit Q's of the individual stages are increased beyond the values required for maximal flatness. The reference gain of the cascade will be increased, but it will be found that the frequency response characteristic has developed a "lumpy" appearance. Instead of being flat, it now exhibits peaks and dips. The circuit is then said to be *overstaggered.*

There is nothing especially undesirable about this lumpiness as long as it is under control. It was stated earlier that there was nothing especially advantageous about the maximally flat function except that it had a good gain-bandwidth product. This product can be increased by introducing small peaks and dips into the response

TABLE 10

TCHEBICHEF POLYNOMIALS

n	$C_n(y)$
1	y
2	$2y^2 - 1$
3	$4y^3 - 3y$
4	$8y^4 - 8y^2 + 1$
5	$16y^5 - 20y^3 + 5y$

characteristic. However, the amplifier phase response is worse and there is considerable overshoot in the transient response (see chapter 6).

The most common form of controlled lumpiness is called the *equal ripple function.* Here n stages of narrow band, single tuned amplifiers are cascaded and overstaggered so that the resulting over-all gain function has an amplitude of the form

$$|g(jy)| = K \frac{1}{\sqrt{1 + \epsilon C_n^2(y)}} \tag{5.98}$$

This is the equal ripple function.

The function $C_n(y)$ appearing in equation (5.98) is a Tchebichef polynomial, and it has the forms given in table 10. The gain functions corresponding to the various values of n are sketched in figure (5.24).

If the equal ripple function is factored, it can be shown that it will have poles of the following form:

(1) When n is even,

$$p_m = \sinh \left[-\alpha + j \frac{\pi}{2n} (1 + 2k) \right] \tag{5.99}$$

where $k = 0, -1$ for $n = 2$; $k = 0, \pm 1, -2$ for $n = 4$.

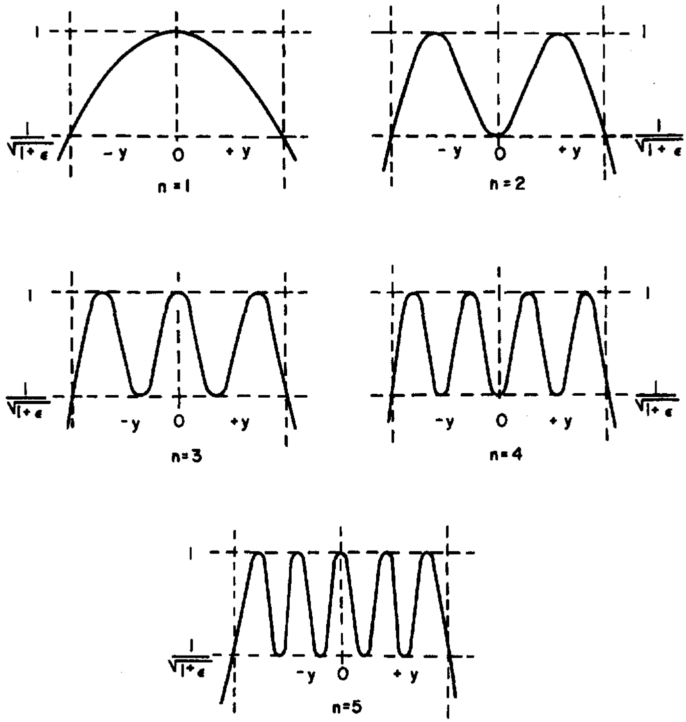


Fig. 5.24. Equal ripple gain functions.

(2) When n is odd,

$$p_m = \sinh \left[-\alpha + j\frac{\pi k}{2n} \right] \tag{5.100}$$

where $k = 0, \pm 1$ for $n = 3$; $k = 0, \pm 1, \pm 2$ for $n = 5$.
It is clear that the poles have the general form

$$p_m = \sinh(-\alpha + j\beta) = -\sinh \alpha \cos \beta + j \cosh \alpha \sin \beta \tag{5.101}$$

Divide through by cosh α:

$$\frac{p_m}{\cosh \alpha} = -\tanh \alpha \cos \beta + j \sin \beta \tag{5.102}$$

If this last equation is compared with the general equation for the poles of the maximally flat function, it is apparent that they have exactly the same form except that the real part is multiplied by tanh α in the equal ripple case. Hence the equal ripple function can be synthesized by systematically overstaggering the stages designed to produce maximal flatness.

In passing it is also important to observe that overstaggering increases the slope of the skirts of the selectivity curve, making it steeper than the skirt of the maximally flat response. This increase in the sharpness of the cutoff characteristic improves the selectivity characteristic so that delineation of adjacent channels is more readily secured.

The design procedure may be outlined briefly as follows:

(1) The design requirements will specify the desired bandcenter ω_c, over-all bandwidth B_n, amount of ripple, and the number of stages n.

(2) Calculate ϵ from the relationship for the ripple

$$\text{volt logits} = 10 \log_{10}\sqrt{1+\epsilon} \tag{5.103}$$

(3) Calculate α from

$$\alpha = \frac{1}{n}\sinh^{-1}\sqrt{\frac{1}{\epsilon}} \tag{5.104}$$

(4) Compute $\tanh \alpha$.

(5) Construct the complex p-plane and locate the poles of the maximally flat function as outlined in section (5.10). For each pole, multiply the real part by $\tanh \alpha$. The resulting pole is a pole of the equal ripple function.

$$\text{real part} = \text{normalized stage bandwidth} = B/B_n$$

$$\text{imaginary part} = \frac{2}{B_n}\begin{pmatrix}\text{deviation of stage resonant}\\ \text{frequency from bandcenter}\end{pmatrix}$$

It is apparent that the center frequencies of the stages used to synthesize the equal ripple function are exactly the same as those used to synthesize the maximally flat function. However, the stage bandwidths are reduced by the $\tanh \alpha$ factor. Although the over-all bandwidth remains the same as in the maximally flat case, the reduction in stage bandwidth increases the stage gain and the over-all figure of merit of the cascade.

It can be shown that the reference gain of the cascade is

$$A_T = \left(\frac{F_a}{B_n}\right)^n \prod_{m=1}^{n} \frac{1}{\sqrt{(\tanh \alpha \cos \beta_m)^2 + (\sin \beta_m)^2}} \tag{5.105}$$

5.15. Cascading of Flat-Staggered n-uples

It has been shown that n high Q, single tuned amplifier stages can be cascaded and staggered to produce a maximally flat frequency response. For a maximally flat n-uple,

$$|A_T(jy)| = \left(\frac{F_a}{B_n}\right)^n \frac{1}{\sqrt{1+y^{2n}}} \tag{5.106}$$

Now consider the staggered n-uple as a single unit in a cascade of m such staggers. In other words, a total of mn tubes are arranged in a cascade of m stagger tuned amplifiers, each stagger containing n tubes. The over-all system gain is

$$|A_T(jy)|^m = \left(\frac{F_a}{B_n}\right)^{mn} \frac{1}{(1+y^{2n})^{m/2}} \tag{5.107}$$

At the over-all cutoff frequencies

$$\frac{1}{(1+y^{2n})^{m/2}} = \frac{1}{2^{1/2}} \tag{5.108}$$

or

$$(1+y^{2n})^m = 2 \tag{5.109}$$

Take the mth root of both sides of this equation and subtract 1. This yields

$$y^{2n} = 2^{1/m} - 1 \tag{5.110}$$

Solve for the cutoff frequencies y.

$$y = \pm(2^{1/m} - 1)^{1/2n} \tag{5.111}$$

Therefore the upper and lower cutoff frequencies are

$$\begin{aligned} y_H &= (2^{1/m} - 1)^{1/2n} \\ &= \text{over-all upper cutoff frequency} \end{aligned} \tag{5.112}$$

$$\begin{aligned} y_L &= -(2^{1/m} - 1)^{1/2n} \\ &= \text{over-all lower cutoff frequency} \end{aligned} \tag{5.113}$$

The over-all bandwidth B_n is the difference between these cutoff frequencies, or

$$B_m = 2(2^{1/m} - 1)^{1/2n} \tag{5.114}$$

But $B_n =$ bandwidth of each n-uple $= 2$ in terms of the normalized frequency y. Hence the over-all bandwidth of the cascade of staggered n-uples is

$$B_m = B_n(2^{1/m} - 1)^{1/2n} \tag{5.115}$$

This is a convenient equation to use for the over-all bandwidth of any cascade connection of high Q, single tuned amplifiers regardless of whether they are synchronous or stagger tuned. For example, in the case of the synchronous connection, the number of tubes in each staggered n-uple is 1. Hence the over-all bandwidth is

$$B_m = B(2^{1/m} - 1)^{1/2} \tag{5.116}$$

which was obtained earlier for the same case by direct evaluation. For a staggered pair, $n = 2$, so that

$$B_m = B_n(2^{1/m} - 1)^{1/4} \tag{5.117}$$

For a staggered triple, $n = 3$ and

$$B_m = B_n(2^{1/m} - 1)^{1/6} \tag{5.118}$$

and so on. It should be clear that the shrinking of the over-all bandwidth proceeds at a slower rate as the number of stages in the staggered n-uple is increased.

5.16. Double Tuned Amplifier: Equal Q, High Q Case

While the double tuned amplifier is only a single stage, it exhibits many of the characteristics of cascaded single tuned amplifiers. Therefore the design of the circuit is included here rather than in chapter 4.

The general equation for the gain function of a double tuned amplifier was derived in chapter 4 and the result is reproduced in equation (5.119). This equation is valid only when the primary and secondary are tuned to the same frequency.

$$A(s) = -\frac{g_m k}{C_1C_2(1 - k^2)\sqrt{L_1L_2}} \cdot \frac{s}{s^4 + a_1s^3 + a_2s^2 + a_3s + a_4} \tag{5.119}$$

where the coefficients of the characteristic function are

$$a_1 = \omega_r\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \tag{5.120}$$

$$a_2 = \omega_r^2\left(\frac{1}{Q_1Q_2} + 2\right) \tag{5.121}$$

$$a_3 = \omega_r^3\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) \tag{5.122}$$

$$a_4 = \omega_r^4(1 - k^2) \tag{5.123}$$

The terms in these coefficients were previously shown to be

$$Q_1 = \omega_0 R_1 C_1 \tag{5.124}$$

$$Q_2 = \omega_0 R_2 C_2 \tag{5.125}$$

$$\omega_r = \frac{\omega_0}{\sqrt{1 - k^2}} \tag{5.126}$$

$$\omega_0 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}} \tag{5.127}$$

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{5.128}$$

The coefficients of the characteristic function can be simplified somewhat when the primary and secondary Q's are equal to one another. Thus if $Q = Q_1 = Q_2$, the coefficients are

$$a_1 = \frac{2\omega_r}{Q} \tag{5.129}$$

$$a_2 = \omega_r^2\left(\frac{1}{Q^2} + 2\right) \doteq 2\omega_r^2 \tag{5.130}$$

$$a_3 = \frac{2\omega_r^3}{Q} \tag{5.131}$$

$$a_4 = \omega_r^4(1 - k^2) \tag{5.132}$$

The approximate value for a_2 given in equation (5.130) is applicable when the circuit Q is so high that $1/Q^2$ is much less than 2. This condition is assumed to exist in the work that follows.

Using the coefficients obtained for the high Q, equal Q case, the characteristic equation for the amplifier can be written directly as

$$s^4 + \frac{2\omega_r}{Q}s^3 + 2\omega_r^2 s^2 + \frac{2\omega_r^3}{Q}s + \omega_r^4(1 - k^2) = 0 \tag{5.133}$$

The four roots of this equation are the four poles of the gain function. The natures of the roots are easily deduced because in the high Q case they must be complex conjugates. That is, the four roots are

$$s_1 = -\sigma_1 + j\omega_1 = \gamma_1\underline{/\theta_1} \tag{5.134}$$

$$s_1' = -\sigma_1 - j\omega_1 = \gamma_1\underline{/-\theta_1} \tag{5.135}$$

$$s_2 = -\sigma_2 + j\omega_2 = \gamma_2\underline{/\theta_2} \tag{5.136}$$

$$s_2' = -\sigma_2 - j\omega_2 = \gamma_2\underline{/-\theta_2} \tag{5.137}$$

where

$$\gamma_1^2 = \sigma_1^2 + \omega_1^2 \tag{5.138}$$

$$\gamma_2^2 = \sigma_2^2 + \omega_2^2 \tag{5.139}$$

If these are the roots of the characteristic equation, it is a relatively simple matter to express the coefficients of this equation in terms of the roots by using the standard relationships between roots and coefficients of polynomials. Therefore,

$$a_1 = -(\text{sum of the roots}) = 2(\sigma_1 + \sigma_2) \tag{5.140}$$

$$\begin{aligned} a_2 &= +(\text{sum of the products of roots 2 at a time}) \\ &= 4(\sigma_1\sigma_2 + \gamma_1^2 + \gamma_2^2) \end{aligned} \tag{5.141}$$

$$\begin{aligned} a_3 &= -(\text{sum of the products of roots 3 at a time}) \\ &= 2(\sigma_1\gamma_2^2 + \sigma_2\gamma_1^2) \end{aligned} \tag{5.142}$$

$$a_4 = +(\text{product of the roots}) = \gamma_1^2\gamma_2^2 \tag{5.143}$$

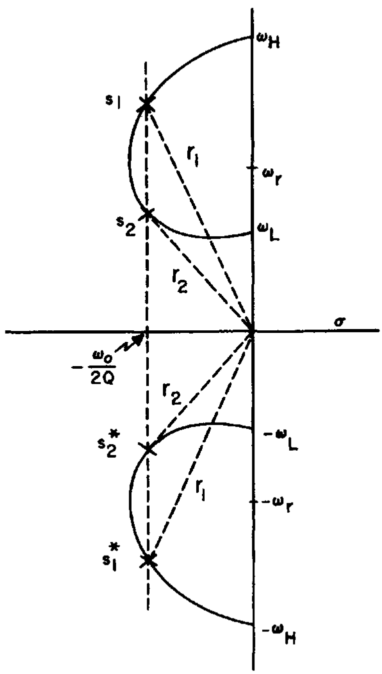


Fig. 5.25. Pole diagram for an equal Q double tuned amplifier.

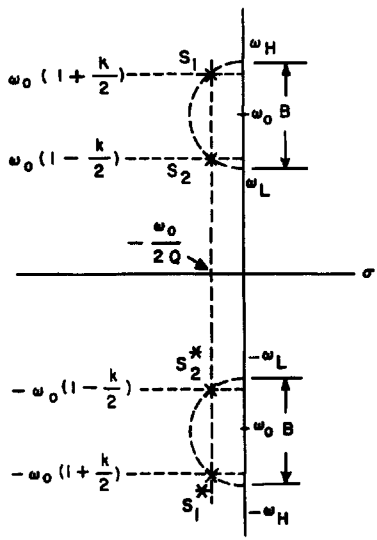


Fig. 5.26. Pole diagram for the high Q, equal Q, double tuned amplifier.

There are now two sets of equations for the coefficients, one involving the components of the poles and one set given in terms of the circuit constants of the amplifier. The poles can then be computed by equating the two sets of relationships and solving the results simultaneously.

In the equal Q case assumed, the real parts of the poles must be equal so that $\sigma = \sigma_1 = \sigma_2$. Thus the results of the computation outlined in the preceding paragraph are

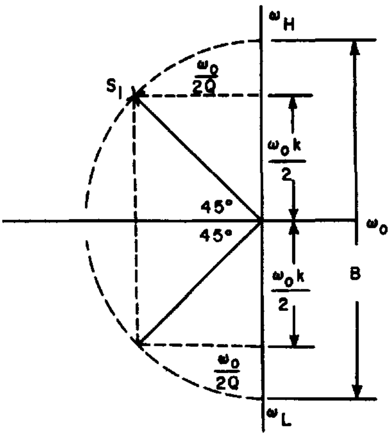


Fig. 5.27. Enlargement of the top part of figure (5.26).

$$\sigma = \frac{\omega_r}{2Q} = \frac{\omega_0}{2Q} \cdot \frac{1}{\sqrt{1-k^2}} \doteq \frac{\omega_0}{2Q} \tag{5.144}$$

$$\gamma_1 = \frac{\omega_0}{\sqrt{1-k}} \doteq \omega_0\left(1 + \frac{k}{2}\right) \tag{5.145}$$

$$\gamma_2 = \frac{\omega_0}{\sqrt{1+k}} \doteq \omega_0\left(1 - \frac{k}{2}\right) \tag{5.146}$$

$$\omega_0^2 = \frac{\omega_r^2}{1-k^2} \doteq \omega_r^2 \tag{5.147}$$

The approximate relationships given in these equations are valid as long as the circuit Q is high, because this makes k^2 much less than 1.

Equations (5.144) through (5.147) give the information necessary to locate the poles in the complex s plane. Figure (5.25) shows a typical pole diagram. Thus the amplifier can be easily designed as soon as the design requirements definitely fix the poles to specific points in the s plane. The high Q case has been assumed so that the real part of the poles, $\omega_0/2Q$, is small compared with the imaginary component; for all practical purposes the imaginary component is equal to the amplitude γ. Under the same conditions, $\omega_r \doteq \omega_0$, so that the poles can be located approximately, but with considerable accuracy as shown in figure (5.26).

Suppose, for example, that a high Q, equal Q, double tuned amplifier is to be designed to give a maximally flat response. Two pairs of conjugate poles in the s plane will be required to produce a flat function corresponding to the case of $n = 2$. For $n = 2$, it was shown in section (5.10) that the poles should be located at 45° angles as shown in figure (5.27), which is simply an enlargement of the region about ω_0.

The radial distance from ω_0 to either pole is clearly

$$\frac{\omega_0 k}{2 \sin 45^\circ} = \frac{\omega_0 k}{\sqrt{2}} \tag{5.148}$$

This is the radius of the half circle in figure (5.27), and the bandwidth B is twice the radius. Thus

$$B = \sqrt{2}\,\omega_0 k \tag{5.149}$$

For the pole angles to be 45°, the real and imaginary parts, measured from ω_0, must be equal. That is,

$$\frac{\omega_0}{2Q} = \frac{\omega_0 k}{2} \tag{5.150}$$

Consequently, for 45° angles, or a maximally flat response,

$$k_t = \frac{1}{Q} \tag{5.151}$$

which corresponds to the condition for transitional coupling.

The design procedure to obtain a maximally flat response is now easily summarized as follows:

(1) The design requirements will specify the values required for ω_r and B. Because the high Q case is assumed, $\omega_r \doteq \omega_0$. Moreover, the values of the primary and secondary capacitances C_1 and C_2 will also be known once the tube is selected.

(2) From equation (5.149), compute the coefficient of coupling.

(3) For maximal flatness,

$$Q = \frac{1}{k} = \omega_0 R_1 C_1 = \omega_0 R_2 C_2 \tag{5.152}$$

Hence R_1 and R_2 can be computed.

$$R_1 = \frac{Q}{\omega_0 C_1} = \frac{1}{k\omega_0 C_1} \tag{5.153}$$

$$R_2 = \frac{Q}{\omega_0 C_2} = \frac{1}{k\omega_0 C_2} \tag{5.154}$$

(4) Finally, because

$$\omega_0^2 = \frac{1}{L_1 C_1} = \frac{1}{L_2 C_2} \tag{5.155}$$

so

$$L_1 = \frac{1}{\omega_0^2 C_1} \tag{5.156}$$

$$L_2 = \frac{1}{\omega_0^2 C_2} \tag{5.157}$$

$$M = k\sqrt{L_1 L_2} \tag{5.158}$$

(5) The amplifier can now be designed with a tuned transformer or tee or pi section, whichever is deemed most appropriate.

If it is desirable to overcouple the circuit so that the response dips in the center, the amplifier should be designed as an equal ripple pair. The design procedure is essentially the same as that for transitional coupling, just given, except that the real parts of the poles are modified by the tanh α factor as outlined in section (5.13).

Still other possibilities exist. High Q, equal Q, double tuned amplifier stages can be cascaded and designed to give the same response as staggered quadruples, sextuples, or octuples. In such cases it turns out that all stages are tuned to the same frequency and the staggering is produced by variations in the loading, or circuit Q. This kind of amplifier is said to be *stagger-damped.*

When the high Q case does not apply, procedures have been developed for designing the amplifier to have the same response as an exact flat staggered pair. The conformal transformation between s and p planes is different from that used for the single tuned case.

5.17. Distributed Amplifiers,[3] General

The steady state figure of merit for a single stage video amplifier was given in chapter 4 as the gain-bandwidth product. That is, $F_a = A_r\omega_H$, where $A_r =$ reference gain; $\omega_H =$ upper cutoff frequency. It was shown that the value of this figure of merit for a resistance coupled amplifier was $F_a = g_m/C_T$, because

$$A_r = g_m R \quad \text{and} \quad \omega_H = \omega_2 = \frac{1}{RC_T}$$

In the analytical work on various amplifier types that followed the resistance coupled amplifier, it was shown that the same reference gain $g_m R$ could be achieved, but with higher cutoff frequencies through the use of improved circuitry. In all cases, the upper cutoff frequency was expressed in terms of a multiple of ω_2, the cutoff frequency for a resistance coupled amplifier. That is, for any video amplifier, it was shown that

$$A_r = g_m R \quad \text{and} \quad \omega_H = K\omega_2 = \frac{K}{RC_T}$$

[3] This discussion is adapted from E. L. Ginzton, W. R. Hewlett, J. H. Jasberg, J. D. Noe, "Distributed Amplifiers," *Proc. IRE,* vol. 36, no. 8, August, 1948, pp. 956–969, with the permission of the authors and the publisher.

where $K =$ constant greater than 1. Hence the figure of merit of any video amplifier can be expressed as

$$F_a = K\frac{g_m}{C_T}$$

Through the use of improved circuitry, shunt or series peaking for example, it was found that K could be made as high as 3. Actually, somewhat larger values can be obtained by using more complex circuits, but it is generally impractical to make it greater than about 4 because

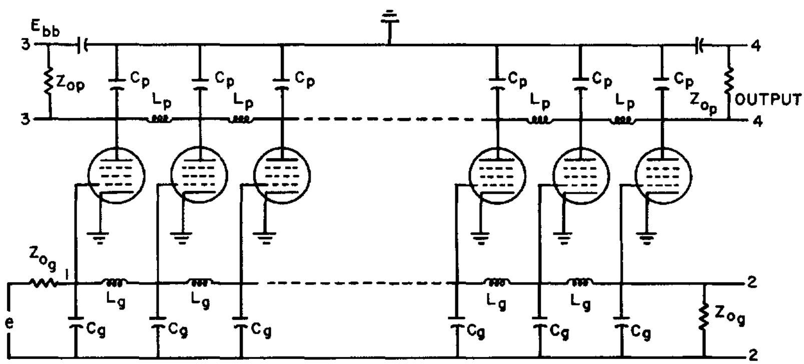


Fig. 5.28. One distributed amplifier stage of n sections.

the circuit complexity increases the wiring capacitance and C_T and tends to offset the improvement. It has been shown[4] that the maximum possible value for K is about 5.1.

From the discussion of cascading presented in the preceding sections it is clear that multistaging can be used to achieve greater gain-bandwidth products, but even here there is an eventual upper limit beyond which further improvement is not produced by cascading. The situation is not encouraging, for the design requirements may occasionally demand gains and bandwidths in excess of anything possible by the methods outlined so far. Clearly some other method of multistaging should be used.

In a sense, cascading is analogous to a series connection. Thus the thought occurs that paralleling tubes might result in an increase in the amplifier figure of merit. Unfortunately, this does not prove to be helpful directly, because paralleling tubes increases both the total g_m

[4] W. W. Hansen, "On the Maximum Gain-Bandwidth Product in Amplifiers," *J. Appl. Phys.*, vol. 16, September, 1945.

and C_T to the same extent, so that no net improvement in F_a is noted. However, this does suggest that a possibility might exist whereby the tube transconductances can be effectively paralleled and added while the capacitances are not. This very effect is successfully accomplished in the *distributed amplifier.*

A representative single *stage* distributed amplifier is shown in figure (5.28). A stage is composed of *n sections*, each section consisting of a tube and associated circuits. In the distributed amplifier the tube capacitances are made to serve as the shunt elements of two transmission lines, a *grid line* and a *plate line.* The tube capacitances involved are then specified as follows: C_p = plate-to-cathode capacitance of the tube; C_g = grid-to-cathode capacitance of the tube.

The characteristic impedances of the two transmission lines are

$$Z_{0_g} = \sqrt{\frac{L_g}{C_g}} \quad \text{and} \quad Z_{0_p} = \sqrt{\frac{L_p}{C_p}} \tag{5.159}$$

if the lines are composed of dissipationless elements. These characteristic impedances are independent of the number of sections in the line, and are therefore independent of the number of tubes. Hence the effect of connecting tubes this way is such that the tube capacitances do not add. We still have to prove that the tube transconductances add as if the tubes were connected in parallel.

The two transmission lines are designed to have equal velocities of propagation. Before describing the operation of the circuit, the following definitions are necessary:

Grid termination = impedance connected across 2–2.

Reverse termination = impedance connected across 3–3.

Plate termination = impedance connected across 4–4.

The input signal supplied to terminals 1–1 causes a wave to travel down the grid line. When it reaches the grid of the first tube the grid voltage causes a plate current to flow, in phase with the grid signal and in both directions along the plate line. The plate current component traveling to the right travels at the same speed as the grid signal voltage, so that both arrive at the second tube at the same time. Hence the plate current of the second tube will be in phase with that arriving from tube 1 and the component of plate current from tube 2, traveling to the right, will add to that from tube 1. This process is repeated at each tube, so that the output taken from terminals 4–4 is directly proportional to the number of tubes in the stage. Thus the *effective transconductance* of the distributed stage is proportional to the number of tubes in the stage and can be raised to any desired value without increasing the total effective shunt capacitance. As a result, the maximum obtainable bandwidth and amplifier figure of merit can be raised to almost any desired value.

The plate current components traveling to the left on the plate line are absorbed in the reverse termination without reflection. Therefore they do not affect the operation of the stage.

5.18. Cascading Distributed Amplifier Stages

After the necessary amplifier figure of merit has been achieved by adding sections to a distributed stage, the separate stages can be cascaded. The derivation of the resulting over-all gain is simplified if the following terms are defined:

Section = each tube with its sections of transmission line.

Stage = n sections.

Gain of a section = A_0.

Gain of a stage = A.

Gain of m cascaded stages = A_T.

Let E_g designate the grid voltage applied to the input of the grid line. The resulting plate currents in each tube will be

$$I = g_m E_g \tag{5.160}$$

if the grid line is lossless and if the r_p of the tube is very much larger than the characteristic impedance of the plate line. This current divides in the plate line; part travels toward the output and the remainder toward the reverse termination. Because the plate line is terminated in its characteristic impedance,

(1) Input impedance of the left part of plate line = Z_{0_p}.

(2) Input impedance of the right part of plate line = Z_{0_p}.

Consequently the plate current from each tube divides evenly, half traveling to the left and half to the right. The gain of the section is

$$A_0 = g_m Z_m = g_m \frac{Z_{0_p} Z_{0_p}}{Z_{0_p} + Z_{0_p}} = \frac{1}{2} g_m Z_{0_p} \tag{5.161}$$

If there are n sections in the stage, the current components and section gains add directly, making the *stage gain measured from grid to plate*,

$$A_{gp} = nA_0 = \frac{n}{2} g_m Z_{0_p} \tag{5.162}$$

In general, the characteristic impedances Z_{0_p} and Z_{0_g} of the plate and grid lines are not the same. Moreover, when stages are cascaded, it is necessary to connect the output from the plate line of one stage to the input of the grid line of the succeeding stage. Because the characteristic impedances of the two lines are unequal, it is necessary to insert an impedance transformer between stages. This device changes the impedance level from Z_{0_p} in the plate line to Z_{0_g} in the succeeding grid line; all lines are then properly terminated. However, this change in impedance level corresponds to a voltage change of $\sqrt{Z_{0_g}/Z_{0_p}}$. Hence the amplification of the stage *measured from grid to grid* is

$$A = A_{gp}\sqrt{\frac{Z_{0_g}}{Z_{0_p}}} \tag{5.163}$$

$$= \frac{n}{2} g_m Z_{0_p}\sqrt{\frac{Z_{0_g}}{Z_{0_p}}} = \frac{n}{2} g_m \sqrt{Z_{0_p} Z_{0_g}} \tag{5.164}$$

The equation for the stage gain given in equation (5.164) can be put into a more convenient form, because the velocities of propagation on the two lines are the same. This means that the cutoff frequencies of the lines must also be the same and given by

$$f_c = \frac{1}{\pi\sqrt{L_p C_p}} = \frac{1}{\pi\sqrt{L_g C_g}} \tag{5.165}$$

where f_c is the cutoff frequency for a low pass, constant-k filter. Solve each equation for the section inductance.

$$L_p = \frac{1}{\pi^2 f_c^2 C_p}; \qquad L_g = \frac{1}{\pi^2 f_c^2 C_g} \tag{5.166}$$

The characteristic impedances of the lossless lines are

$$Z_{0_p} = \sqrt{\frac{L_p}{C_p}}; \qquad Z_{0_g} = \sqrt{\frac{L_g}{C_g}} \tag{5.167}$$

Hence the product of the characteristic impedances is

$$Z_{0_p} Z_{0_g} = \sqrt{\frac{L_p}{C_p}}\sqrt{\frac{L_g}{C_g}} = \sqrt{\frac{L_p L_g}{C_p C_g}} \tag{5.168}$$

Substitute for L_p and L_g from equation (5.166).

$$Z_{0_p} Z_{0_g} = \frac{1}{\pi^2 f_c^2} \cdot \frac{1}{C_p C_g} \tag{5.169}$$

so that

$$\sqrt{Z_{0_p} Z_{0_g}} = \frac{1}{\pi f_c}\sqrt{\frac{1}{C_p C_g}} \tag{5.170}$$

Substitute this expression into the equation for the stage gain in (5.164).

$$A = \frac{ng_m}{2\pi f_c}\sqrt{\frac{1}{C_pC_g}} \tag{5.171}$$

or

$$A = \frac{n}{\omega_c}\cdot\frac{g_m}{\sqrt{C_pC_g}} \tag{5.172}$$

Now, if m such stages are cascaded, the over-all gain of the system is

$$A_T = A^m = \left(\frac{n}{\omega_c}\right)^m\left(\frac{g_m}{\sqrt{C_pC_g}}\right)^m \tag{5.173}$$

Take the mth root of both sides of this equation.

$$A_T^{1/m} = \frac{n}{\omega_c}\cdot\frac{g_m}{\sqrt{C_pC_g}} \tag{5.174}$$

and the equation can be solved for n to obtain

$$n = A_T^{1/m}\omega_c\frac{\sqrt{C_pC_g}}{g_m} \tag{5.175}$$

The total number of tubes in the cascade is denoted by N. Because there are m stages of n sections each, it is clear that $N = mn$. Hence, equation (5.175) can be written as

$$N = mA_T^{1/m}\omega_c\frac{\sqrt{C_pC_g}}{g_m} \tag{5.176}$$

The least number of tubes N required to produce a specified over-all gain A_T can now be computed. Differentiate equation (5.176) with respect to m and equate the result to zero.

$$\frac{\partial N}{\partial m} = \frac{\omega_c\sqrt{C_pC_g}}{g_m}\frac{\partial}{\partial m}mA_T^{1/m}$$

or

$$\frac{\partial}{\partial m}mA_T^{1/m} = 0 \tag{5.177}$$

Carry out the indicated differentiation.

$$A_T^{1/m} - \frac{A_T^{1/m}\ln A_T}{m} = 0 \tag{5.178}$$

Consequently,

$$mA_T^{1/m} = A_T^{1/m}(\ln A_T) \tag{5.179}$$

or

$$m = \ln A_T \tag{5.180}$$

which is the number of stages that will make the number of tubes a minimum. Solve equation (5.180) for the over-all gain by taking the antilogarithm of both sides.

$$A_T = \varepsilon^m \tag{5.181}$$

However, it will be recalled that the stage gain is

$$A = A_T^{1/m} \tag{5.182}$$

so that the stage gain required to make N a minimum is

$$A = (\varepsilon^m)^{1/m} = \varepsilon = 2.718 \tag{5.183}$$

Hence, if enough sections are added to the stage so that the stage gain is equal to ε, this will automatically result in a cascade that has the fewest tubes for a specified over-all gain.

The number of tubes actually required is easily computed, because

$$A = \varepsilon = \frac{n}{\omega_c} \cdot \frac{g_m}{\sqrt{C_pC_g}} \tag{5.184}$$

Therefore

$$n = \varepsilon\omega_c \frac{\sqrt{C_pC_g}}{g_m} = 17.1 f_c \frac{\sqrt{C_pC_g}}{g_m} \tag{5.185}$$

the number of tubes in each stage required to make the stage gain equal to ε. Now, because, $m = \ln A_T =$ number of stages required, the total number of tubes required for a certain over-all gain A_T is

$$N = mn = 17.1 f_c (\ln A_T) \frac{\sqrt{C_pC_g}}{g_m} \tag{5.186}$$

5.19. Frequency Response of a Distributed Amplifier

It was shown in equation (5.164) of the preceding section that the stage gain of a distributed amplifier is

$$A = \frac{ng_m}{2} \sqrt{Z_{0_p}Z_{0_g}} \tag{5.187}$$

where the Z_{0_p} and Z_{0_g} are the characteristic impedances of the plate and grid transmission lines. It is interesting to observe that the form of this equation is essentially the same as that used for all amplifier types. That is,

$$A = g_m' Z_m$$

where $g_m' =$ effective transconductance $= ng_m/2$; $Z_m =$ mutual impedance $= \sqrt{Z_{0_p}Z_{0_g}}$.

Each transmission line is actually made up of a cascade of low pass, constant-k, pi type filters as shown in figure (5.29). The characteristic impedance of a symmetrical pi section is given by the general equation

$$Z_0 = \frac{\sqrt{Z_1 Z_2}}{\sqrt{1 + Z_1/4Z_2}} \tag{5.188}$$

where, for the distributed amplifier,

$$Z_1 = j\omega L \quad \text{and} \quad L = L_p \text{ or } L_g \tag{5.189}$$

$$Z_2 = \frac{1}{j\omega C} \quad \text{and} \quad C = C_p \text{ or } C_g \tag{5.190}$$

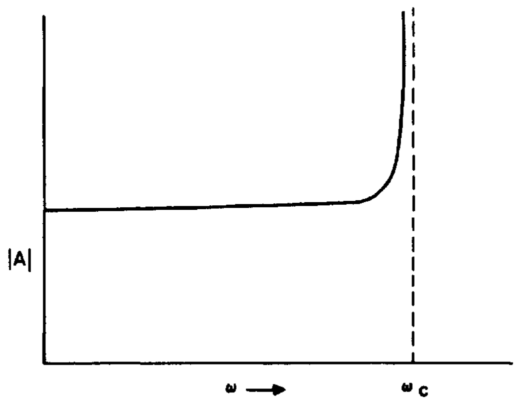


Fig. 5.29. Frequency response of a distributed amplifier.

Hence the characteristic impedance of the section is

$$Z_0 = \frac{\sqrt{L/C}}{\sqrt{1 - \omega^2 LC/4}} \tag{5.191}$$

or

$$Z_0 = \frac{R}{\sqrt{1 - \pi^2 f^2 LC}} \tag{5.192}$$

where

$$R = \sqrt{\frac{L}{C}} \quad \text{and} \quad \omega = 2\pi f \tag{5.193}$$

The cutoff frequency of the filter was previously defined as

$$f_c = \frac{1}{\pi\sqrt{LC}} \tag{5.194}$$

Hence, the equation for the characteristic impedance can be written

$$Z_0 = \frac{R}{\sqrt{1 - (f/f_c)^2}} \tag{5.195}$$

For the particular cases of the plate and grid lines of the distributed amplifier, this equation for the characteristic impedance becomes

$$Z_{0_p} = \frac{R_p}{\sqrt{1 - (f/f_c)^2}} \quad \text{where} \quad R_p = \sqrt{L_p/C_p} \tag{5.196}$$

$$Z_{0_g} = \frac{R_g}{\sqrt{1 - (f/f_c)^2}} \quad \text{where} \quad R_g = \sqrt{L_g/C_g} \tag{5.197}$$

Substitute these last two expressions into the gain equation given in (5.187).

$$A = \frac{ng_m}{2} \sqrt{R_p R_g} \frac{1}{\sqrt{1 - (f/f_c)^2}} \tag{5.198}$$

It is apparent that the reference gain of a distributed amplifier stage is

$$A_r = \frac{ng_m}{2} \sqrt{R_p R_g} \tag{5.199}$$

so that the general gain equation becomes

$$A = A_r \frac{1}{\sqrt{1 - (f/f_c)^2}} \tag{5.200}$$

If you examine this last equation you can see that the stage gain increases rapidly as the operating frequency of the amplifier approaches the cutoff frequency f_c. Theoretically, the amplification will be infinite at the cutoff point. This large increase in gain for frequencies near cutoff is generally undesirable. Methods of eliminating the peak are discussed at length in the basic reference cited.

Because the gain increases near cutoff instead of decreasing as all the cascaded amplifiers did, the usual definition of the upper cutoff frequency is not valid in this case. In other words, $\omega_H \neq \omega_c$. However, the cutoff characteristic is relatively sharp, and the bandwidth of the amplifier can be closely approximated by ω_c.

PROBLEMS

5.1. A low pass amplifier is to be designed to have an over-all upper cutoff frequency of 10 mcps and a reference gain of 20 volt logits. A cascade of identical resistance coupled amplifier stages is to be used with 6AK5 tubes, for which $g_m = 4500\ \mu$mhos and $C_T = 11.0\ \mu\mu$f. Is this amplifier feasible? What is the maximum obtainable upper cutoff frequency if the gain is 20 volt logits? For an over-all upper cutoff frequency of 12 mcps, what is the maximum possible gain?

5.2. An amplifier is to be designed to have a bandwidth of 3 mcps with a 60 mcps center frequency. Assume that $g_m = 5000\ \mu$mhos and $C_T = 11.0\ \mu\mu$f per stage. Design a maximally flat-staggered triple, computing the L and R required for each stage. Compute the over-all gain at 60 mcps.

5.3. Redesign the amplifier of problem (5.2) using the same tubes, but a synchronous connection of three stages. Compute the L and R for each stage and the over-all gain at 60 mcps. Compare the gain with that of the staggered triple.

5.4. Design an exact flat-staggered triple to have cutoff frequencies of 5 mcps and 20 mcps. Find the resonant frequency and Q of each stage.

5.5. Design an equal ripple, staggered triple to have a bandwidth of 1.5 with a center frequency of 30 mcps; the amount of ripple is to be 0.25 volt logit. Here 6SJ7 tubes are to be used, so that $F_a = 110 \times 10^6$. Compute the reasonant frequency and Q of each stage. Compute the over-all gain at band center and the gain that would be obtained at 30 mcps from a maximally flat-staggered triple.

5.6. An audio preamplifier is to be designed for use between a microphone and a class A power amplifier. The over-all response of the amplifier from microphone to the grid of the power amplifier is to be flat, ± 1.5 volt logits, from 20 c to 20 kc. The open circuit output voltage from the microphone is 2-mv, rms, and it has an internal impedance of 100,000 ohms. A 25-v peak-to-peak grid signal is required to drive the power amplifier. Assume standard resistance coupling throughout and interstage wiring capacitances of 15 $\mu\mu$f. Best results are achieved if the cutoff frequencies of all coupling circuits are the same. Assume C_{in} of the power amplifier to be 50 $\mu\mu$f.

An infinite number of designs could be obtained. To restrict the problem, assume that the first tube after the microphone is a 6SJ7 and all other tubes are 6J5's. Design the amplifier.

5.7. A radar beacon receiver is to be designed to have an over-all bandwidth of 3.5 mcps measured from the mixer output to the video output. The over-all system gain is to be 55 volt logits. For particular economic reasons, 6AC7 tubes are to be used throughout, so that $g_m = 9000\ \mu$mhos, $C_T = 23\ \mu\mu$f. The voltage transfer function of the video detector may be taken to be

0.8 and it is decided that two video stages will be used. The detector may be treated as a third video stage. Best operation results if the bandwidths of all stages are the same, so that the upper cutoff frequency of the video stages should be equal to the half bandwidth of the tuned stages. Determine the total number of stages required if synchronous intermediate frequency stages and resistance coupled video stages are used. What are the design requirements for the tuned (IF) section, the detector-video cascade? What would be the effect of using only one video stage and increasing the tuned cascade by one tube?

5.8. Design a maximally flat double tuned amplifier to have a bandwidth of 3 mcps centered about 60 mcps. Compute the mid-band gain. Determine L_1, L_2, k, R_1, and R_2, assuming that $g_m = 4500$ μmhos, $C_1 = 5$ $\mu\mu$f, $C_2 = 6$ $\mu\mu$f. If a double tuned pi section is desired, are the coil inductances reasonable? Why might a pi section be preferred to a transformer?

5.9. Redesign the amplifier of problem (5.8), but overcouple it so that an equal ripple response is obtained with 0.2 volt logits of ripple.

Chapter 6

TRANSIENT RESPONSE OF MULTISTAGE AMPLIFIERS

When the specifications for the design of a new amplifier are drawn up, the details of the amplifier as a separate unit are not considered. That is, the amplifier, regardless of type, or number of stages involved, is treated as a single unit in the over-all block diagram of the system of which the amplifier is a component part. The performance characteristics of the amplifier are then formulated so that it fulfills its mission in the system. The over-all performance properties of the amplifier are then specified and the circuit designer must answer questions of the following types:

(1) How many stages should be used?

(2) What kind of stages should be used?

(3) What tube is best suited to the amplifier?

(4) What are the gain, rise time, overshoot, and sag requirements of each individual stage in the amplifier?

It seems clear that these questions cannot be answered satisfactorily until the over-all performance of the multistage amplifier can be specified in terms of the behavior of the individual stages; or conversely, when the characteristics of the individual stages can be expressed in terms of the over-all performance properties desired, the individual stages can be designed.

The purpose of this chapter is to determine the effects of multistaging upon the over-all transient response of amplifiers. Certain rules of behavior will be established to express the characteristics of multistage amplifiers in terms of the corresponding characteristics of the individual stages.

The chapter has three main subdivisions, as follows:

(1) Transient response of cascaded video amplifiers.

(2) Envelope response of band pass amplifiers.

(3) Transient response of distributed amplifiers.

Rules of behavior are quickly formulated for the characteristics of cascaded low pass amplifiers. The *band pass, low pass* analogy is then used to describe the behavior of band pass amplifiers in terms of these same rules. The distributed amplifier is treated as a special case.

6.1. Composition of Rise Times, No Overshoot

The transient response characteristics of various single stage, low pass amplifiers were evaluated and discussed in chapter 4. It was found that resistance coupled and shunt peaked amplifiers, with m less than 0.25, did not overshoot. These are the most common amplifiers with *monotonic* (nonovershooting) transient edge response characteristics. Further study of the subject eventually proved that amplifiers do not overshoot unless their functions have conjugate poles in the s plane. While this rule was formulated for single stages, it can be generalized to include *any* amplifier regardless of the number of stages involved.

Attention was directed to a paper by Elmore at the beginning of chapter 4. According to Elmore, any amplifier, regardless of its type or the number of stages, has a gain function that can be expressed in general terms as

$$A(s) = A_r \frac{1 + a_1 s + a_2 s^2 + \dots}{1 + b_1 s + b_2 s^2 + \dots} \tag{6.1}$$

If the amplifier does not overshoot, the rise time T_R can be computed from Elmore's formula

$$T_R = \sqrt{2\pi[b_1^2 - a_1^2 + 2(a_2 - b_2)]} \tag{6.2}$$

This last expression will be used to determine how the rise time of a cascade depends upon the rise times of the individual stages, assuming that none of the stages overshoots.

In chapter 4 it was shown that the transient edge response characteristics of a single stage resistance coupled amplifier were

$$A_1(s) = -A_r \frac{1}{1 + (1/\omega_2)s} \tag{6.3}$$

or, in an alternative form, this could be written

$$A_1(s) = -A_r \frac{1}{1 + c_1 s} \tag{6.4}$$

where $c_1 = 1/\omega_2$.

By analogy to equation (6.1), the terms in the rise time formula for this gain function are

$$\left.\begin{aligned} a_1 &= 0; \qquad a_2 = 0 \\ b_1 &= c_1; \qquad b_2 = 0 \end{aligned}\right\} \tag{6.5}$$

Consequently, Elmore's rise time for a single stage is

$$T_R = \sqrt{2\pi c_1^2} = \sqrt{2\pi}\, c_1 \tag{6.6}$$

Suppose that two other resistance coupled amplifiers are available with gain functions

$$A_2(s) = -A_r \frac{1}{1 + c_2 s} \tag{6.7}$$

$$A_3(s) = -A_r \frac{1}{1 + c_3 s} \tag{6.8}$$

The rise times of these two stages are

$$T_{R_2} = \sqrt{2\pi} c_2 \tag{6.9}$$

$$T_{R_3} = \sqrt{2\pi} c_3 \tag{6.10}$$

Now suppose that all three of these separate stages are connected in cascade. As a result, the over-all gain function is

$$A(s) = A_1(s) A_2(s) A_3(s) \tag{6.11}$$

$$= -A_r^3 \frac{1}{(1 + c_1 s)(1 + c_2 s)(1 + c_3 s)} \tag{6.12}$$

Multiply the factors in the denominator and collect coefficients of like powers of s. The resulting equation is

$$A(s) = -A_r^3 \frac{1}{1 + (c_1 + c_2 + c_3)s + (c_1 c_2 + c_1 c_3 + c_2 c_3)s^2 + c_1 c_2 c_3 s^3} \tag{6.13}$$

By analogy to equation (6.1),

$$\left.\begin{aligned} a_1 &= 0; \qquad b_1 = c_1 + c_2 + c_3 \\ a_2 &= 0; \qquad b_2 = c_1 c_2 + c_1 c_3 + c_2 c_3 \end{aligned}\right\} \tag{6.14}$$

Therefore the over-all rise time of the cascade is

$$T_R = \sqrt{2\pi[(c_1 + c_2 + c_3)^2 - 2(c_1 c_2 + c_1 c_3 + c_2 c_3)]} \tag{6.15}$$

Carry out all the indicated algebraic steps.

$$T_R = \sqrt{2\pi(c_1^2 + c_2^2 + c_3^2)} \tag{6.16}$$

But

$$T_{R_1}^2 = 2\pi c_1^2 \tag{6.17}$$

$$T_{R_2}^2 = 2\pi c_2^2 \tag{6.18}$$

$$T_{R_3}^2 = 2\pi c_2^3 \tag{6.19}$$

so that the over-all rise time of the cascade is

$$T_R = \sqrt{T_{R_1}^2 + T_{R_2}^2 + T_{R_3}^2 + \ldots} \tag{6.20}$$

Thus the rise time of a cascade of nonovershooting amplifiers is equal to the square root of the sum of the squares of the rise times of the individual stages.

An examination of the coefficients involved in the rise time formula will show that the same results would be obtained for any number of stages. That is,

$$T_R = \sqrt{T_{R_1}^2 + T_{R_2}^2 + T_{R_3}^2 + \ldots + T_{R_n}^2} \qquad (6.21)$$

Equation (6.21) is true for *any* amplifier having a monotonic transient response as long as Elmore's definition of rise time is used. It is also approximately true for the 10–90% rise time. The approximation is a good one because the error seldom exceeds about 10% even when only two stages are involved.[1]

If n identical stages are connected in cascade, the individual stage rise times are all equal and specified by T_{R_s}. The over-all rise time is then

$$T_R = T_{R_s}\sqrt{n} \qquad (6.22)$$

6.2. Composition of Rise Times, Overshoot Present

In the analysis of single stage video amplifiers it was found that the rise time could be reduced by allowing the amplifier to overshoot slightly. For example, some curves were given in figure (4.23) for the shunt peaked amplifier, showing that the rise time could be reduced from $2.2RC_T$ to about $1.3RC_T$, by changing m from 0 to 0.38, if the amplifier is allowed to overshoot by about 2%. The increase in amplifier speed is appreciable. Of course, further reductions in rise time could be achieved by permitting larger overshoots.

There are many applications in which overshoots in excess of 2% would be excessive. Because the technique of introducing overshoot to decrease the rise time is useful, it is important to determine what happens to the over-all rise time and overshoot when n stages of small overshoot are connected in cascade.

Unfortunately, Elmore's rise time formula does not apply to amplifiers that overshoot. The only way to determine the effects of cascading is to compute the transient response characteristics of a number of cascaded amplifiers, with the overshoot as a parameter. When the results are plotted as a function of time, it will be possible to determine the over-all rise time and overshoot for these cases. Fortunately, this

[1] See Valley and Wallman, *Vacuum Tube Amplifiers*, vol. 18, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, p. 78.

computation has been done by Bedford and Fredendall.[2] According to their results, as long as the overshoot of the individual stages in the cascade is equal to or less than 2%, the following statements are true:

(1) The over-all rise time is given approximately by equation (6.21).

(2) The over-all overshoot remains nearly constant, increasing slightly if at all.

It is found that the overshoot increases rapidly if the overshoot per stage exceeds approximately 2%. Computed results lead to the following rules[3] for amplifiers in this category:

(1) The over-all overshoot of a cascade of identical amplifiers having overshoots in excess of 2% increases approximately as $\sqrt{n}$, where $n =$ number of stages.

(2) The over-all rise time increases much less rapidly than $\sqrt{n}$.

6.3. Condition for Minimum Over-all Rise Time

It was shown in section (5.6) of the preceding chapter that a cascade of identical amplifiers, either resistance coupled or single tuned, had the maximum over-all bandwidth obtainable with a specified gain when the stage gain was equal to $\varepsilon^{1/2}$. Because the bandwidth and rise time are governed by the same physical factors, it is reasonable to infer that minimum rise time will be produced by the same conditions that yield maximum bandwidth. The validity of this inference will be proven in the following derivation.

The amplifier figure of merit used in comparing the transient edge response characteristics of video amplifiers was defined in chapter 4 as the gain/rise time ratio. That is,

$$f_a = \frac{\text{reference gain}}{\text{rise time}} = \frac{A_r}{T_{R_s}} \tag{6.23}$$

If n identical stages are connected in cascade, the over-all reference gain is

$$A_T = A_r^n \tag{6.24}$$

Or, conversely, the stage gain can be expressed in terms of the over-all gain as

$$A_r = A_T^{1/n} \tag{6.25}$$

[2] A. V. Bedford and G. L. Fredendall, "Transient Response of Multi-Stage Video Amplifiers," *Proc. IRE*, vol. 27, April, 1939, pp. 277–284.

[3] Valley and Wallman, *op. cit.*, p. 78. Of further interest is an outstanding paper by H. E. Kollman, R. E. Spencer, and C. P. Singer, "Transient Response," *Proc. IRE*, vol. 33, March, 1945, pp. 169–195.

If it is assumed that the amplifier stages are free from overshoot, or do not overshoot by more than 2%, the over-all rise time is

$$T_R = T_{R_s} n^{1/2} \tag{6.26}$$

Consequently, the figure of merit for one stage in the cascade becomes

$$f_a = \frac{A_r}{T_{R_s}} = \frac{A_T^{1/n} n^{1/2}}{T_R} \tag{6.27}$$

The over-all rise time can now be expressed in terms of the over-all gain A_T, number of stages n, and amplifier figure of merit f_a, as

$$T_R = \frac{A_T^{1/n} n^{1/2}}{f_a} \tag{6.28}$$

Equation (6.28) shows how the over-all rise time depends upon the over-all gain, number of stages, and figure of merit. Both the gain and figure of merit are independent of n. Hence the rise time can be minimized with respect to the number of stages by standard calculus methods; differentiate T_R with respect to n, set the result equal to zero and solve for n. Therefore

$$\frac{\partial T_R}{\partial n} = A_T^{1/n} \frac{\partial}{\partial n}(n^{1/2}) + n^{1/2} \frac{\partial}{\partial n}(A_T^{1/n}) = 0 \tag{6.29}$$

As a result,

$$n = 2 \ln A_T \tag{6.30}$$

This is the number of stages that will make T_R a minimum for a specified over-all gain.

Consequently, the over-all gain of the cascade is

$$A_T = \varepsilon^{n/2} \tag{6.31}$$

and the over-all rise time is

$$T_R = \frac{(n\varepsilon)^{1/2}}{f_a} \tag{6.32}$$

which is the minimum rise time for a specified A_T. The over-all gain can be expressed in terms of the stage gain as

$$A_T = A_r^n \tag{6.33}$$

Set this equal to equation (6.31).

$$A_r = \varepsilon^{1/2} = 1.65 \tag{6.34}$$

This is the same result as that predicted at the beginning of this section.

The use of equations (6.30) and (6.32) can be illustrated by a simple example. Suppose that a cascade of resistance coupled amplifiers, using

6AK5's, is to be built to have an over-all gain of 40 volt logits. The minimum possible rise time is desired.

The gain/rise time ratio for a resistance coupled amplifier is

$$f_a = \frac{1}{2.2} \cdot \frac{g_m}{C_T} \tag{6.35}$$

It is assumed that (g_m/C_T) for the 6AK5 stages is 400×10^6.

Hence
$$f_a = \frac{400 \times 10^6}{2.2} = 182 \times 10^6$$

The over-all gain is specified as 40 volt logits. This corresponds to a voltage ratio of 10^4. According to equation (6.30), the number of stages required for minimum over-all rise time is

$$n = 2 \ln A_T = 2 \ln 10^4 = 8 \ln 10 \doteq 18 \tag{6.36}$$

Then, substituting into equation (6.32) for the rise time,

$$T_R\,(\text{min}) \doteq 0.0384 \quad \mu\text{sec}$$

While the rise time is desirably small, the necessity of using 18 stages may prove impractical. Assuming this to be the case, consider the possibility of using 9 stages. By using the formulas given in section (6.1), the resulting rise time is found to be 0.0459 μsec. The rise time is increased by about 20% while the number of tubes is cut in half and the over-all gain remains the same. This might be a better solution.

6.4. Flat Top Response of Cascaded Video Amplifiers

When all degenerative effects are neglected and the amplifier is assumed to be uncompensated for sag, it was shown in chapter 4 that all the video amplifiers discussed had the same flat top response. The flat top response functions were all of the form

$$A(s) = -A_r \frac{s}{s + \omega_1} \tag{6.37}$$

It was shown that the ω_1 factor was determined primarily by the coupling circuit between stages, mainly by R_g and C_c, and to a lesser extent by r_p and R_L.

If n identical amplifiers are connected in cascade, the overall gain equation is

$$A_T(s) = \pm A_r^n \left(\frac{s}{s + \omega_1}\right)^n \tag{6.38}$$

Because the over-all gain is the ratio of the output voltage to the input voltage,

$$A_T(s) = \frac{E_o(s)}{E_i(s)} \tag{6.39}$$

and the output voltage from the cascade is

$$E_o(s) = E_i(s)A_T(s) \tag{6.40}$$

If the input voltage as a function of time is assumed to be a unit step function,

$$e_i(t) = 1 \quad \text{for all} \quad t \geqq 0 \tag{6.41}$$

Consequently,

$$E_i(s) = \frac{1}{s} \tag{6.42}$$

Therefore the output voltage is

$$E_o(s) = \frac{1}{s} A_T(s) \tag{6.43}$$

or

$$E_o(s) = \pm A_r^n \frac{s^{n-1}}{(s + \omega_1)^n} \tag{6.44}$$

The output voltage as a function of time is the inverse Laplace transform of $E_o(s)$. This can be computed from a partial fraction expansion of equation (6.44) or by reference to published tables[4] of function-transform pairs. Regardless of the method, the result may be expressed as

$$e_o(t) = \pm A_r^n \varepsilon^{-\omega_1 t} \sum_{k=0}^{n-1} \frac{(n-1)!(-\omega_1)^k}{(n-1-k)!(k!)^2} t^k \tag{6.45}$$

This general equation is easily worked out for a few special cases by assuming various values for n. For example:

when $n = 1$,

$$e_o(t) = -A_r \varepsilon^{-\omega_1 t} \tag{6.46}$$

when $n = 2$,

$$e_o(t) = A_r^2 \varepsilon^{-\omega_1 t}(1 - \omega_1 t) \tag{6.47}$$

when $n = 3$,

$$e_o(t) = -A_r^3 \varepsilon^{-\omega_1 t}\left(1 - 2\omega_1 t + \frac{\omega_1^2 t^2}{4}\right) \tag{6.48}$$

[4] See Gardner and Barnes, *Transients in Linear Systems*, John Wiley & Sons, Inc. New York, 1942, p. 346, pair no. 2.1362.

when $n = 4$,

$$e_o(t) = A_r^4 \varepsilon^{-\omega_1 t}\left(1 - 3\omega_1 t + \frac{3\omega_1^2 t^2}{2} - \frac{2\omega_1^3 t^3}{6}\right) \tag{6.49}$$

and so on for any value of n selected. The initial slopes of the curves are needed to evaluate the sag. So differentiate $e_o(t)$ with respect to t and let $t = 0$. As a result,

when $n = 1$,

$$\text{slope} = -\omega_1 A_r \tag{6.50}$$

when $n = 2$,

$$\text{slope} = -2\omega_1 A_r^2 \tag{6.51}$$

when $n = 3$,

$$\text{slope} = -3\omega_1 A_r^3 \tag{6.52}$$

when $n = 4$,

$$\text{slope} = -4\omega_1 A_r^4 \tag{6.53}$$

and so on. By induction it is clear that if there are n stages, the initial slope will be $-n\omega_1 A_r^n$. Consequently, as long as the reciprocal of the total initial slope is small compared with the duration of the pulse, the over-all fractional sag is given by the following equations:

(1) Identical stages in cascade

$$\text{total fractional sag} = n\omega_1 T \tag{6.54}$$

where T = pulse duration.

(2) Nonidentical stages in cascade

total fractional sag = arithmetic sum of stage sags

Although the proof of these two rules has been carried through for the comparatively simple case of sag caused by the coupling circuit only, the rule applies equally well to all amplifiers with small sags, regardless of the cause of the sag.

6.5. Summary, Rules of Video Amplifier Behavior

The transient response characteristics of multistage video amplifiers have been discussed in the preceding sections, and the behavior of the cascade has been formulated in terms of the characteristics of the individual stages. For the sake of convenience and as a formal summary of the results obtained up to this point, the following rules of behavior are stated:

Rule 1. For video amplifiers that do not overshoot, the over-all rise time of an n stage cascade is closely approximated by

$$T_R = \sqrt{T_{R_1}^2 + T_{R_2}^2 + T_{R_3}^2 + \ldots + T_{R_n}^2} \tag{6.55}$$

Rule 2. The over-all rise time of a cascade of video amplifiers that have individual stage overshoots of 2% or less is given by equation (6.55).

Rule 3. If n video stages with overshoots in excess of 2% are cascaded, then (a) the overshoot increases approximately as $n^{1/2}$; (b) the rise time increases less rapidly than $n^{1/2}$.

Rule 4. If the cascade does not overshoot, the minimum over-all rise time is obtained when the gain of each stage is 1.65.

Rule 5. The total sag in the output of a video cascade is approximately equal to the arithmetic sum of the sags of each stage when these sags are small.

A similar set of rules will be found elsewhere[5].

6.6. The Band Pass, Low Pass Analogy

In many electronic systems used for pulse transmission, both low pass and band pass amplification are required. That is, it may be necessary to amplify high frequency signals that have a pulse *envelope* such as that shown in figure (6.1). Therefore, pulse transmission systems often

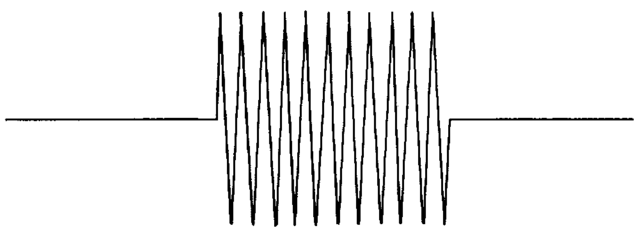


Fig. 6.1. Rectangular high frequency voltage pulse.

involve cascades of both low pass and band pass amplifiers. A typical system might appear as shown in figure (6.2). In such cases the transient response characteristics of band pass amplifiers are of interest.

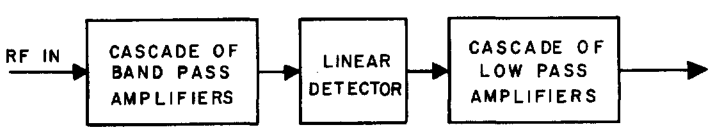


Fig. 6.2. Some components in a pulse transmission system.

Although there are notable exceptions, the response of the band pass amplifier to the *envelope* of the high frequency pulse is the matter of

[5] Valley and Wallman, *op. cit.*, pp. 77–78, 86.

most concern. This will be called the *envelope response* of the amplifier. In other words, the performance of a band pass amplifier used for pulse transmission is most frequently evaluated by determining the faithfulness with which the envelope of the input pulse is reproduced in the output. A possible response to the high frequency pulse of figure (6.1) is shown in figure (6.3). Note that a finite rise time and overshoot are

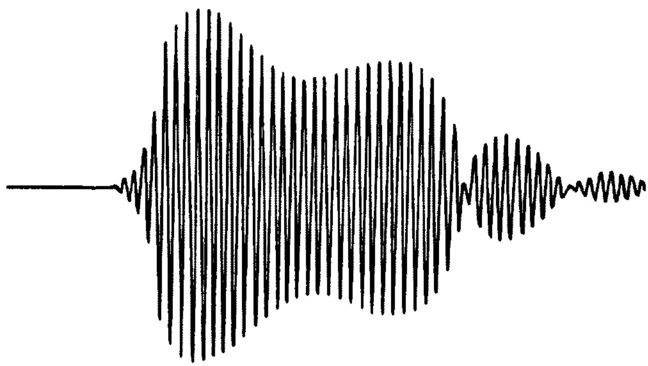


Fig. 6.3. Possible transient response of a band pass amplifier driven by a rectangular high frequency pulse.

indicated, both at the beginning and the end of the pulse. This figure also shows that the effects produced by overshoot on the leading and trailing edges of the pulse are different.

Assume that the envelope response characteristics of band pass amplifiers are to be determined. Because the envelope of a high frequency pulse is a video signal when considered separately, it should be possible to study the envelope response of a band pass amplifier in terms of the transient response of some video amplifier. In other words, it is suggested that a video amplifier could be specified to have the same transient response characteristics as the envelope response characteristics of the band pass amplifier. If such an equivalence can be established, the determination of the envelope response will be greatly simplified, because the transient response of the equivalent video amplifier can be evaluated fairly easily with the aid of the rules given in the preceding section.

A useful concept can be informally deduced by comparing the characteristics and gain functions of representative low pass and band pass amplifiers. For example, consider the simplest circuits in each category, a resistance coupled and a single tuned amplifier. The essential features of interest in these two amplifiers are shown in figure (6.4). If you examine the curves in this figure you can see that they are related. That is, the complete frequency response characteristic of a resistance coupled amplifier would be nearly identical to that of a single tuned amplifier if the *negative* frequency range could be obtained in a practical case. One important difference exists because the total

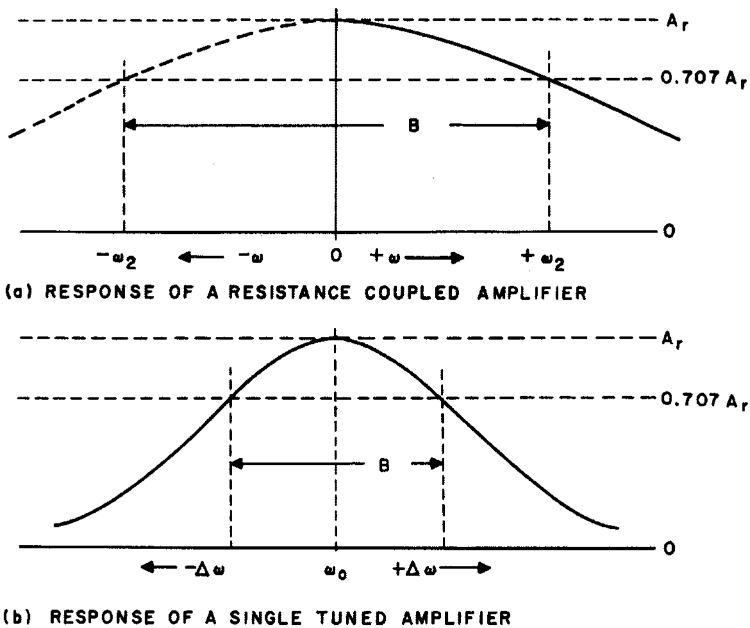


Fig. 6.4. Comparison of amplifier characteristics.

bandwidth B of this theoretical response curve for the resistance coupled amplifier will be equal to $2\omega_2$, and this is twice the bandwidth of the band pass amplifier. This correspondence might be specified the other way around. That is, the response characteristic of the single tuned amplifier would be the same as that of the resistance coupled amplifier if the center frequency is translated down the frequency scale from ω_0 to 0. The difference between the two amplifier response curves is simply that the resistance coupled amplifier will have twice as high an upper cutoff frequency.

Therefore a low pass or *video analogue* of the band pass amplifier can be constructed by making a video amplifier to have the same response as that of the band pass amplifier when the band pass characteristic is translated down the frequency scale from ω_c to zero. The transient response of this synthetic, proposed video analogue, which has an upper cutoff frequency of $B/2$, will be identically the same as the envelope response of the actual band pass amplifier. This is one special example of a relatively general principle called the *band pass, low pass analogy*[6].

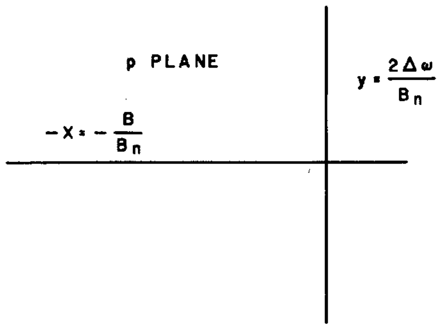


Fig. 6.5. Complex p plane; low pass analogue of the s plane.

This process of using a low pass analogue for a band pass circuit was used extensively in chapter 5 in developing the design procedure for stagger-tuned amplifiers. It was shown in section (5.7) that a new complex plane could be constructed, called the p plane, in which a high Q single tuned amplifier was characterized by a single pole. This pole was made to fall in this plane by translating the center frequency of the stage down the frequency scale from ω_c to zero by the use of a conformal transformation. Because a single pole is characteristic of a low pass amplifier, the p plane represents a convenient device whereby the properties of the low pass analogues of single tuned band pass amplifiers can be specified. The coordinates of the p plane are shown in figure (6.5) where $\Delta\omega =$ frequency difference between bandcenter and the resonant frequency of the stage; $B =$ stage bandwidth; $B_n =$ over-all bandwidth.

Because the pole location in the p plane defines the low pass analogue of a particular band pass amplifier, the transient response of this video analogue will be exactly the same as the envelope response of the band pass amplifier. If it is a high Q, single tuned amplifier stage, there will be one pole on the negative real axis of the p plane. Hence the transient

[6] V. D. Landon, "The Band-Pass Low-Pass Analogy," *Proc. IRE*, vol. 24, December, 1936, pp. 1582–1584; also P. R. Aigrain, B. R. Teare Jr., and E. M. Williams, "Generalized Theory of the Band-Pass Low-Pass Analogy," *Proc. IRE*, vol. 37, October, 1949, pp. 1152–1155.

response of the video analogue will have the form of a single exponential. The envelope response is the same. The rise time computed from the response of the video analogue will be the rise time of the envelope of the high frequency pulse.

6.7. Synchronous Single Tuning

In chapter 4, the poles of the gain functions of various amplifier types were computed and located in the complex s plane. The coordinates σ and ω of this plane have the dimensions of frequency or reciprocal time.

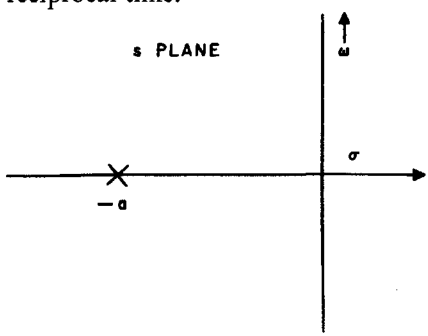


Fig. 6.6. s plane coordinates scaled off in units of 1/(time).

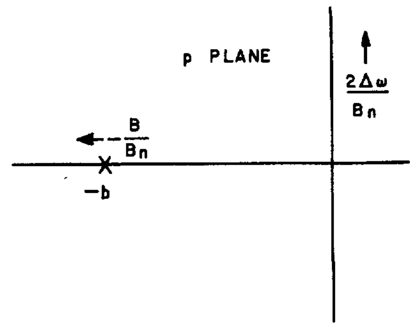


Fig. 6.7. p plane coordinates marked off in dimensionless numbers.

Consider a pole located in the s plane at some point such as $-a$ as shown in figure (6.6). This pole would come from a function of the general form

$$F(s) = K\frac{1}{s + a} \tag{6.56}$$

The inverse transform of this function is

$$f(t) = K\varepsilon^{-at} \tag{6.57}$$

Now, because a has the units of reciprocal time, the product at is a dimensionless number. This is a necessary condition because the exponent of ε must be dimensionless.

Now consider the p plane and while doing so, keep the properties of the s plane in mind for comparative purposes. In the p plane, the coordinates of the plane have been *normalized* by dividing through by $(B_n/2)$. Consequently, the coordinates are dimensionless numbers. As a result, the inverse Laplace transform of a function of p will involve normalized time of the form $B_n t/2$.

For example, assume that a pole is located at a point $-b$ in the p plane as shown in figure (6.7). Such a pole must have come from a function of the form

$$F(p) = K\frac{1}{p+b} \tag{6.58}$$

The inverse transform of this function is

$$f\left(\frac{B_n t}{2}\right) = K\varepsilon^{-bB_n t/2} \tag{6.59}$$

Because b is dimensionless, as is the product $B_n t/2$, the entire exponent is dimensionless as required. Everything is mathematically correct and analogous to the results obtained from the s plane.

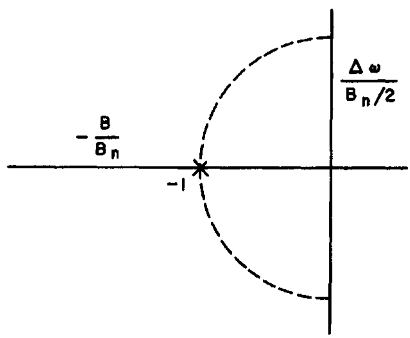


Fig. 6.8. Poles of a maximally flat amplifier, $n = 1$, in the p plane.

This information can now be applied to the determination of the envelope response of band pass amplifiers. It was shown in section (5.10) that the poles of the maximally flat-staggered amplifier all fall upon the periphery of a unit circle in the left half of the complex p plane. The angles of the poles, relative to the negative real axis, were specified. If a single stage single tuned amplifier is to be studied, it can be assumed that the stagger has a single stage with the pole located in the p plane as shown in figure (6.8). Hence $p = -1$. This pole arises from a function

$$A(p) = -A_r\frac{1}{p+1} \tag{6.60}$$

and this is the gain function of the low pass video analogue of the band pass amplifier. Thus the transient response computed from this function is the same as the envelope response of the single tuned amplifier.

The transient response of the video analogue is easily computed because

$$A(p) = \frac{E_o(p)}{E_i(p)} \tag{6.61}$$

or

$$E_o(p) = E_i(p)A(p) \tag{6.62}$$

If the envelope of the voltage input to the tuned amplifier is a unit, high frequency, step function, such as that shown in figure (6.1), the envelope input can be expressed as

$$e_i\left(\frac{B_n t}{2}\right) = 1 \quad \text{for all} \quad \frac{B_n t}{2} \geq 0 \tag{6.63}$$

Therefore the Laplace transform of the envelope input is

$$\mathcal{L}\left[e_i\left(\frac{B_n t}{2}\right)\right] = \frac{1}{p}$$

and the output voltage is then

$$E_o(p) = \frac{1}{p}A(p) = -A_r\frac{1}{p(p+1)} \tag{6.64}$$

Then the inverse transform of the envelope output, or the transient response of the video analogue, is

$$e_o\left(\frac{B_n t}{2}\right) = -A_r(1 - \varepsilon^{-B_n t/2}) \tag{6.65}$$

where B_n = overall bandwidth = stage bandwidth = B. Hence

$$e_o\left(\frac{Bt}{2}\right) = -A_r(1 - \varepsilon^{-Bt/2}) \tag{6.66}$$

A graph of this equation is given in figure (6.9). If the 10–90% rise time is computed from this curve, the result is

$$T_R = \frac{4.4}{B} = 4.4RC_T \tag{6.67}$$

The rise time of an actual resistance coupled video amplifier was previously shown to be

$$T_R = 2.2RC_T \tag{6.68}$$

It is clear that the rise time of the envelope of a band pass amplifier is exactly twice the rise time in a video amplifier having the same R and C_T.

Figure (6.9) shows that the envelope response of a single tuned amplifier does not overshoot. Consequently, when identical stages are cascaded to form a synchronous single tuned amplifier, the over-all rise time can be computed from *Rule* 1. That is, if $T_R =$ over-all envelope rise time, $T_{R_s} =$ envelope rise time of one stage $= 4.4RC_T$, $n =$ number of stages, then

$$T_R = T_{R_s} n^{1/2} \tag{6.69}$$

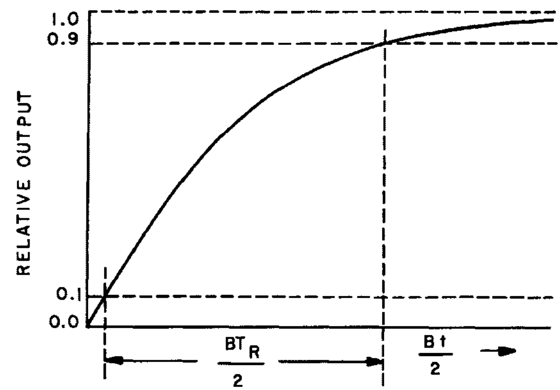


Fig. 6.9. Envelope response of a single tuned amplifier.

6.8. Staggered Pairs

The poles of a maximally flat-staggered pair are located in the complex p plane as shown in figure (6.10); this was proven in section

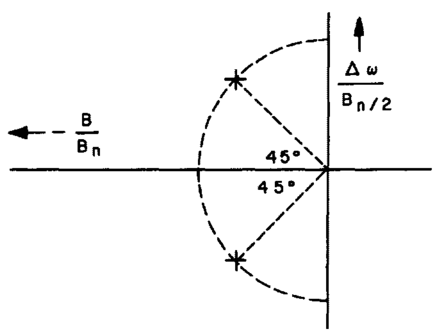


Fig. 6.10. Poles in the p plane for a maximally flat-staggered pair.

(5.10). The two poles are

$$p_1 = -0.71 + j0.71$$
$$p_2 = -0.71 - j0.71$$

or, in generalized notation

$$p_m = -x_m \pm jy_m \tag{6.70}$$

These two poles define the gain function of the low pass video analogue of the maximally flat-staggered pair. The transient response of this analogue is the same as the envelope response of the pair. Because the video analogue has conjugate poles in the p plane, its response will be oscillatory and will overshoot. Combining the factors involving the poles yields the gain function of the video analogue as follows:

$$A(p) = -A_r \frac{1}{(p + p_1)(p + p_2)} = -A_r \frac{1}{(p + x_m - jy_m)(p + x_m + jy_m)} \tag{6.71}$$

or

$$A(p) = -A_r \frac{1}{(p + x_m)^2 + y_m^2} \tag{6.72}$$

The method outlined in the preceding section can be used to determine the transient response of this video analogue. That is, if the envelope of the signal input to the stagger tuned amplifier is a unit step function, then

$$e_i\left(\frac{B_n t}{2}\right) = 1 \quad \text{for all} \quad \frac{B_n t}{2} \geq 0 \tag{6.73}$$

The Laplace transform of the input voltage is then

$$E_i(p) = \mathcal{L}\left[e_i\left(\frac{B_n t}{2}\right)\right] = \frac{1}{p} \tag{6.74}$$

Consequently, the output from the video analogue is

$$E_o(p) = E_i(p)A(p) = \frac{1}{p} A(p) \tag{6.75}$$

or

$$E_o(p) = -A_r \frac{1}{[(p + x_m)^2 + y_m^2]p} \tag{6.76}$$

The envelope response is the inverse transform of equation (6.76); this may be found by expanding the equation by the partial fraction expansion into a series of recognizable factors. The eventual result is

$$e_o\left(\frac{B_n t}{2}\right) = -A_r\left[\frac{1}{|p_m|^2} + \frac{1}{|p_m|\, y_m}\, \varepsilon^{-x_m B_n t/2} \sin\left(\frac{y_m B_n t}{2} - \phi\right)\right] \tag{6.77}$$

where

$$\phi = \tan^{-1} \frac{y_m}{-x_m} \tag{6.78}$$

$$|p_m|^2 = x_m^2 + y_m^2 \tag{6.79}$$

Although it will be desirable to compute the rise time eventually, for the present the amount of overshoot in the envelope response is of the greatest interest, because this will determine the manner in which the over-all rise time of a cascade of staggered pairs can be computed. Fortunately, for the case of the staggered pair, this is a relatively simple calculation. The overshoot is defined as the difference between the maximum value of the output voltage and its final value. Hence the first step is to determine this maximum value, and this is easily accomplished by standard methods from calculus. Simply differentiate $e_o(B_nt/2)$ with respect to $(B_nt/2)$, set the result equal to zero, and solve for $B_nt/2$. The numerical work can be simplified a little by defining a term t' to represent the $B_nt/2$ factor.

The output voltage from the video analogue is

$$e_o(t') = -A_r\left[\frac{1}{|p_m|^2} + \frac{1}{|p_m|\,y_m}\,\varepsilon^{-x_mt'}\sin(y_mt' - \phi)\right] \tag{6.80}$$

Therefore

$$\frac{de_o(t')}{dt'} = -A_r\left\{\frac{1}{|p_m|\,y_m}\,\varepsilon^{-x_mt'}[y_m\cos(y_mt' - \phi) - x_m\sin(y_mt' - \phi)]\right\} \tag{6.81}$$

The derivative will be equal to zero and the function will have its maximum or minimum value when

$$y_m\cos(y_mt' - \phi) = x_m\sin(y_mt' - \phi) \tag{6.82}$$

or, rearranging terms, when

$$\tan(y_mt' - \phi) = \frac{y_m}{x_m} \tag{6.83}$$

Expressed in an alternative form, this is

$$y_mt' = \tan^{-1}\frac{y_m}{x_m} + \phi \tag{6.84}$$

However, it was previously shown that

$$\phi = \tan^{-1}\frac{y_m}{-x_m} \tag{6.85}$$

Hence
$$y_mt' = \tan^{-1}\frac{y_m}{x_m} + \tan^{-1}\frac{y_m}{-x_m} = \pi \tag{6.86}$$

Therefore the time at which the response of the video analogue reaches its maximum value is

$$t' = \frac{\pi}{y_m} \tag{6.87}$$

It is a relatively easy matter now to evaluate the overshoot, because the final value of $e_o(t')$ is $A_r/|p_m|^2$. So, substituting equation (6.87) for t', the overshoot is easily computed from

$$\begin{aligned} \gamma &= \text{overshoot} \\ &= \left.\frac{\text{second term in equation (6.80)}}{\text{first term in equation (6.80)}}\right|_{t'=\pi/y_m} \end{aligned} \tag{6.88}$$

Hence
$$\%\gamma = \frac{p_m}{y_m}\varepsilon^{-x_m/y_m}\sin(\pi - \phi) \times 100\% \tag{6.89}$$

$$= \frac{p_m}{y_m}\varepsilon^{-x_m\pi/y_m}\sin\phi \times 100\% \tag{6.90}$$

The poles of the maximally flat-staggered pair were given at the beginning of this section as $p_m = -0.71 \pm j0.71$, so that

$$x_m = 0.71; \qquad y_m = 0.71; \qquad p_m = 1.0.$$

Therefore the overshoot evaluated from equation (6.90) is

$$\%\gamma = (\varepsilon^{-\pi})100\% = 4.3\% \tag{6.91}$$

Because the overshoot exceeds 2%, the rise times of a cascade of stages do not compose according to the simple relationship given in rules 1 and 2. Instead, rule 3 must be used. According to this rule, the overshoot of a cascade of maximally flat-staggered pairs will increase approximately as $m^{1/2}$, where m is the number of staggered pairs. That is,

$$\text{total overshoot} = 4.3m^{1/2}\ \% \tag{6.92}$$

For example, if 4 staggered pairs are cascaded, the total overshoot is approximately 8.6% according to this formula. This agrees quite closely with the 8.4% evaluated from the actual response.

If equation (6.80) for the envelope response is actually plotted, the 10–90% rise time figures out to be

$$T_R = \frac{4.34}{B_n} \tag{6.93}$$

Following rule 3 again, the overall rise time of a cascade of maximally flat-staggered pairs will increase considerably less rapidly than the overshoot. According to data computed by others[7], the over-all rise time increases by less than 15% when the number of staggered pairs in cascade is increased from 1 to 6.

The effects of over-staggering a staggered pair may be evaluated in a general way from the overshoot formula given in equation (6.90). An overstaggered pair adjusted to give an equal ripple gain function has p plane poles of

$$p_m = -0.71 \tanh \alpha \pm j0.71$$

where α is governed by the amount of ripple. Hence

$$\begin{aligned} x_m &= 0.71 \tanh \alpha \\ y_m &= 0.71 \\ p_m &= 0.71\,(1 + \tanh^2 \alpha)^{1/2} \\ \phi &= \tan^{-1}\left(\frac{1}{\tanh \alpha}\right) \end{aligned}$$

Consequently, substitution into the overshoot formula yields

$$\%\gamma = (1 + \tanh^2 \alpha)^{1/2}\, \varepsilon^{-\pi \tanh \alpha} \sin\left(\tan^{-1} \frac{1}{\tanh \alpha}\right) 100\% \qquad (6.94)$$

where

$$\alpha = \frac{1}{2} \sinh^{-1} (\epsilon^{-1/2}) \qquad (6.95)$$

and the ϵ in this last expression is governed by the ripple as given in chapter 5.

If $\alpha = 0.589$, the overshoot works out to be 18.9%. Thus, in general, the overshoot of an overstaggered amplifier will exceed that of a maximally flat-staggered amplifier.

6.9. Envelope Response, Concluding Remarks

While the discussion of the envelope response of band pass amplifiers has been confined to three relatively simple cases: (1) synchronous single tuning, (2) maximally flat-staggered pairs, (3) overstaggered, equal ripple pairs, most of the points of importance have been raised. For example, it was shown that a single tuned amplifier had a rise time of $4.4/B$ with zero overshoot, while a staggered pair had 4.3% overshoot with a rise time of $4.34/B_n$.

To compare the two cases, consider a staggered pair relative to two synchronous stages. For the two synchronous stages the overshoot

[7] Valley and Wallman, *op. cit.*, p. 281, table 7.1.

remains zero, but the rise time is increased to $6.2/B$, where B is the stage bandwidth. Because

$$B_n = B(2^{1/2} - 1)^{1/2} \doteq 0.64B \tag{6.96}$$

the rise time of the cascade of two stages is approximately $3.96/B_n$. This figure compares with $4.34/B_n$ for the staggered pair.

Now consider the over-all gain of the two amplifier cascades.

$$A_T \text{ (staggered pair)} = \left(\frac{F_a}{B_n}\right)^2 \tag{6.97}$$

$$A_T \text{ (synchronous pair)} = \left(\frac{0.64F_a}{B_n}\right)^2 \tag{6.98}$$

Therefore the gain/rise time ratios are

$$f_a \text{ (staggered pair)} = 0.231 \frac{F_a^2}{B_n} \tag{6.99}$$

$$f_a \text{ (synchronous pair)} = 0.104 \frac{F_a^2}{B_n} \tag{6.100}$$

The gain/rise time ratio of the maximally flat-staggered pair is more than twice that of a synchronously connected pair of the same figure of merit F_a and over-all bandwidth. Thus it would be possible to achieve a shorter rise time for a given gain by using a staggered pair if 4.3% overshoot can be tolerated.

The large overshoot computed for the overstaggered pair looks bad. However, by proper choice of the ripple it is possible to achieve a better gain with an overstaggered pair than with a staggered triple, assuming each amplifier to have the same rise time and overshoot.

It should be clear now that the selection of the best possible amplifier required to meet certain gain, rise time, and overshoot specifications can be an extremely difficult job. An extraordinarily high degree of technical ability is required to engineer the design properly. Certainly, the so-called "practical" man would be hopelessly lost and completely impractical.

The envelope response characteristics of staggered n-uples of any kind, of complexity greater than the staggered pair, are relatively difficult to evaluate. The difficulty is not one of theory, but simply the mechanical problem of computing the envelope response.

The flat top response has not been explored at all because the effects of the coupling capacitor are usually unimportant in band pass amplifiers. This is also generally true of degenerative effects, though there are some special problems in this connection.

6.10. Transient Response of a Distributed Amplifier

The general theory of the distributed amplifier and its steady state frequency response were covered in chapter 5. The equation for the voltage gain of a single distributed amplifier stage was derived and shown to be

$$A = \frac{ng_m}{2}(Z_{0_p}Z_{0_g})^{1/2} \tag{6.101}$$

This equation is perfectly general because the kind of frequency involved is left unspecified. The characteristic impedances are given by

$$Z_0 = \sqrt{\frac{Z_1Z_2}{1 + Z_1/4Z_2}} \tag{6.102}$$

The values for these impedances in the distributed amplifier are easily determined by looking back at figures (5.28) and (5.29) in the preceding chapter. After substitution of the appropriate values you can easily show that

$$Z_{0_p} = R_p \frac{\omega_c}{\sqrt{s^2 + \omega_c^2}} \tag{6.103}$$

$$Z_{0_g} = R_g \frac{\omega_c}{\sqrt{s^2 + \omega_c^2}} \tag{6.104}$$

where ω_c, R_p, and R_g were all defined in chapter 5.

Substitute these equations into (6.101). The resulting expression for the voltage gain is

$$A(s) = \frac{ng_m}{2}\sqrt{R_pR_g}\frac{\omega_c}{\sqrt{s^2 + \omega_c^2}} \tag{6.105}$$

or

$$A(s) = A_r \frac{\omega_c}{\sqrt{s^2 + \omega_c^2}} \tag{6.106}$$

The output voltage as a function of time must be computed before the rise time and overshoot of the amplifier can be determined. Because $E_o(s) = E_i(s)A(s)$, the response transform to a unit step input function will be

$$E_o(s) = \frac{1}{s}A(s) = A_r \frac{\omega_c}{s(s^2 + \omega_c^2)^{1/2}} \tag{6.107}$$

It is impossible to expand this by partial fractions because there are an infinite number of poles arising from the irrational fraction. It is simpler to use a theorem from the basic theory of the Laplace transform.

This theorem states that

$$\mathcal{L}^{-1}\left(\frac{1}{s}F(s)\right) = \int_0^t f(t)\,dt \tag{6.108}$$

where

$$f(t) = \mathcal{L}^{-1}[F(s)] \tag{6.109}$$

Therefore, in equation (6.107), let

$$F(s) = \frac{1}{(s^2 + \omega_c^2)^{1/2}} \tag{6.110}$$

The inverse transform of this function may be found in several tables of function-transform pairs.[8] The form is

$$f(t) = \mathcal{L}^{-1}[F(s)] = J_0(\omega_c t) \tag{6.111}$$

where $J_0(\omega_c t) =$ Bessel function of the first kind and order zero. Therefore the output voltage is

$$e_o(t) = \omega_c A_r \int_0^t J_0(\omega_c t)\,dt \tag{6.112}$$

The Bessel function can be expressed in terms of an infinite series as follows:[9]

$$J_0(\omega_c t) = 1 - \frac{(\omega_c t)^2}{2^2} + \frac{(\omega_c t)^4}{2^2\,4^2} - \frac{(\omega_c t)^6}{2^2\,4^2\,6^2} + \cdots \tag{6.113}$$

To evaluate the output voltage, integrate this equation term by term, multiply through by $\omega_c A_r$. The transient response is then

$$e_o(t) = A_r\left[\omega_c t - \frac{(\omega_c t)^3}{12} + \frac{(\omega_c t)^5}{320} - \frac{(\omega_c t)^7}{16{,}128} + \cdots\right] \tag{6.114}$$

To determine the overshoot, several things must be done. It is necessary to find the time at which the first maximum voltage is attained. This is not difficult, because we only need to differentiate the output voltage, set the result equal to zero, and solve for t. Because the derivative of the output voltage is

$$\frac{de_o(t)}{dt} = \omega_c A_r J_0(\omega_c t) \tag{6.115}$$

it is fairly simple to find the value of $\omega_c t$ that will make the derivative zero. The zeros of $J_0(\omega_c t)$ have been tabulated in several places.[10] From these we find that the first zero occurs when $\omega_c t = 2.405$. Substitute this into equation (6.114) for $e_0(t)$ and the result is

$$e_o(t)_{\max} = (1.47)A_r \tag{6.116}$$

[8] See for example, Gardner and Barnes, *op. cit.*, pp. 317–320, and also p. 352, pair no. 5.01.

[9] Ruel V. Churchill, *Fourier Series and Boundary Value Problems*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1941, p. 145.

[10] See, for example, *ibid.*, p. 157.

The evaluation of the overshoot requires that the final value of $e_o(t)$ be known. This can be evaluated as follows:

$$e_o(t) = \omega_c A_r \int_0^\infty J_0(\omega_c t)\, dt = \text{final value at } t = \infty$$

$$= A_r \int_0^\infty J_0(\omega_c t)\, d(\omega_c t) \tag{6.117}$$

This is a definite integral having a value of 1.[11] Therefore the overshoot is

$$\%\,\gamma = \frac{\text{max. value} - \text{final value}}{\text{final value}} \times 100\% = 47\%$$

This tremendous overshoot is exceedingly undesirable, and is reduced to practical limits in actual circuits by several different methods. The problem is closely tied into the rapid increase in gain near the cutoff frequency ω_c. Measures taken to remove this unfortunate steady state characteristic will also improve the overshoot, and readers interested in the practical circuitry are referred to footnote reference (3) of chapter 5.

It is interesting that this 47% overshoot is independent of the value of ω_c.

PROBLEMS

6.1. A synchronous single tuned amplifier operating at 30 mc is to be designed to provide an over-all mid-band gain of 35 volt logits. Type 6AK5 tubes are to be used with $g_m = 4500$ μmhos, and the individual total interstage capacitances may be taken to be 12 $\mu\mu$f. Compute the minimum possible envelope rise time. Compute the number of stages required for minimum envelope rise time.

6.2. The number of stages required for minimum rise time in problem (6.1) is excessive. On the basis of weight, size, heat dissipation problems, and so on, it is decided that no more than seven tubes can be used and that the gain must be 35 volt logits. Compute the possible envelope rise time.

6.3. The rise time of the amplifier in problem (6.2) proves to be too large. It is then decided that the envelope rise time cannot exceed 0.08 μsec. Calculate the resulting over-all mid-band gain and compare it with the specified value.

6.4. From the foregoing calculations it is clear that the required gain and rise time cannot be achieved through the use of only seven 6AK5 tubes, synchronously connected. You talk your supervisor into letting you use one

[11] See S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand Company, Inc., New York, 1943, p. 56, formula 7–27.

more stage even though this requires a redesign of the equipment container. Can the 35 volt logits and 0.08 μsec requirements be met now?

6.5. In the event that only seven tubes were permissible under any circumstances, how much of an increase in the tube g_m would be required to meet specifications? Assume that the interstage shunt capacitance is at an irreducible minimum of 12 $\mu\mu$f. How could you increase the g_m of the 6AK5 tubes? What dangers are involved in this and what precautions should be taken?

6.6. A high gain, fast video amplifier is required for use in a laboratory oscilloscope deflection system. Assume 6AK5 tubes are used with $g_m = 4500$ μmhos and $C_T = 11$ $\mu\mu$f. An over-all gain of 25 volt logits is required. Calculate the over-all rise time available, using a two-stage cascade of identical

(*a*) Resistance coupled amplifiers.
(*b*) Shunt peaked amplifiers with $m = 0.25$.
(*c*) Shunt peaked amplifiers with $m = 0.30$.
(*d*) Shunt peaked amplifiers with $m = 0.50$.
(*e*) Series peaked amplifiers with $K = 0.50$.

Compute or estimate the overshoot in these cases and make a critical evaluation of the five possibilities.

6.7. A radar receiver is to be designed to operate with an intermediate frequency of 60 mc. The following receiver requirements must be met: signal input to the receiver = 10 μv; signal output to CRT (max) = 10 v; range measurement to be accurate to ± 50 feet; 1 radar mile = 10.73 μsec. The voltage transfer function of the detector, E(video)/E(IF) is 0.8 and its rise time characteristics are the same as for a resistance coupled amplifier. Assume that 6AK5 tubes are used in the IF amplifier and a 6AC7 and 6AG7 in the video section. Synchronous single tuning is used in the IF and identical resistance coupled stages in the detector-video section. All stages are to have the same rise time. The following data apply:

Tube	g_m (μmhos)	C_T ($\mu\mu$f)
6AK5	4500	12
6AC7	8000	25
6AG7	7700	25

(a) Determine the number of stages required following the mixer.
(b) Break the over-all requirements into stage requirements.
(c) Comment upon the feasibility of the system.
(d) Is the design conservative or will the system be subject to tube variability effects?
(e) Indicate the direction of the changes required if the receiver sensitivity is to be increased or if greater range accuracy is desired.

Chapter 7

FEEDBACK CIRCUITS

In chapter 3 it was shown that the general equation for the voltage amplification of any class A vacuum tube amplifier was of the form

$$A = \pm g_m' Z_m$$

where g_m' = effective transconductance of the tube; Z_m = mutual impedance of the network in the equivalent plate circuit.

Both g_m' and Z_m depend upon the properties of the tube. This is unfortunate, because the tube constants such as r_p, g_m, and the interelectrode capacitances, vary from tube to tube, time to time, and are generally affected by changes in environment and polarizing potentials. As a result, the response characteristics of the amplifier change because of the influence of a large number of factors over which little control is exercised by the circuit designer. This obviously complicates circuit design. However, it has been found that some of these effects can be minimized by taking a small fraction of the amplifier output and feeding it back to the input. This is called *feedback*, and the general problem is briefly treated here. More extensive discussions may be found in any book on servomechanisms.

Stabilization of gain is not the only effect achieved through feedback. For example, in chapter 11 it is shown that nonlinear, or harmonic, distortion is produced in large signal circuits such as class A power amplifiers. The distortion results from the motion of the operating point of the tube over a nonlinear part of its characteristics. It will be shown that feedback can be used to reduce the amount of harmonic distortion.

It will also be shown that feedback can be used to reduce circuit noise, improve the frequency response characteristics, and fulfill other useful functions.

The foregoing are cases of deliberate feedback used to achieve some particular operational advantage. However, inadvertent feedback nearly always occurs in electronic circuits. It may occur through tube interelectrode capacitances, a common impedance in the power supply, or stray electric and magnetic fields. Therefore it is essential that the general characteristics and effects of feedback be understood, and that tools be provided for attacking specific problems analytically and experimentally.

7.1. General Gain Equation, Single Closed Loop

For any ordinary voltage amplifier, the over-all voltage gain of a cascade of stages is defined simply as

$$A_o = \frac{\text{amplifier output voltage}}{\text{amplifier input voltage}} = \frac{E_o}{E_i} \tag{7.1}$$

This ratio, or *voltage transfer function*, is usually frequency dependent and can be written in terms of frequency as

$$A_o(s) = \frac{E_o(s)}{E_i(s)} \tag{7.2}$$

The terminology is illustrated in figure (7.1a).

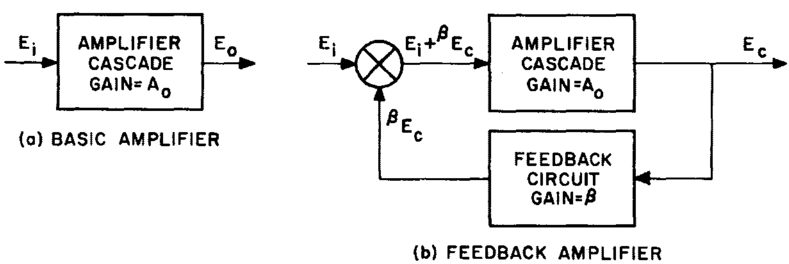


Fig. 7.1. Definitions of terms.

Now consider figure (7.1b). The same basic amplifier cascade is used, but the output voltage is connected back into the input circuit through the feedback network; this circuit has a voltage transfer function denoted by β. Therefore the voltage fed back into the input is βE_c.

The symbol used in figure (7.1b), consisting of a circle with an $\times$ through it, indicates an adding device. That is, the output from the *adder* is the sum of the arrow marked inputs. Hence the input to the amplifier of figure (7.1b) is $E_i + \beta E_c$.

The over-all amplification equation for the *feedback* amplifier of figure (7.1b) is

$$A_c = \frac{\text{system output voltage}}{\text{system input voltage}} = \frac{E_c}{E_i} \tag{7.3}$$

The new output voltage E_c is easily computed because $A_0 =$ gain of the basic amplifier.

or
$$A_o = \frac{E_c}{E_i + \beta E_c} \tag{7.4}$$

Solve this equation for E_c.

$$E_c = \frac{A_o}{1 - \beta A_o} E_i \tag{7.5}$$

As a result, the over-all amplification of the system with feedback is

$$A_c = \frac{A_o}{1 - \beta A_o} \tag{7.6}$$

where $A_o =$ voltage transfer function of the basic amplifier; $\beta =$ voltage transfer function of the feedback circuit; $\beta A_o =$ feedback factor. The equation can be expressed in terms of the complex frequency as

$$A_c(s) = \frac{A_o(s)}{1 - \beta(s)A_o(s)} \tag{7.7}$$

This is a basic and fundamental relationship in the field of electronic circuit theory.

A feedback system such as that shown in figure (7.1b) is called a *closed loop* system because there is a closed signal transmission path. Because there is only one feedback path, it is called a *single* closed loop system. Multiple loop systems are common.

A system that does not employ feedback is called an *open loop* system because there is no closed signal transmission path. For convenience in notation, all quantities characteristic of the open loop system will carry a subscript o; a subscript c will be used for all closed loop quantities.

From equation (7.7) it is clear that the over-all voltage transfer function $A_c(s)$ of a closed loop system is governed by the character of the open loop transfer function and the transfer function of the feedback circuit. Both β and A_o are usually dependent upon frequency, as indicated in equation (7.7) so that the over-all gain function is also frequency dependent.

As a result of the feedback, the mathematical form for the gain function of the feedback amplifier is different from that of the open loop amplifier. This means that the two amplifiers will have different poles and zeros in the s plane, but it is difficult to generalize about their new locations. Moreover, the scale factor, or gain, is usually altered. If the scale factor is reduced, negative feedback is in use. Conversely, an increase in scale factor usually results from positive feedback.

7.2. Effects of Feedback on Gain, Distortion, and Noise

In the preceding section we described some effects of positive and negative feedback without defining what we mean. These terms are defined as follows: positive feedback makes the feedback factor $\beta(s)A_o(s)$ positive while the factor is negative when the feedback is negative. From this you can see that the gain of a negative feedback

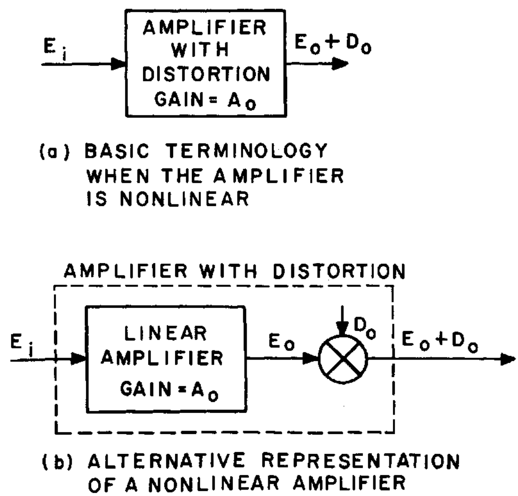


Fig. 7.2. Nonlinear distortion in amplifiers.

amplifier is less than the gain of the open loop amplifier regardless of the value of s. This is evident from the following equation:

$$A_c\,(\text{negative FB}) = \frac{A_o(s)}{1 + \beta(s)A_o(s)}$$

Conversely, positive feedback increases the gain of the amplifier as long as $\beta(s)A_o(s)$ is less than 1. When $\beta(s)A_o(s)$ is exactly $+1$ the gain is infinite and the amplifier becomes an *oscillator*. This circuit is covered in some detail later.

The effect of feedback on the frequency response of an amplifier can be assessed from the new pole and zero locations. Negative feedback generally increases the amplifier *bandwidth*, but this does not necessarily mean improvement at *both* ends of the response.

Suppose that an open loop amplifier introduces nonlinear distortion as shown schematically in figure (7.2a). The nonlinearity is indicated by showing the total amplifier output voltage as the sum of two components, the signal output voltage, E_o, and the distortion voltage D_o. An alternative representation of the circuit is shown in figure (7.2b), and this is more useful for present purposes.

Now consider the case when a single feedback loop is added to supply *negative* feedback. This condition is indicated schematically in figure (7.3). In this case the total distortion output D_c is the sum of two input

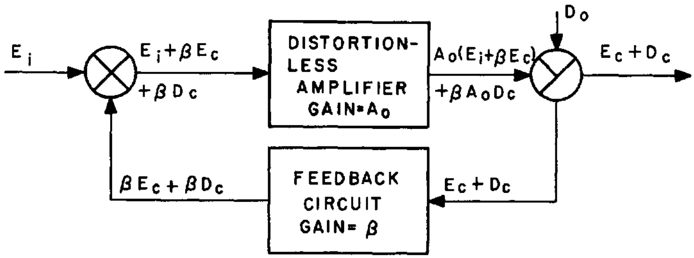


Fig. 7.3. Nonlinear distortion in a closed loop system.

terms: D_o = distortion introduced by the amplifier; $\beta A_o D_c$ = distortion fed back. Hence,

$$D_c = D_o + \beta A_o D_c$$

Or, rearranging terms,

$$D_c = \frac{D_o}{1 - \beta A_o} \tag{7.8}$$

Thus the nonlinear distortion is altered in exactly the same way as the over-all gain. As a result, negative feedback will reduce the nonlinear distortion, while positive feedback will cause it to increase.

Because the use of negative feedback reduces the gain by the same amount as it reduces the distortion, it would seem that no particular advantage is obtained. However, nonlinear distortion is usually important only in the latter stages of an amplifier cascade where the signal amplitudes are large and where the *power* output is mainly important. Negative feedback can be used in such stages to reduce nonlinear distortion, and the decreased gain can be made up in preceding stages. This permits the power output to be held constant while the nonlinear distortion is reduced.

Noise, like nonlinear distortion, is changed by the $1/(1 - \beta A_o)$ factor when feedback is applied. This is true regardless of the character of the noise. A typical case might be indicated schematically as shown in figure (7.4) where only noise voltage components are shown. Signal voltages have been omitted because the Principle of Superposition applies. The various terms appearing in figure (7.4) are defined as follows: $N =$ noise voltage input; $N_o = NA_o =$ noise voltage output, open loop case; $N_c =$ noise voltage output, closed loop case.

From figure (7.4) it is clear that

$$N_c = (N + \beta N_c) = NA_o + \beta A_o N_c$$

or

$$N_c = N_c + \beta A_o N_c$$

Therefore the closed loop noise voltage output is

$$N_c = \frac{N_o}{1 - \beta A_o} \tag{7.9}$$

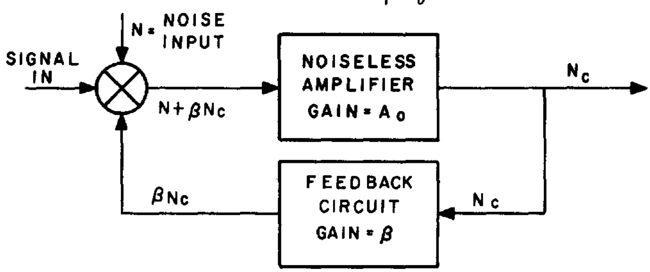


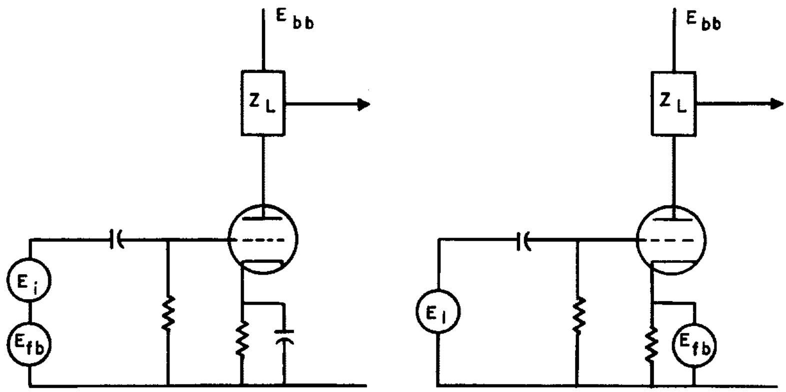
Fig. 7.4. Noise in a feedback amplifier.

7.3. Feedback Circuit Connections

The feedback voltage for the type of amplifiers being discussed here is always inserted in series with the input voltage. This feedback voltage E_{fb} can be introduced in series with the original signal input in the grid circuit or in the cathode circuit. Typical feedback circuit connections are shown in figure (7.5).

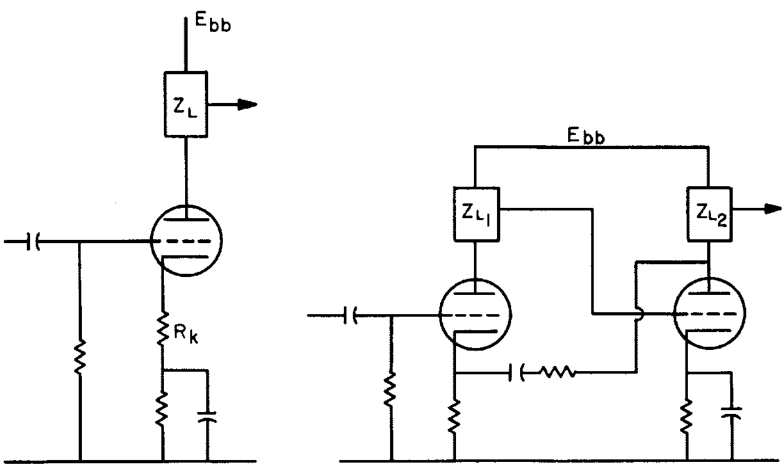
The feedback voltage E_{fb} shown in figure (7.5) may be developed at any point in an amplifier chain and fed back through some circuit so that the over-all voltage transfer function of the feedback loop is β. Nearly any kind of circuit can be used for this purpose. Various types of frequency selective circuits such as filters and bridges are commonly used.

If the voltage fed back is proportional to a current flowing through an impedance other than the load, the system is said to use *current controlled feedback*. If the feedback voltage is proportional to some signal voltage, the system uses *voltage controlled feedback*. Representative circuits are shown in figure (7.6).



(a) GRID FEEDBACK (b) CATHODE FEEDBACK

Fig. 7.5. Feedback connections.



(a) NEGATIVE CURRENT FEEDBACK IN THE CATHODE CIRCUIT; CATHODE DEGENERATION

(b) VOLTAGE FEEDBACK IN THE CATHODE CIRCUIT

Fig. 7.6. Representative feedback circuits.

There is another type of feedback, previously encountered in the discussion of the Miller effect, that has considerable practical importance. In this case, the plate and grid of the tube are connected together through some arbitrary admittance as shown in figure (7.7). It will be shown that this causes the input admittance of the tube to vary with frequency, thereby changing the loading of the driving stage and causing a corresponding change in the response of the system. This type of circuit is discussed in some detail in later sections of the chapter.

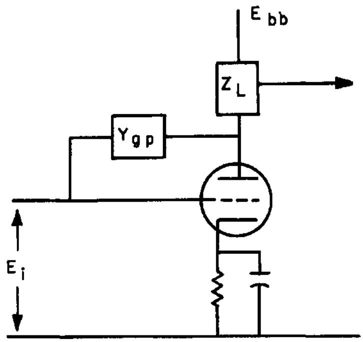


Fig. 7.7. Grid-to-plate feedback.

7.4. Cathode Degeneration Treated by Feedback Analysis

The most common type of current-controlled feedback is cathode degeneration. A typical circuit was given in figure (7.6a). The voltage gain equation of the amplifier with cathode degeneration was derived

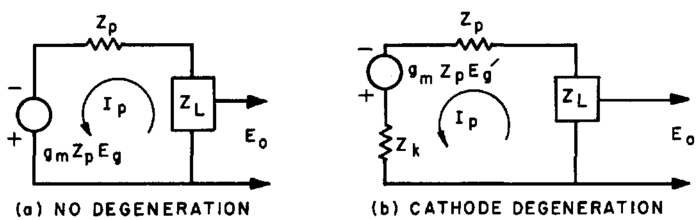


Fig. 7.8. Equivalent plate circuits.

by standard methods in chapter 3. The same equation will be developed in this section as an illustration of the use of the general feedback equation.

The class A equivalent circuits of a grounded cathode amplifier with and without cathode degeneration are shown in figure (7.8). The voltage source equivalent circuit is the more convenient for this derivation, and the signal input voltages in the two circuits are adjusted to give the same plate current I_p. This also produces the same output voltage E_o.

For the open loop circuit, the voltage loop equation is

$$g_m z_p E_g = I_p(Z_L + z_p) \tag{7.10}$$

while the loop equation in the degenerative case is

$$g_m z_p E_g' = I_p(Z_L + z_p + Z_k) \tag{7.11}$$

The grid voltage in the degenerative circuit is the difference between the signal input voltage E_i and the cathode voltage E_k. That is,

$$E_g' = E_i - E_k = E_i - I_p Z_k \tag{7.12}$$

Consequently, the loop equation for the feedback amplifier is

$$g_m z_p E_i = I_p(Z_L + z_p) + I_p Z_k(g_m z_p + 1) \tag{7.13}$$

However, according to equation (7.10),

$$I_p(Z_L + z_p) = g_m z_p E_g \tag{7.14}$$

so that equation (7.13) can be written

$$g_m z_p E_i = g_m z_p E_g + I_p Z_k(g_m z_p + 1) \tag{7.15}$$

or, in an alternative form,

$$g_m z_p(E_i - E_g) = I_p Z_k(g_m z_p + 1) \tag{7.16}$$

Consider the terms in this equation: E_i = signal input voltage with feedback; E_g = signal input without feedback. Both of these voltages produce the same output voltage. Clearly then, the difference between E_g and E_i is the feedback voltage, E_{fb}. That is,

$$E_{fb} = E_g - E_i = -I_p Z_k \frac{g_m z_p + 1}{g_m z_p} \tag{7.17}$$

It is especially interesting to note that the feedback voltage is *not* equal to the cathode voltage $I_p Z_k$.

If Z_m designates the mutual impedance of the amplifier plate circuit, the current flowing through this impedance is related to the plate current I_p as follows:

$$I = I_p \frac{z_p + Z_L}{z_p} \tag{7.18}$$

Therefore the output voltage is

$$E_o = -IZ_m = -I_p \frac{z_p + Z_L}{z_p} Z_m \tag{7.19}$$

The voltage transfer function of the feedback circuit is

$$\beta = \frac{E_{fb}}{E_o} = + \frac{Z_k}{Z_m} \cdot \frac{g_m z_p + 1}{g_m z_p} \cdot \frac{z_p}{z_p + Z_L} \tag{7.20}$$

This can be written in various forms as follows:

$$\beta = \frac{Z_k(g_m z_p + 1)}{(Z_L + z_p)g_m Z_m} \tag{7.21}$$

$$\beta = \frac{Z_k(g_m z_p + 1)Z_{in}}{Z_L g_m z_p Z_m}; \qquad Z_{in} = \frac{z_p Z_L}{z_p + Z_L} \tag{7.22}$$

Because the gain of the open loop amplifier is

$$A_o = -g_m Z_m$$

the feedback factor of the degenerative amplifier is

$$\beta A_o = -\frac{Z_k(g_m z_p + 1)}{z_p + Z_L} \tag{7.23}$$

The general equation for the voltage amplification of a feedback amplifier is

$$A_c = \frac{A_o}{1 - \beta A_o} \tag{7.24}$$

Hence the gain of the degenerative amplifier is

$$A_c = -\frac{g_m Z_m}{1 + Z_k(g_m z_p + 1)/(z_p + Z_L)} \tag{7.25}$$

$$= -g_m' Z_m \tag{7.26}$$

This is precisely the same result as that derived in chapter 3.

The grounded cathode amplifier with cathode degeneration is a general amplifier type of considerable practical importance. The effects of this kind of feedback are assessed for a particular case in the next section as an illustration of the general procedure that should be followed in other cases.

7.5. Effect of Cathode Degeneration on Response and Gain

It has been shown that the general voltage amplification equation for a grounded cathode amplifier with cathode degeneration is

$$A_c = -\frac{g_m Z_m}{1 + Z_k(g_m z_p + 1)/(z_p + Z_L)}$$

The various terms in this equation are defined as: Z_m = mutual impedance of the passive network in the equivalent plate circuit; z_p = plate impedance of the amplifier tube $= r_p\omega_p/(s + \omega_p)$; Z_L = input impedance of the connected load circuit.

The feedback occurs in the cathode impedance Z_k. This impedance can have nearly any circuit configuration. For the purpose of this analysis, it is assumed to be a pure resistance so that $Z_k = R_k$.

The nature of the connected load circuit must be specified before we can proceed any further with the problem. We will assume a resistance coupled amplifier. Therefore, for the resistance coupled amplifier in the mid-frequency region, it was shown in chapter 4 that

$$A_o = -A_r = -g_m R$$

where
$$R = \frac{1}{1/r_p + 1/R_L + 1/R_g}$$

At the low frequency end of the band, the gain equation for the nondegenerative amplifier was derived in chapter 4 and shown to be

$$A_o \text{ (low frequency)} = -A_r \frac{s}{s + \omega_1} \tag{7.27}$$

In the low frequency region the plate impedance is virtually equal to the plate resistance because the effect of the tube interelectrode capacitance is negligible. That is,

$$z_p \text{ (low frequency)} = r_p \tag{7.28}$$

Other terms required in the calculation are easily defined from the equivalent circuit of the amplifier as follows:

$$Z_m = R \frac{s}{s + \omega_1}; \qquad \omega_1 = \frac{1}{(R_1 + R_g)C_c}$$

$$Z_L = R_x \frac{s + \omega_g}{s + \omega_y}; \qquad \omega_g = \frac{1}{R_g C_c}$$

$$R_x = \frac{R_L R_g}{R_L + R_g}; \qquad \omega_y = \frac{1}{(R_L + R_g)C_c}$$

$$R_1 = \frac{r_p R_L}{r_p + R_L}$$

Substitute these expressions into the general gain equation for the degenerative amplifier and the result is

$$A_c = -\frac{A_r \left(\dfrac{s}{s + \omega_1}\right)}{1 + \dfrac{R_k}{R_x}\left(\dfrac{\mu + 1}{\mu}\right)\left(\dfrac{s + \omega_y}{s}\right)\left(\dfrac{s}{s + \omega_1}\right) A_r} \tag{7.29}$$

After some rearrangement of terms, this can be written

$$A_c = -\left[\frac{A_r}{1 + \dfrac{R_k}{R_x}\left(\dfrac{\mu + 1}{\mu}\right) A_r}\right] \cdot \left[\frac{s}{s + \dfrac{\omega_1 + \dfrac{R_k}{R_x}\left(\dfrac{\mu + 1}{\mu}\right) A_r \omega_y}{1 + \dfrac{R_k}{R_x}\left(\dfrac{\mu + 1}{\mu}\right) A_r}}\right] \tag{7.30}$$

Or, in an alternative form,

$$A_c = -A_r'\left(\frac{s}{s + \omega_L}\right) \tag{7.31}$$

where
$$A_r' = \frac{A_r}{1 + \dfrac{R_k}{R_x}\left(\dfrac{\mu + 1}{\mu}\right) A_r} \tag{7.32}$$

$$\omega_L = \omega_1 \frac{1 + R_k\left(\dfrac{R_1 + R_g}{R_L R_g}\right)\left(\dfrac{\mu + 1}{\mu}\right) A_r}{1 + R_k\left(\dfrac{R_L + R_g}{R_L R_g}\right)\left(\dfrac{\mu + 1}{\mu}\right) A_r} \tag{7.33}$$

From equation (7.31) you can easily see that the reference gain A_r' is always less than that of the nondegenerative amplifier. Also, because

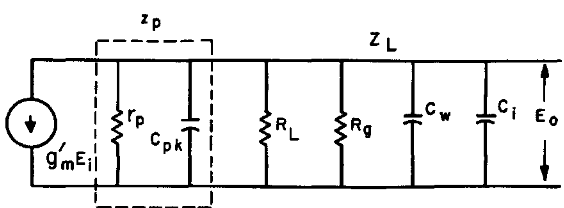


Fig. 7.9. High frequency equivalent plate circuit.

R_1 is always less than R_L, the lower cutoff frequency is always less than that in the nondegenerative case, though the difference is very small.

In the high frequency region the situation is more complex and more obscure. The general gain equation for voltage amplification of an amplifier with cathode degeneration was given earlier. From the high frequency equivalent circuit given in figure (7.9), the various terms in this equation can be shown to have the following values:

$$z_p = r_p \frac{\omega_p}{s + \omega_p} \quad \text{where} \quad \omega_p = \frac{1}{r_p C_{pk}} \tag{7.34}$$

$$Z_L = R_x \frac{\omega_x}{s + \omega_x} \quad \text{where} \quad R_x = \frac{R_L R_g}{R_L + R_g} \tag{7.35}$$

$$\omega_x = \frac{1}{R_x C_x}; \qquad C_x = C_i + C_w \tag{7.36}$$

$$Z_m = \frac{z_p Z_L}{z_p + Z_L} \tag{7.37}$$

These expressions should be substituted into the voltage amplification equation. After a considerable amount of algebraic manipulation, the final result can be expressed as

$$A(s) = \frac{A_r'}{1 + s\dfrac{\omega_x(R_x/R_k+1)+\omega_p(\mu+1+r_p/R_k)}{\omega_p\omega_x\left(\mu+1+\dfrac{r_p+R_x}{R_k}\right)} + s^2\dfrac{1}{\omega_p\omega_x\left(\mu+1+\dfrac{r_p+R_x}{R_k}\right)}}$$

The A_r' factor shown is given by equation (7.32), the reference gain of the amplifier.

The high frequency gain function just given above has the general form

$$A(s) = -\frac{A_r'}{1 + as + bs^2} \tag{7.38}$$

where a and b are the coefficients of s and s^2 in the gain equation. In the steady state this becomes

$$A(j\omega) = \frac{A_r' \underline{/\theta_H}}{\sqrt{(1 - b\omega^2)^2 + a^2\omega^2}} \tag{7.39}$$

At the upper cutoff frequency ω_H the radical must equal $\sqrt{2}$.

Hence
$$(1 - b\omega_H^2)^2 + a^2\omega_H^2 = 2 \tag{7.40}$$

Eventually this can be written

$$\omega_H^4 - \frac{2b - a^2}{2b^2}\omega_H^2 - \frac{1}{b^2} = 0 \tag{7.41}$$

Solve for ω_H and rearrange terms somewhat to obtain

$$\omega_H = \sqrt{\frac{1}{b}\left(1 - \frac{a^2}{2b}\right) + \sqrt{\left(1 - \frac{a^2}{2b}\right)^2 + 1}} \tag{7.42}$$

It is easily proved that + signs are required for each radical if ω_H is to be a real and realizable number.

In amplifiers of this type it is nearly always true that

$$\frac{a^2}{2b} \gg 1 \quad \text{or} \quad \frac{2b}{a^2} \ll 1 \tag{7.43}$$

so that the upper cutoff frequency is approximately

$$\omega_H \doteq \sqrt{\frac{a^2}{2b^2}\left[-1 + \sqrt{1 + \left(\frac{2b}{a^2}\right)^2}\right]} \tag{7.44}$$

The factor inside the radical has the general form of

$$(1 + x^2)^{1/2}$$

where x is much less than 1. This can be closely approximated by $1 + x^2/2$. Then, using this approximation, the expression for the upper cutoff frequency becomes

$$\omega_H \doteq \sqrt{\frac{a^2}{2b^2}\left[-1 + 1 + \frac{1}{2}\left(\frac{2b}{a^2}\right)^2\right]} \doteq \frac{1}{a} \tag{7.45}$$

Therefore the upper cutoff frequency is approximately

$$\omega_H \doteq \frac{\omega_p\omega_x\,[\mu + 1 + (r_p + R_x)/R_k]}{\omega_x(R_x/R_k + 1) + \omega_p(\mu + 1 + r_p/R_k)} \tag{7.46}$$

The amplifier figure of merit, which is $A_r'\omega_H = F_a'$, works out to be

$$F_a' = F_a \frac{1}{1 + (C_{pk}/C_T)(R_k/R_x) + (C_x/C_T)\left(\dfrac{\mu + 1}{\mu}\right)g_m R_k} \tag{7.47}$$

where $F_a = g_m/C_T =$ figure of merit of the nondegenerative amplifier.

From this last equation it is clear that the amplifier figure of merit is degraded through the use of cathode degeneration. In most cases the amplifier cutoff frequency will be only slightly affected compared with the reduction in reference gain.

The main advantage of cathode degeneration is obviously unrelated to the figure of merit, because the reference gain without degeneration is $A_r = g_m R$, while with degeneration it is approximately

$$A_r' \doteq \frac{g_m R}{1 + g_m R_k} \tag{7.48}$$

Thus, it is clear that the reference gain is less dependent upon variations in transconductance when cathode degeneration is used. This is more apparent when the rate of change of gain as a function of transconductance is evaluated. That is,

$$\frac{\partial A_r}{\partial g_m}\text{(no degeneration)} \quad = R$$

$$\frac{\partial A_r'}{\partial g_m}\text{(with degeneration)} = \frac{R}{(1 + g_m R_k)^2}$$

It is clear that the degenerative amplifier will have a reference gain less dependent upon transconductance changes than the nondegenerative amplifier.

The expression for the voltage gain of a *grounded grid* amplifier is

$$A = g_m'' Z_m$$

where g_m'' = effective transconductance of the grounded grid amplifier circuit; $g_m'' = (\mu + 1)g_m'/\mu$, where g_m' = effective transconductance of the degenerative grounded cathode amplifier. Consequently, the response characteristics of a grounded grid amplifier are identical to those of a degenerative grounded cathode amplifier with the gain and figure of merit modified by the $(\mu + 1)/\mu$ factor.

7.6. Internal Impedance with Cathode Degeneration

The two preceding sections dealt with cathode degeneration as an example of negative, current-controlled feedback. The effect of this feedback on the output impedance of the amplifier is of some interest. The standard method for computing this impedance is:

(1) Short circuit the input terminals of the amplifier.

(2) Place a generator across the output terminals.

(3) Measure the voltage and current at the output terminals.

(4) The ratio of the output circuit voltage and current is the output impedance of the amplifier.

The circuit connections required to determine the amplifier output impedance are shown in figure (7.10) for the amplifier with and without cathode degeneration.

The output impedance of the amplifier in the absence of cathode degeneration and with a two-terminal load circuit is easily computed from figure (7.10a) to be

$$Z_o = \frac{z_p Z_L}{z_p + Z_L} \tag{7.49}$$

When cathode degeneration is used, the situation is a little more complex. The application of the generator voltage E_c across the amplifier output causes a voltage E_k to be generated across Z_k, as shown in figure (7.10b). The total grid voltage E_g is $E_g = E_i - E_k$, where E_i is the input signal voltage. However, in this case, E_i is zero because the input terminals are short circuited. Hence $E_g = -E_k$ and the equivalent generator in the plate circuit is

$$-g_m z_p E_g = +g_m z_p E_k$$

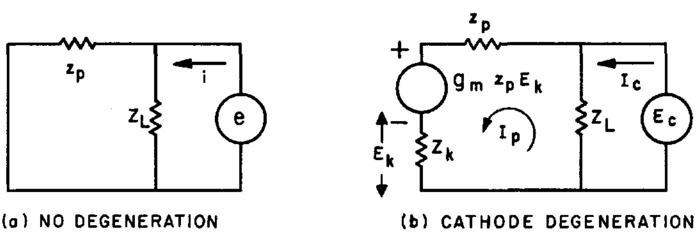


Fig. 7.10. Connections for measuring output impedance.

As a result, the voltage loop equation around the plate circuit of figure (7.10b) is

$$I_p[Z_k(g_m z_p + 1) + z_p + Z_L] = I_c Z_L \tag{7.50}$$

Therefore

$$I_p = I_c \frac{Z_L}{z_p + Z_L + Z_k(g_m z_p + 1)} \tag{7.51}$$

The generator voltage is

$$E_c = (I_c - I_p)Z_L$$

or

$$E_c = I_c\left[1 - \frac{Z_L}{z_p + Z_L + Z_k(g_m z_p + 1)}\right]Z_L \tag{7.52}$$

After a little standard algebraic manipulation it is possible to write

$$Z_c = \frac{E_c}{I_c} = \frac{z_p Z_L/(z_p + Z_L) + Z_k(g_m z_p + 1)Z_L/(z_p + Z_L)}{1 + Z_k(g_m z_p + 1)/(z_p + Z_L)} \tag{7.53}$$

However, a more general terminology can be used because it was previously shown that $Z_o = z_p Z_L/(z_p + Z_L) =$ output impedance of the open loop amplifier, and $-\beta A_o = Z_k(g_m z_p + 1)/(z_p + Z_L)$ $=$ feedback factor. Consequently, equation (7.53) reduces to

$$Z_c = \frac{Z_o}{1 - \beta A_o} - \frac{\beta A_o}{1 - \beta A_o} Z_L \tag{7.54}$$

Of course, this is a negative feedback circuit, so that

$$Z_c = \frac{Z_o}{1 + \beta A_o} + \frac{\beta A_o}{1 + \beta A_o} Z_L \tag{7.55}$$

This can be put into an alternative form as

$$Z_c = Z_o \frac{1 + \beta A_o(1 + Z_L/z_p)}{1 + \beta A_o} \tag{7.56}$$

It is obvious that the numerator will always be larger than the denominator, so that the internal impedance of the amplifier with current-controlled feedback is always larger than that obtained from the same amplifier without feedback.

The result should be expected, because the feedback acts to stabilize the plate current, thereby making the tube operate more like a constant current device. Constant current sources have high internal impedances. Essentially the same result is achieved with the grounded grid amplifier.

7.7. A Simple Case of Voltage Feedback, Cathode Follower

The simplest and commonest type of negative *voltage-controlled* feedback is the cathode follower shown in figure (7.11). Actually, as long as Z_L is a three-terminal network, this is not a true case of voltage-controlled feedback. However, it is true voltage controlled feedback when Z_L has only two terminals.

The equation for the voltage gain of a cathode follower was derived in chapter 3 and expressed as:

$$A = +g_m' Z_m = \frac{g_m Z_m}{1 + \mu Z_L/(z_p + Z_L)} = \frac{g_m Z_m}{1 + g_m Z_{in}} \tag{7.57}$$

where Z_m = mutual impedance of the passive network in the equivalent plate circuit; Z_{in} = input impedance of the passive network in the equivalent plate circuit; Z_L = input impedance of the connected load; $g_m Z_m$ = gain of the circuit if the amplifier had been connected as a grounded cathode circuit.

The general equation for the voltage amplification of a feedback amplifier has been proven to be

$$A_c = \frac{A_o}{1 - \beta A_o}$$

or, for the negative feedback case,

$$A_c = \frac{A_o}{1 + \beta A_o} \tag{7.58}$$

Comparison of this relationship with that for the cathode follower shows that

$$A_o = g_m Z_m \tag{7.59}$$

$$\beta A_o = \frac{\mu Z_L}{z_p + Z_L} = g_m Z_{in} \tag{7.60}$$

The feedback factor can also be expressed as

$$\beta A_o = \beta g_m Z_m \tag{7.61}$$

so that the voltage transfer function of the feedback circuit is

$$\beta = \frac{Z_{in}}{Z_m} \tag{7.62}$$

Now consider a resistance coupled cathode follower as shown in figure (7.12). This amplifier will have the same equivalent plate circuits

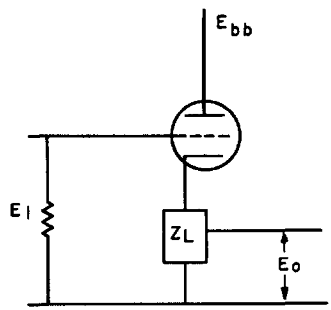


Fig. 7.11. Cathode follower.

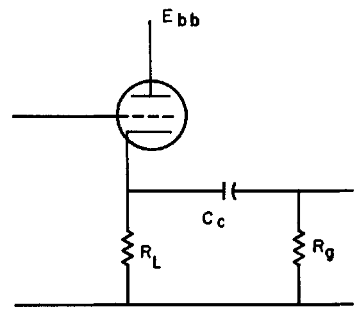


Fig. 7.12. Resistance coupled cathode follower.

as the resistance coupled grounded cathode amplifier except for the change in amplifier transconductance. In chapter 4 for the grounded cathode amplifier, it was shown that

$$A_r = g_m R \quad \text{where} \quad R = \frac{1}{1/r_p + 1/R_L + 1/R_g}$$

$$\omega_1 = \text{lower cutoff frequency} = \frac{1}{(R_1 + R_g)C_c}$$

$$R_1 = \frac{r_p R_L}{r_p + R_L}$$

$$\omega_2 = \text{upper cutoff frequency} = \frac{1}{RC_T}$$

$$C_T = C_o + C_w + C_i$$

Consider the mid-frequency, or reference, case for the cathode follower. Here $A_o = A_r = g_m R$ and $Z_{in} = Z_m = R$, so that $\beta = 1$. Therefore the reference gain of a cathode follower is

$$A_r' = \frac{A_o}{1 + \beta A_o} = \frac{A_r}{1 + A_r} \tag{7.63}$$

In the low frequency case for the grounded cathode, resistance coupled amplifier, it was shown in chapter 4 that

$$A_o = A_r \frac{s}{s + \omega_1}; \qquad Z_m = R \frac{s}{s + \omega_1}$$

It is easily shown that

$$Z_{in} = \frac{z_p Z_L}{z_p + Z_L} = \frac{r_p Z_L}{r_p + Z_L} = R \frac{s + \omega_g}{s + \omega_1} \tag{7.64}$$

where $z_p \doteq r_p$

and
$$\omega_g = \frac{1}{R_g C_c} \tag{7.65}$$

Thus
$$\beta = \frac{Z_{in}}{Z_m} = \frac{s + \omega_g}{s} \tag{7.66}$$

Therefore the low frequency gain function of a resistance coupled cathode follower is

$$A_c\,(\text{low}) = \frac{A_r s/(s + \omega_1)}{1 + (s + \omega_g)A_r/(s + \omega_1)} \tag{7.67}$$

or
$$A_c\,(\text{low}) = \frac{A_r}{1 + A_r} \cdot \frac{s}{s + (\omega_1 + \omega_g A_r)/(1 + A_r)} \tag{7.68}$$

and finally
$$A_c\,(\text{low}) = A_r \frac{s}{s + \omega_L} \tag{7.69}$$

Therefore the lower cutoff frequency of the cathode follower is

$$\omega_L = \frac{\omega_1}{1 + A_r}\left(1 + \frac{\omega_g}{\omega_1} A_r\right) \tag{7.70}$$

This frequency is always larger than ω_1 because ω_g is always larger than ω_1. However, the change will usually be rather small because ω_1 and ω_g are nearly equal in most cases.

For the high frequency equivalent circuit, it was shown in chapter 4 that

$$A_o = A_r \frac{\omega_2}{s + \omega_2}$$

for the grounded cathode, resistance coupled amplifier. Also, because the high frequency equivalent plate circuit is a two-terminal network, $Z_{in} = Z_m$. Therefore $\beta = 1$ and the gain equation for the corresponding cathode follower is

$$A_c\,(\text{high}) = \frac{A_r\omega_2/(s+\omega_2)}{1 + A_r\omega_2/(s+\omega_2)} = \frac{A_r\omega_2}{s + \omega_2(1 + A_r)} \tag{7.71}$$

or

$$A_c\,(\text{high}) = \frac{A_r}{1 + A_r}\,\frac{1}{1 + s/[\omega_2(1 + A_r)]} = A_r'\,\frac{1}{1 + s/\omega_H} \tag{7.72}$$

Thus the upper cutoff frequency is

$$\omega_H = \omega_2(1 + A_r) \tag{7.73}$$

and the gain-bandwidth product is

$$F_a = A_r'\omega_H = A_r\omega_2 = \frac{g_m}{C_T} \tag{7.74}$$

As you can easily see, the upper cutoff frequency of a cathode follower is many times larger than the cutoff frequency of a grounded cathode amplifier using the same tube and circuit parameters.

7.8. Internal Impedance with Voltage Feedback

The output, or internal, impedance of an amplifier using voltage feedback can be computed by the same method as that used for the case of current controlled feedback. Thus in the open loop case, the amplifier and its equivalent representation would appear as shown in figure (7.13).

When the input terminals of the amplifier are short circuited this makes $E_i = E_iA_o = E_o = 0$. Therefore, from the circuit in figure (7.13) you can easily see that the output impedance of the open loop amplifier is Z_o.

The corresponding circuits and representations for the voltage-controlled feedback amplifier are shown in figure (7.14). Here the equation around the output circuit is

$$E_c = I_cZ_o + \beta A_oE_c$$

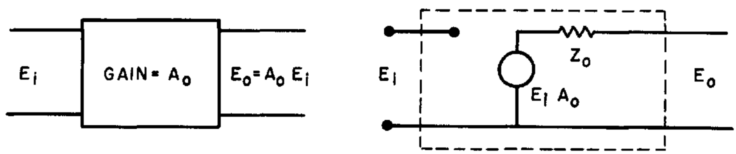
or

$$Z_c = \frac{E_c}{I_c} = \frac{Z_o}{1 - \beta A_o} \tag{7.75}$$

When negative feedback is used, this becomes

$$Z_c = \frac{Z_o}{1 + \beta A_o} \tag{7.76}$$

This shows that the output impedance of the feedback amplifier will always be less than that of the open loop amplifier. This should be anticipated because voltage-controlled feedback acts to stabilize the output voltage and make the circuit act more nearly like a constant voltage source. Constant voltage sources have zero internal impedance.

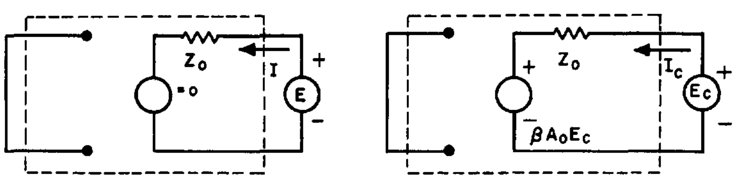


(a) OPEN LOOP AMPLIFIER (b) EQUIVALENT REPRESENTATION

Fig. 7.13. Open loop amplifier.

This formula is easily applied to the cathode follower because $Z_o = R$ in the mid-frequency region and $\beta A_o = A_r$, so that

$$Z_c \text{ (mid-frequency)} = \frac{R}{1 + A_r} \tag{7.77}$$



(a) OPEN LOOP AMPLIFIER (b) VOLTAGE CONTROLLED FEEDBACK AMPLIFIER

Fig. 7.14. Impedance measurements.

7.9. Grid-to-Plate Feedback

All the feedback systems discussed so far have used a feedback voltage inserted in series with the input signal voltage. An entirely different situation obtains when a current is fed back into the input circuit.

One of the simplest circuit arrangements designed to feed current back is the grid-to-plate feedback circuit shown in figure (7.15). In this case, conductive coupling between the grid and plate circuits is used and R_{12} is the feedback resistor. The equivalent plate circuit, neglecting the tube interelectrode capacitances and assuming an ideal input coupling circuit is shown in figure (7.16).

This is a grounded cathode amplifier. The general equation for the voltage gain of a grounded cathode amplifier was shown to be

$$A = -g_m' Z_m'$$

In the absence of feedback, Z_m' is readily computed because it is exactly equal to the Z_m shown in the circuit of figure (7.16). However, when the feedback resistor R_{12} is inserted, the grid circuit is coupled directly into the plate circuit. In the assumed case shown in figure (7.16), this simply adds R_{12} in parallel with Z_m, assuming that the internal impedance of the grid signal generator is zero. Therefore a new and lower value for Z_m' is obtained. As a result, the gain of the amplifier is reduced from the value without feedback.

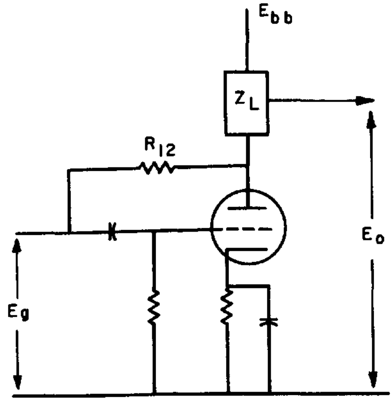


Fig. 7.15. Basic grid-to-plate feedback circuit.

We can easily determine the input impedance of the amplifier; it is necessary only to apply the compensation theorem to the plate circuit. That is, the complex network between the plate and cathode terminals of the circuit of figure (7.16) can be replaced by a generator of terminal

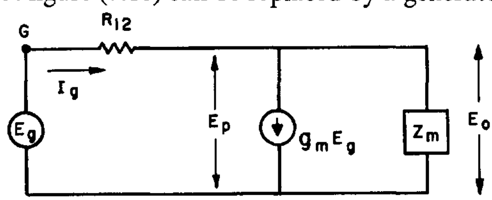


Fig. 7.16. Equivalent plate circuit of the amplifier in figure (7.15).

voltage $E_p = -E_g A_{gp}$ where $A_{gp} = -E_p/E_g$. The voltage loop equation around the resulting circuit, which is shown in figure (7.17), is

$$E_g = I_g R_{12} + E_g A_{gp} \tag{7.78}$$

Therefore the input impedance of the amplifier is

$$Z_g = \frac{E_g}{I_g} = \frac{R_{12}}{1 - A_{gp}} \tag{7.79}$$

or, the input admittance is

$$Y_g = G_{12}(1 - A_{gp}) \tag{7.80}$$

Thus, in the case of the Miller effect discussed in chapter 3, the input impedance of the amplifier is altered because of the grid-to-plate feedback.

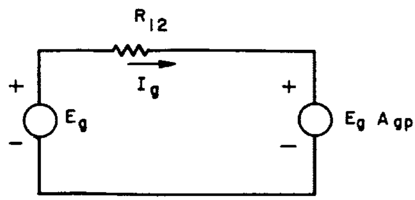


Fig. 7.17. Application of the compensation theorem to the circuit of figure (7.16).

No particular problem exists so far because the meanings of A_{gp} and the stage gain A are perfectly clear and can be calculated without too much difficulty. However, when such an amplifier is driven by another

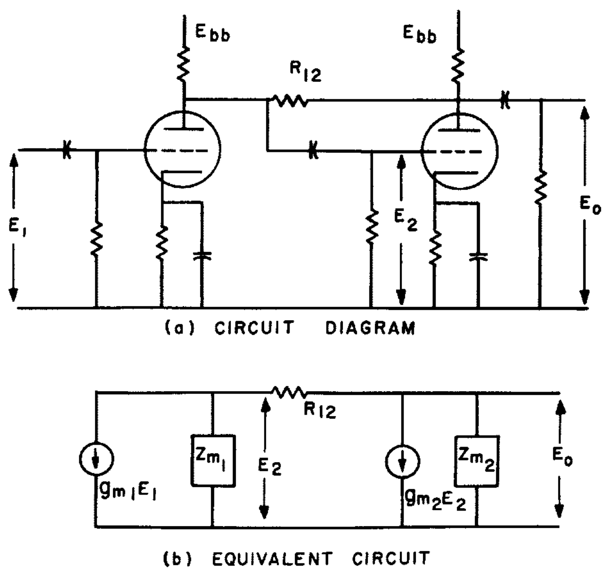


Fig. 7.18. Feedback pair.

amplifier stage, as in the circuit of figure (7.18), the problem is greatly complicated by the removal of the isolation between stages. Because of the conductive coupling between the input and output circuits of the second stage, both tubes have common plate circuits and involve two generators. As a result, the meaning of the gain of each stage is obscure, so that it is virtually impossible to calculate the individual stage gains for the following reasons:

(1) To calculate the gain of stage 1, the input impedance of stage 2 must be known.

(2) The input impedance of stage 2 depends upon the gain of stage 2 and this depends upon the total impedance in the plate circuits of both stages.

It should be clear that the loss of isolation between stages makes gain evaluations by standard techniques a difficult proposition. As a result, it is more convenient to treat the two amplifiers as a single amplifier stage and to compute the over-all system gain by direct analysis of the circuit.

7.10. The Feedback Pair

The discussion in this section is confined to an examination of the properties of a two-tube amplifier in which there is conductive coupling between the input and output of the second tube. Such an amplifier is called a *feedback pair*. If three tubes are used in a similar arrangement, the amplifier is called a *feedback triple*. Any feedback *n-uple* is possible, but the feedback pair is possibly the most common and certainly the easiest to analyze.

To simplify the discussion and mathematics in the exploratory steps, the interstage networks used in the feedback pair will be restricted to two-terminal types. Consequently, a typical feedback pair will appear as shown in figure (7.18a). The corresponding equivalent circuit is shown in figure (7.18b).

The circuit has two current nodes and the nodal equations are clearly

$$g_{m_1}E_1 + \frac{E_2}{Z_{m_1}} + \frac{E_2 - E_o}{R_{12}} = 0 \qquad (7.81)$$

$$g_{m_2}E_2 + \frac{E_o}{Z_{m_2}} + \frac{E_o - E_2}{R_{12}} = 0 \qquad (7.82)$$

In nodal analysis, it is more convenient to use admittances so that the node equations can be written

$$g_{m_1}E_1 + Y_{m_1}E_2 + G_{12}(E_2 - E_o) = 0 \qquad (7.83)$$

$$g_{m_2}E_2 + Y_{m_2}E_o + G_{12}(E_o - E_2) = 0 \qquad (7.84)$$

Solve these two equations simultaneously for the voltage gain of the pair. The result is

$$A = \frac{E_o}{E_i} = \frac{g_{m_1}(g_{m_1} - G_{12})}{G_{12}(g_{m_2} - G_{12}) + (Y_{m_1} + G_{12})(Y_{m_2} + G_{12})} \tag{7.85}$$

This formula is general as long as two-terminal load circuits are used, and this generality makes it difficult to interpret the effects of feedback. Therefore we must discuss specific cases from this point on.

Assume a feedback pair of resistance coupled amplifiers in which the high frequency response characteristics are of interest. Assume that the two stages are identical. Hence the interstage mutual admittances are

$$Y_{m_1} = Y_{m_2} = G + sC_T \tag{7.86}$$

Therefore the equation for the voltage amplification of the feedback pair is

$$A = \frac{g_{m_1}(g_{m_2} - G_{12})}{G_{12}(g_{m_2} - G_{12}) + (G + G_{12} + sC_T)(G + G_{12} + sC_T)} \tag{7.87}$$

A little standard algebraic manipulation will reduce equation (7.87) to the following form:

$$A = \left(\frac{g_{m_1}}{C_T}\right)\left(\frac{g_{m_2} - G_{12}}{C_T}\right)\frac{1}{s^2 + 2\left(\dfrac{G + G_{12}}{C_T}\right)s + \dfrac{G^2 + G_{12}(g_{m_2} + 2G)}{C_T^2}} \tag{7.88}$$

Note that the scale factor of the gain function,

$$\frac{g_{m_1}}{C_T} \cdot \frac{g_{m_2} - G_{12}}{C_T}$$

is less than the scale factor

$$\frac{g_{m_1}}{CT} \cdot \frac{g_{m_2}}{C_T}$$

obtained when two resistance coupled stages are connected in cascade, without feedback.

The equation for the reference gain of the amplifier is easily derived and expressed as

$$A_r' = \frac{g_{m_1}(g_{m_2} - G_{12})}{G^2 + G_{12}(g_{m_2} + 2G)} \tag{7.89}$$

The poles of the gain function are readily evaluated from equation (7.88) and expressed as

$$s_{1,2} = -\frac{G + G_{12}}{C_T} \pm \frac{1}{C_T}\sqrt{G_{12}(G_{12} - g_{m_2})} \tag{7.90}$$

The poles will both be real and the response will be nonoscillatory as long as $G_{12} \geq g_{m_2}$. If this condition prevails, the scale factor will acquire a minus sign and there will be phase inversion through the amplifier.

The response will be oscillatory and will exhibit overshoot when the poles are complex conjugates. This situation occurs when $g_{m_2} > G_{12}$. The scale factor then acquires a positive sign. The poles are

$$s_{1,2} = -\frac{G + G_{12}}{C_T} \pm \frac{j}{C_T}\sqrt{G_{12}(g_{m_2} - G_{12})} \tag{7.91}$$

It is apparent that the poles of the feedback pair can be located at almost any physically realizable pair of points in the complex s plane by suitable adjustment of the three independent circuit parameters, G, G_{12}, and g_{m_2}. This is important because the response characteristics of any amplifier are governed by the locations of the poles of the gain function. Thus an indefinite number of responses can be obtained from a resistance coupled feedback pair.

It should be noted that specification of the pole location requires a knowledge of two factors:

(1) α = real part of the pole = $(G + G_{12})/C_T$.

(2) The radical term, which can be either real or imaginary. However, there are three independent variables, G, G_{12}, and g_{m_2}. As a result, there are an infinite number of solutions for any given problem unless one of the three variables is specified. Ordinarily, g_{m_2} will be specified because the amplifier tubes will be selected with a view toward a high figure of merit. It is fairly common to make both tubes in the pair identical so that $g_{m_1} = g_{m_2}$.

7.11. Bandpass Feedback Pair

The general expression for the over-all gain of a feedback pair was derived in the preceding section for the case of two-terminal load circuits and expressed as

$$A = \frac{g_{m_1}(g_{m_2} - G_{12})}{G_{12}(g_{m_2} - G_{12}) + (Y_{m_1} + G_{12})(Y_{m_2} + G_{12})} \tag{7.92}$$

Now consider the case where both stages of the feedback pair have identical single tuned interstage networks of the form shown in figure (7.19). In this case

$$Y_{m_1} = Y_{m_2} = G + sC_T + \frac{1}{sL} \tag{7.93}$$

Then let

$$Y = Y_{m_1} + G_{12} = Y_{m_2} + G_{12} \tag{7.94}$$

$$= (G + G_{12}) + sC_T\left(1 + \frac{1}{s^2LC_T}\right) \tag{7.95}$$

The $1/LC_T$ factor is the square of the undamped natural frequency of the resonant circuit. This is identified as ω_0^2. Then, in the steady state, the admittance Y becomes

$$Y(j\omega) = (G + G_{12}) + jC_T\frac{(\omega + \omega_0)(\omega - \omega_0)}{\omega} \tag{7.96}$$

If the circuit is a narrow band or high Q circuit, all frequencies of interest will fall near ω_0. Therefore

$$\omega + \omega_0 \doteq 2\omega$$

and

$$\omega - \omega_0 = \pm\Delta\omega$$

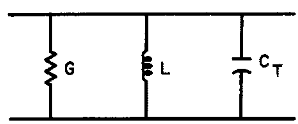


Fig. 7.19. Single tuned circuit.

The admittance function can now be written approximately as

$$Y(j\omega) \doteq (G + G_{12}) \pm j2\,\Delta\omega\, C_T \tag{7.97}$$

This approximate equation for the admittance should be substituted into the amplifier gain function. The result is

$$A(j\omega) = \frac{g_{m_1}(g_{m_2} - G_{12})}{G_{12}(g_{m_2} - G_{12}) + (G + G_{12} \pm j2\Delta\omega\, C_T)^2} \tag{7.98}$$

Multiply terms in the denominator; collect real and imaginary parts. The resulting gain function is

$$A(j\omega) = \frac{g_{m_1}(g_{m_2} - G_{12})/C_T^2}{\dfrac{G_{12}(g_{m_2} - G_{12}) + (G + G_{12})^2}{C_T^2} \pm j\dfrac{4(G + G_{12})\,\Delta\omega}{C_T} + (2j\,\Delta\omega)^2} \tag{7.99}$$

The denominator is a quadratic function in $\Delta\omega$, so the gain function has two poles in terms of $\Delta\omega$ as a variable. These poles can be either real or complex, depending upon the relative magnitudes of the variables. To obtain general results, it is best to assume that the poles will be complex, and if they are complex, they must be complex conjugates. Hence the poles are specified as

$$-P_1 = \alpha + j\beta \quad \text{and} \quad -P_2 = \alpha - j\beta \tag{7.100}$$

Then the gain function can be written

$$A(j\omega) = \frac{g_{m_1}}{C_T} \cdot \frac{g_{m_2} - G_{12}}{C_T} \cdot \frac{1}{(P_1 \pm j2\ \Delta\omega)(P_2 \pm j2\ \Delta\omega)} \tag{7.101}$$

For practical reasons it is necessary to establish an equivalence between these poles and the circuit constants. This is easily done by multiplying terms out in equation (7.101) so that the gain function is

$$A(j\omega) = \frac{g_{m_1}(g_{m_2} - G_{12})/C_T^2}{P_1P_2 \pm 2j(P_1 + P_2)\ \Delta\omega + (2j\ \Delta\omega)^2} \tag{7.102}$$

If the *component forms* of P_1 and P_2 are used, the denominator of equation (7.102) reduces to

$$(\alpha^2 + \beta^2) \pm j(2\alpha)2\ \ \Delta\omega + (2j\ \ \Delta\omega)^2 \tag{7.103}$$

Equate coefficients of like powers of $\Delta\omega$ in this equation to those in equation (7.99). This process yields two equations:

$$\alpha^2 + \beta^2 = \frac{G_{12}(g_{m_2} - G_{12}) + (G + G_{12})^2}{C_T^2} \tag{7.104}$$

$$\alpha = -\frac{G + G_{12}}{C_T} \tag{7.105}$$

The imaginary part of the pole is now easily calculated to be

$$\beta = \frac{\sqrt{G_{12}(g_{m_2} - G_{12})}}{C_T} \tag{7.106}$$

Hence the poles in the complex P plane are

$$-P_{1,2} = -\frac{G + G_{12}}{C_T} \pm j\frac{\sqrt{G_{12}(g_{m_2} - G_{12})}}{C_T} \tag{7.107}$$

These poles are of exactly the same form as those calculated for the low pass feedback pair in the s plane. The result might have been anticipated from a consideration of the bandpass, low pass transformation discussed in chapter 6.

The poles can be positioned to nearly any desired point in the complex P plane. Nearly any response characteristic can be obtained from a feedback pair.

For example, suppose that a feedback pair is to be used to provide the same response as a maximally flat staggered pair. According to chapter 5, the P plane poles of a flat staggered pair are

$$P_{1,2} = (-0.707 \pm j0.707)B_n \tag{7.108}$$

where B_n = over-all bandwidth of the pair. To make the response of the feedback pair identical to that of the staggered pair, it is necessary only to equate pole components. As a result,

$$\frac{G + G_{12}}{C_T} = 0.707B_n \tag{7.109}$$

$$\frac{1}{C_T}\sqrt{G_{12}(g_{m_2} - G_{12})} = 0.707B_n \tag{7.110}$$

Any attempt to develop useful design equations directly from these two expressions is fruitless because there are three independent variables and only two equations. However, in many cases it can be assumed that g_{m_2} is much larger than G_{12}. As a result, equation (7.110) becomes

$$\frac{1}{C_T}\sqrt{G_{12}g_{m_2}} \doteq 0.707B_n \tag{7.111}$$

Solve equations (7.109) and (7.111) to obtain

$$G_{12} = \frac{(B_n C_T)^2}{2g_{m_2}} \tag{7.112}$$

$$G = 0.707B_n C_T - G_{12} \tag{7.113}$$

These equations can be expressed more conveniently as

$$R_{12} = \frac{2}{g_m}\left(\frac{F_a}{B_n}\right)^2 \tag{7.114}$$

$$R = \frac{\sqrt{2}F_a/g_m B_n}{1 - 0.707B_n/F_a} \tag{7.115}$$

where g_m = tube transconductances; $F_a = g_m/C_T$.

From this example you can see that the response of any maximally flat staggered n-uple can be duplicated by a cascade of feedback pairs and single tuned stages. The method of duplicating the response of over staggered amplifiers should also be apparent now.

7.12. Stability in Feedback Amplifiers

The preceding sections discussed the effects of feedback on the operating characteristics of amplifiers. A number of desirable effects were noted.

In practical cases the problem is always complicated by the variation in phase shift with frequency in the feedback circuit. This can often result in positive feedback at certain frequencies, even though negative feedback is desired. This positive feedback may become so large that it will regenerate any disturbance in the amplifier and cause the circuit to burst into sustained oscillation. Thus the payment for the advantages of negative feedback is generally made in the increased effort necessary to make sure that the amplifier is reliably stable at all frequencies.

On the other hand, it may be that an *oscillator* is desired rather than an amplifier; in this case it is important to know the conditions necessary to guarantee circuit oscillation.

If the term *stability* is taken to mean that the output from a circuit will eventually die out when the input is removed, interest centers upon the conditions that make an amplifier stable and an oscillator unstable.

A method for determining the relative stability of a circuit was given in chapter 2. It was shown there that if the system transfer function does not have right half plane poles, the system is stable. Right half plane poles give rise to instability of the form of continuously increasing currents and voltages. Hence, once the poles of the transfer function are known, it is an easy matter to evaluate the stability of the system. The test is simple and direct.

The transfer function of a feedback amplifier is

$$A_c(s) = \frac{A_o(s)}{1 - \beta(s)A_o(s)}$$

The circuit will be stable if this function does not involve right half plane poles. It is generally valid to assume that the open loop amplifier is stable, so that A_o does not involve right half plane poles. The same assumption is generally true for the feedback circuit. Consequently, if instability is to be produced, it must arise from the *right half plane zeros of the denominator* of the foregoing gain equation. Thus the problem of studying the stability of a feedback circuit is reduced to a consideration of the zeros of $1 - \beta A_o$.

Superficially it would appear that it is a relatively easy matter to determine the stability of a feedback amplifier by the foregoing technique. However, the simplicity is more apparent than real, because the zeros, or roots, of $1 - \beta A_o$ must be known. This can be a difficult and tedious evaluation. In most practical cases it would require that specific values be assigned to the various elements of the networks. Then, when the zeros have been determined and the system is found to be unstable, the locations of the zeros do not provide any clues as to how the instability can be removed by adjustment of the circuit parameters.

An ideal technique for analyzing feedback systems would: (1) indicate the relative stability; (2) indicate stability boundaries and show how operation should be changed to insure stability; (3) not be entirely dependent upon numerical methods.

There are two methods that partially meet these requirements: (1) the Nyquist criterion; (2) the Routh-Hurwitz criterion. These two methods are treated briefly in the next two sections.

7.13. The Nyquist Stability Criterion

Proof of the Nyquist criterion is omitted because it can be found elsewhere;[1] it involves mathematical methods that cannot be introduced easily. However, the basic mechanical details involved in its use will be explained.

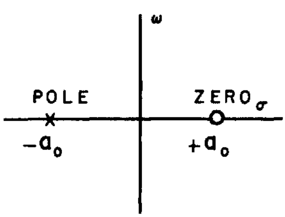


Fig. 7.20. Right half plane zero.

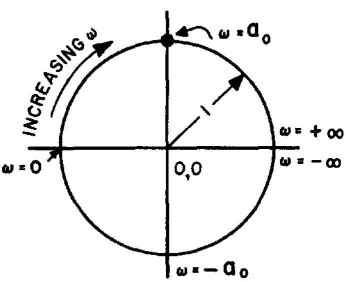


Fig. 7.21. Polar plot of $F(j\omega)$ as a function of frequency.

As a simple illustrative example, assume a specific function of the form

$$F(s) = 1 - \beta(s)A_o(s) = \frac{s - a_o}{s + a_o} \tag{7.116}$$

This function has a pole at $-a_o$ and a zero at $+a_o$ as shown in figure (7.20). It is known that the system will be unstable because of the existence of the right half plane zero.

Now rewrite equation (7.116) in terms of the steady state frequency as

$$F(j\omega) = \frac{j\omega - a_o}{j\omega + a_o} = 1 \,\underline{/\tan^{-1}\frac{\omega}{-a_o} - \tan^{-1}\frac{\omega}{a_o}} \tag{7.117}$$

[1] H. Nyquist, "Regeneration Theory," *Bell System Tech. J.*, vol. 11, pp. 126–147, January, 1932; also E. Peterson, J. G. Kreer, and L. A. Ware, "Regeneration Theory and Experiment," *Bell System Tech. J.*, vol. 13, pp. 680–700, October, 1934.

When this function is plotted in polar coordinates as a function of frequency, it will appear as shown in figure (7.21). The plot is made by locating the magnitude of the function at the proper angle. It is clear from figure (7.21) that the resulting contour encircles the origin once in a clockwise direction as the frequency is varied from $-\infty$ to $+\infty$. This encirclement is caused by the right half plane zero. It is a general rule and it can be proven, that the polar plot of $F(j\omega)$ will encircle the origin once in a clockwise direction for each right half plane zero.

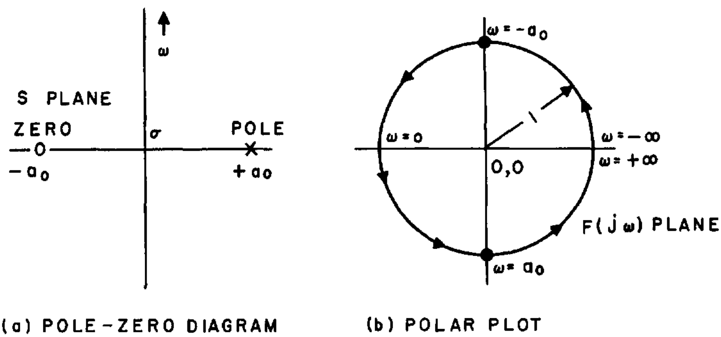


Fig. 7.22. Effect of a right half plane pole.

Now consider the reverse case where a function is defined so that it has a *right half plane pole.* Thus, assume

$$F(s) = 1 - \beta(s)A_o(s) = \frac{s + a_o}{s - a_o} \tag{7.118}$$

and therefore $$F(j\omega) = 1 \angle \tan^{-1}\frac{\omega}{a_o} - \tan^{-1}\frac{\omega}{-a_o} \tag{7.119}$$

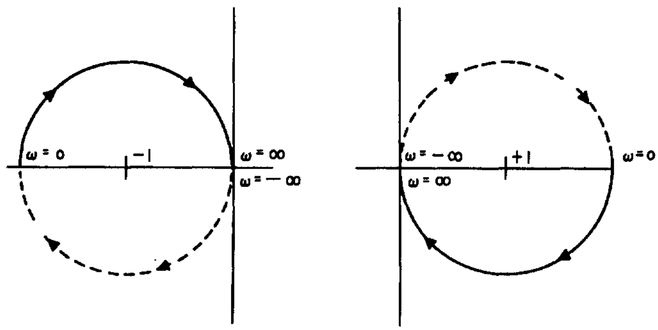
The pole-zero diagram and corresponding polar plot are shown in figure (7.22). In this case it is found that the origin is encircled once in the *counterclockwise* direction for each right half plane pole.

Therefore the number of right half plane poles and zeros can be determined by counting the number and direction of the encirclements of the origin by the polar plot of $1 - \beta A_o$. Of course, for the determination of stability, only the right half plane zeros of this function are of interest, and the Nyquist stability criterion deals with them.

The polar plot of $[1 - \beta(j\omega)A_o(j\omega)]$ is called a *Nyquist diagram,* and the Nyquist criterion is stated as follows:

If the polar plot of $[1 - \beta(j\omega)A_o(j\omega)]$ encircles the origin in a clockwise traverse from $\omega = -\infty$ to $+\infty$, the system is unstable.

In practice, the determination of the Nyquist diagram can be simplified by the transformation shown in figure (7.23). First, the polar plot of $[1 - \beta(j\omega)A_o(j\omega)]$ is changed to a plot of $-\beta(j\omega)A_o(j\omega)$ by translating the origin of coordinates a unit distance to the right. Then encirclements about the point $(-1 + j0)$ are counted. If the entire plane is rotated through 180°, as shown in figure (7.23b), the result is a



(a) POLAR PLOT OF $-\beta(j\omega)A_0(j\omega)$ (b) POLAR PLOT OF $+\beta(j\omega)A_0(j\omega)$

Fig. 7.23. Transformations of the Nyquist diagram.

plot of $+\beta(j\omega)A_o(j\omega)$ and stability is determined by counting encirclements of the point $(1 + j0)$.

The particularly unique advantage of the Nyquist criterion is clear from figure (7.23b). Here it is necessary only to determine the *open loop* transfer function of $A_o(j\omega)$ and $\beta(j\omega)$ connected in cascade. This may be done analytically or experimentally. Then a polar plot is made and the stability question is answered by counting encirclements of the point $+1$. Thus a relatively simple set of computations or measurements are required on an open loop system and the stability of the closed loop system can be predicted with certainty. A more detailed treatment will be found elsewhere.[2]

One practical matter remains to be discussed. It was stated in chapter 2 that nearly all common amplifiers are of the minimum phase shift

[2] Gordon S. Brown and Donald P. Campbell, *Principles of Servomechanisms*, John Wiley & Sons, Inc., New York, 1948, chaps. 6–8.

type. Methods of computing the phase response from the amplitude response have been worked out.[3] Hence it is necessary only to compute or measure the amplitude reponse, because the phase response can be computed from that and the Nyquist diagram plotted.

7.14. Routh-Hurwitz Criterion[4]

Two major disadvantages are involved in the use of the Nyquist criterion.[4] The most important is that it is primarily a numerical process. Specific values for circuit constants must be known before the Nyquist diagram can be computed or experimentally determined. Second, the problem of stabilizing an unstable circuit is essentially a cut-and-try determination of the stability boundaries. The great advantage of the method is that it permits computations and analyses to be made from experimentally determined data, and this often overrides all other considerations.

The method developed by Routh and Hurwitz working independently is essentially analytical, and cannot usually be used with experimentally determined data. Instead, it depends upon the relationships between the coefficients in the denominator of the system transfer function.

It was shown earlier that the transfer functions of all circuits have the general form of the ratio of two polynomials, such as

$$F(s) = K\frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_o}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_o} \tag{7.120}$$

The *characteristic equation* is obtained when the polynomial in the denominator is equated to zero. Clearly, the roots of the characteristic equation are the poles of the system transfer function and thereby determine the nature of the response.

To use the Routh-Hurwitz method, the coefficients of the characteristic equation should be arranged in two rows as follows:

1st row s^n	a_n	a_{n-2}	a_{n-4}	$\cdots$	etc.
2nd row s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	$\cdots$	etc.

[3] See reference (1) of chapter 4.

[4] Frank E. Bothwell, "Nyquist Diagrams and the Routh-Hurwitz Stability Criterion," *Proc. IRE*, vol. 38, November, 1950, pp. 1345–1348; also M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, John Wiley & Sons, Inc., New York, 1942, pp. 197–201.

A third row is formed by cross multiplication of the terms in the first two rows, as follows:

$$\begin{array}{ll|c|c|c}
\text{1st row} & s^n & a_n & a_{n-2} & a_{n-4} \\
\text{2nd row} & s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} \\
\text{3rd row} & s^{n-2} & \dfrac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} & \dfrac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} & \text{etc.}
\end{array}$$

This same process is followed, using rows 2 and 3 to form row 4. Then, rows 3 and 4 are used to form row 5, and so on. The last row should correspond to s^0 so that there are $(n + 1)$ rows. Under normal conditions, the procedure is straightforward.

There are two special cases that might arise in the formation of the rows following the first two. For example, the characteristic equation might be missing a power of s. If this is so, a zero is placed in the table to account for the missing term. If this zero is the first term in any row, but the other terms are not zero, then the next row cannot be formed because all the terms will be infinite. This difficulty can be circumvented by a slight artifice. Replace the zero by a differentially small quantity designated ϵ. Then coefficients are formed in the usual manner. Ordinarily terms involving ϵ^2 need not be retained.

Another special case arises when all the terms in a given row are zero. Any attempt to form the following row then fails. The procedure is then:

(1) Form an *auxiliary equation* with the coefficients of the last non-vanishing row and commencing with the power of s specified by that row. The roots of this equation are also roots of the characteristic equation.

(2) Differentiate the auxiliary equation with respect to s and enter the coefficients of the resulting equation into the table in place of the zeros.

After the complete *Routh-Hurwitz determinant* is formed by the methods just outlined, the coefficients in the *first column* are inspected. The number of times that the signs of these coefficients change is numerically equal to the number of roots of the characteristic equation in the right half plane. Hence a stable system will *not* exhibit any sign changes in the first column coefficients, nor should any roots fall on the vertical axis. The constants of the physical system should be adjusted to assure this condition.

For example, consider a physical system having a cubic characteristic equation as follows:

$$a^3s^3 + a_2s^2 + a_1s + a_0 = 0$$

The Routh-Hurwitz determinant is easily constructed as follows:

$$\begin{array}{ll|c|c|}
\text{1st row} & s^3 & a_3 & a_1 \\
\text{2nd row} & s^2 & a_2 & a_0 \\
\text{3rd row} & s^1 & \dfrac{a_1a_2 - a_3a_0}{a_2} & 0 \\
\text{4th row} & s^0 & a_0 & 0
\end{array}$$

If this is a physical system, the coefficients themselves are all positive. If the system is to be stable, there cannot be any sign changes in the first column, so that $a_1a_2 > a_3a_0$. This is the condition for stability and also defines the stability boundaries.

PROBLEMS

7.1. An amplifier has a voltage gain of 46.2 volt-logits at the mid-frequency without feedback. If negative feedback is added in an amount such that $\beta = 0.01$, compute the closed loop gain of the amplifier.

7.2. An amplifier has a voltage gain of 35 volt-logits in the reference case without feedback. When feedback is applied, the gain is reduced to 25 volt-logits. Compute β.

7.3. An amplifier without feedback produces an output voltage of 125 v with 15% harmonic distortion. The input signal voltage is 65 mv. The amplifier frequency response is flat over the range of interest. If 1% of the output is fed back to produce negative feedback, compute the resulting harmonic distortion in the output and the new value for the input voltage required to keep the output at 125 v.

7.4. Without feedback and at rated power supply voltage, a given amplifier with a flat frequency response characteristic is found to have a voltage gain of 25 volt-logits. When the plate supply voltage drops by 25% the gain of the amplifier drops to 24 volt-logits. Negative feedback is used to stabilize the amplifier. Compute the closed loop gain, if $\beta = 0.01$, at rated power supply voltage and when the supply voltage drops by 25%.

7.5. In a single stage resistance coupled amplifier of the grounded type, the following data apply: $g_m = 4000\ \mu$mhos; $r_p = 10{,}000$ ohms; $R_L = 6000$ ohms; $C_c = 0.001\ \mu$f; $C_T = 30\ \mu\mu$f; $R_g = 500{,}000$ ohms; $C_{pk} = 4\ \mu\mu$f. Compute the reference gain and the upper and lower cutoff frequencies.

7.6. An unbypassed cathode resistor of 600 ohms is added to the amplifier of problem (7.5). Compute the new values for the reference gain and cutoff frequencies.

7.7. The data given in problem (7.5) apply here to a resistance coupled cathode follower. Compute the reference gain and cutoff frequencies of this circuit.

7.8. Suppose the tubes in the amplifiers of problems (7.5) through (7.7) are replaced by tubes having transconductances of 3500 μmhos. Compute the changes in the reference gains assuming all other parameters remain fixed.

7.9. Calculate the output impedances of the three amplifiers of problems (7.5) through (7.7). Assume mid-frequency operation.

7.10. Design a feedback pair to have the same poles as a shunt peaked amplifier with $m = 0.5$. Assume the same tube throughout and total shunt capacities for all stages of 15 $\mu\mu$f. Let $g_m = 4000$ μmhos; R_L(shunt peaked) $= 2000$ ohms; $r_p = 700{,}000$ ohms; $R_g = 1$ megohm. Calculate G_{12} and A_r.

7.11. Compute the rise time and overshoot of the amplifier of problem (7.10). How do these figures compare with those for the shunt peaked amplifier with $m = 0.5$? What is the reason for the difference?

7.12. Design a maximally flat feedback pair to have an over-all bandwidth of 2 mcps centered about 30 mcps. It may be assumed that $g_{m_1} = g_{m_2} = 4500$ μmhos; $C_T = 15$ $\mu\mu$f. It is also assumed that g_{m_2} is much larger than G_{12}. Compute the values required for G_{12}, R, and L. Find A_r. What possible practical problems might be encountered in building this amplifier?

7.13. Repeat problem (7.12), designing a cascade of a feedback pair and one single tuned stage to have the same response as that of a maximally flat-staggered triple.

7.14. The following characteristic equations are given. Using the Routh-Hurwitz criterion, determine whether these equations represent stable systems. Specify the number of right half plane roots.

(a) $s^4 + 5s^3 + 13s^2 + 19s + 10 = 0.$
(b) $s^3 + s^2 + s + 1 = 0.$
(c) $s^4 + 2s^3 + 4s^2 - 2s - 5 = 0.$
(d) $s^5 + 4s^4 + 7s^3 + 8s^2 + 6s + 4 = 0.$
(e) $s^4 + 2s^3 + s + 2 = 0.$
(f) $s^5 - 9s^3 - 22s^2 - 22s - 8 = 0.$

Chapter 8

TRANSISTOR AMPLIFIERS

Although the transistor is a recent invention, the circuit applications of these little devices have developed with remarkable speed. Several different techniques have been used by various people to analyze the circuits and to derive design formulas. Each method[1] has its advantages and drawbacks. Still another method is presented here. It has particular advantages to be noted later in the analysis and design of transistor amplifiers and feedback oscillators.

It will be shown in this chapter that the voltage gain of a transistor amplifier can be expressed

$$A = \pm g_t' Z_m$$

where g_t' = effective transconductance of the transistor amplifier circuit; Z_m = mutual impedance of the passive network in the equivalent plate circuit. Therefore, as in the study of vacuum tube amplifiers, there are two natural subdivisions of inquiry: (1) factors affecting g_t'; (2) factors affecting Z_m. This chapter is subdivided pretty much along these lines, and closes with some comments regarding cascade connections.

8.1. Establishing the Q Point

A transistor will operate satisfactorily *only* if there is a complete d-c path around both the emitter and collector circuits. This is shown for one simple type of transistor amplifier in figure (8.1). As you can see, the base is grounded, so this is called a *grounded base amplifier*.

The symbolism in figure (8.1) is standard, with terms defined as follows:

$V_{\varepsilon\varepsilon}$ = emitter bias supply voltage.
V_{cc} = collector bias supply voltage.
I_ε = total emitter current $= i_\varepsilon + I_{\varepsilon\varepsilon}$.
I_c = total collector current $= i_c - I_{cc}$.

[1] R. L. Wallace Jr., "Duality as a Guide in Transistor Circuit Design," *Bell System Tech. J.*, vol. 30, April, 1951, pp. 381–418; also R. L. Wallace Jr. and W. J. Pietenpol, "Some Circuit Properties and Applications of *n-p-n* Transistors," *Bell System Tech. J.*, vol. 30, July, 1951, pp. 530–563.

Both the emitter and collector currents consist of two terms: bias or d-c components $I_{\varepsilon\varepsilon}$ and I_{cc}, and variational or signal components i_ε and i_c.

The voltage loop equations around the emitter and collector circuits in the *quiescent*, or no signal, state are

$$V_{\varepsilon\varepsilon} = I_{\varepsilon\varepsilon}R_\varepsilon + V_\varepsilon$$

$$V_{cc} = I_{cc}R_c + V_c$$

or, with the terms rearranged somewhat,

$$V_\varepsilon = V_{\varepsilon\varepsilon} - I_{\varepsilon\varepsilon}R_\varepsilon \tag{8.1}$$

$$V_c = V_{cc} - I_{cc}R_c \tag{8.2}$$

These are the *load line equations* for the emitter and collector circuits;

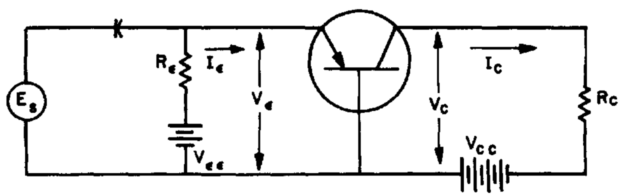


Fig. 8.1. Elementary transistor amplifier; grounded base connection.

they are used to construct the load lines on the transistor input and output characteristics in exactly the same way as the load line is drawn on the static plate characteristics of vacuum tubes.

The transistor manufacturer will usually recommend values for V_c and $I_{\varepsilon\varepsilon}$. These are the *polarizing* potentials and currents and will fix the the quiescent point of the transistor to the recommended position. Actually, two Q points are obtained, one on the output characteristics and one on the input characteristics. However, the specification of V_c and $I_{\varepsilon\varepsilon}$ automatically fixes V_ε and I_{cc}.

Other circuit connections used to obtain the necessary polarizing potentials and currents will be discussed later.

Transistor amplifier operation can be analyzed directly from the load line plots on the static characteristics when all circuit elements are purely resistive. However, the transistor circuit, and especially the load circuit, are rarely simple resistances; analytical treatment is then preferred.

The transistor can be made to operate either in class A or in the switching mode by suitable adjustment of the Q points and signal amplitude. Only class A operation will be covered in this chapter, so that the class A equivalent circuit derived in chapter 1 and shown in figure (1.16b) can be used.

8.2. Types of Transistor Amplifiers

Suppose that the Q point and signal amplitude are adjusted so that a given transistor operates in class A. The equivalent circuit of the

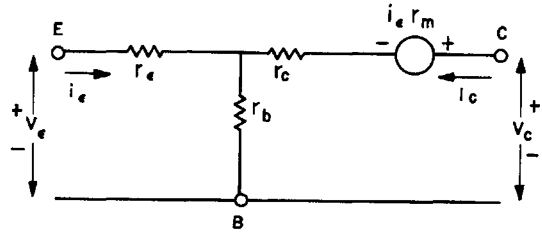


Fig. 8.2. Class A equivalent circuit of a transistor showing assumed voltage polarities and current directions.

transistor then has the form shown in figure (8.2). Some representative values of the four transistor resistances are tabulated below.

Resistance	*Point contact*	*Junction type*
r_ε	250 ohms	25 ohms
r_b	300 ohms	250 ohms
r_c	20 kilo-ohms	10 megohms
r_m	35 kilo-ohms	9.75 meghoms

The equivalent circuit given in figure (8.2) can be used for either n-type or p-type transistors with the currents and voltages specified on the drawing. Of course, in an actual transistor *one* of the currents is always in the wrong direction, so that it is negative on the characteristic curves. The *true* directions of current flow associated with the equivalent circuits are shown in figure (8.3) for both the n-type and p-type transistor.

When single amplifier stages are involved, the direction of current flow actually used in the analysis will not affect the results. However, when more than one stage is involved, it is generally desirable to show the true direction of current flow. A good illustration of this appears in the discussion of a push-pull amplifier in chapter 11.

As far as variational or signal frequency terms are concerned, you can see from these equivalent circuits that the transistor is a three-terminal circuit element. The three terminals are the emitter, base, and collector. Any one of three *signal* electrodes can be used as the reference or *grounded* terminal for signal currents. Therefore, there are three basic amplifier circuit configurations: (1) grounded base amplifier; (2) grounded emitter amplifier; (3) grounded collector amplifier.

When an electrode is said to be *grounded*, the term means grounded for signal frequency components only. The grounded electrode may be

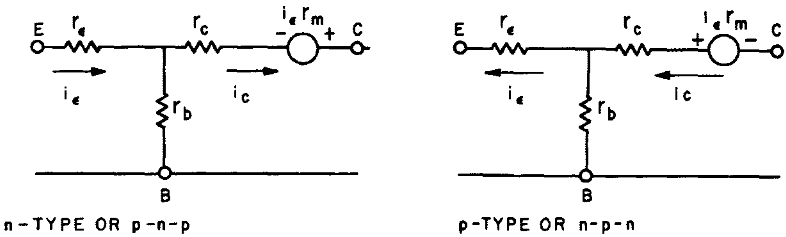


Fig. 8.3. True directions of current flow in the two types of transistors.

at a high d-c potential above the d-c ground.

Each of the three basic amplifier circuits will be discussed in this chapter.

8.3. Interelectrode Capacitances

The class A equivalent circuit of a transistor, showing the emitter and collector capacitances, is given in figure (8.4). Although we will treat

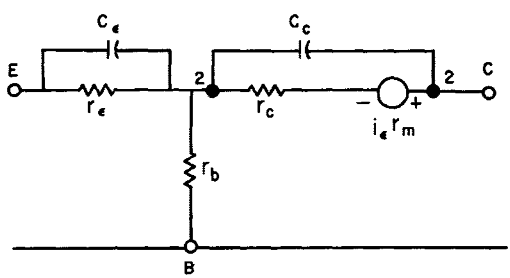


Fig. 8.4. Transistor interelectrode capacitances.

these capacitances as constants, they actually are not; each depends upon the electrode voltage, generally decreasing as the voltage is increased.

The circuit can be simplified. Designate the parallel combination of r_ε and C_ε as z_ε, so that

$$z_\varepsilon = r_\varepsilon \frac{\omega_\varepsilon}{s + \omega_\varepsilon} \quad \text{where} \quad \omega_\varepsilon = \frac{1}{r_\varepsilon C_\varepsilon}$$

Now construct the voltage source equivalent circuit of the collector circuit at the terminals marked 2–2. You can easily show that this circuit consists of a generator and a series impedance as follows:

$$E_{oc} = I_\varepsilon r_m \frac{\omega_c}{s + \omega_c} = I_\varepsilon z_m \tag{8.3}$$

$$Z = z_c = r_c \frac{\omega_c}{s + \omega_c} \tag{8.4}$$

where
$$\omega_c = \frac{1}{r_c C_c} = \textit{collector cut-off} \tag{8.5}$$

Therefore the modified equivalent circuit of the transistor, including

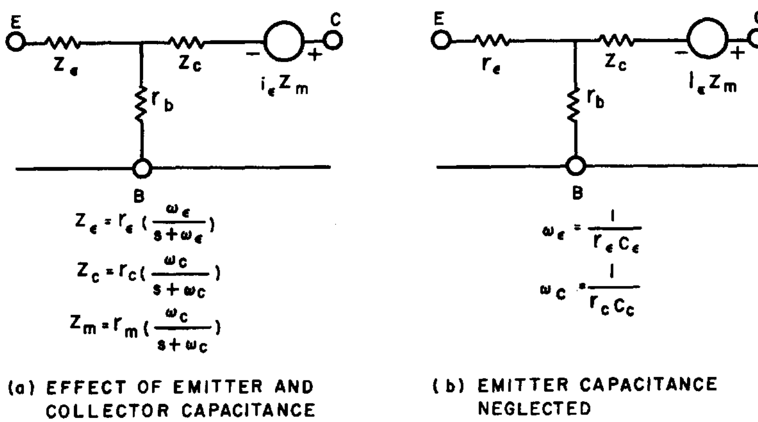


Fig. 8.5. Equivalent circuits including the effect of interelectrode capacitance.

the effects of interelectrode capacitances, has the form shown in figure (8.5a).

Even if C_ε and C_c are nearly equal, ω_ε is always much larger than ω_c because r_ε is always much less than r_c. Hence the collector capacitance becomes important at much lower frequencies than the emitter capacitance, and C_ε is usually neglected. The resulting equivalent circuit is shown in figure (8.5b), and this is used in all subsequent work.

8.4. The Method of Attack

The three basic types of *vacuum tube* amplifiers and several variations were analyzed in chapter 3. In this analysis it was found convenient to postulate an equivalent circuit of the form shown in figure (8.6). As you can see, the amplifier is represented by isolated input and output circuits. The input circuit consists of the signal generator, its internal impedance R_s, and the input impedance Z_i of the amplifier. The output circuit consists of a current source $\pm g'_m E_i$, shunted by an impedance z_p, and this is connected to an arbitrary three-terminal load circuit of input impedance Z_L. Terms are defined as follows:

$\pm g'_m =$ effective transconductance of the amplifier.

$z_p =$ parallel combination of r_p and C_{pk}(or C_o for a pentode).

$Z_L =$ input impedance of the load circuit.

The total *passive* network in the ouput circuit, consisting of z_p and the associated load, was always kept the same regardless of the type of

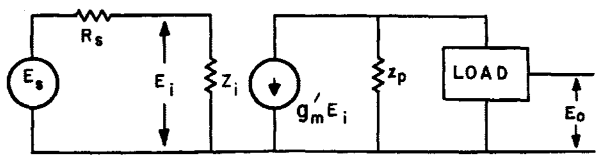


Fig. 8.6. General equivalent circuit of a vacuum tube voltage amplifier as derived in chapter 3.

amplifier under study. In other words, this passive network remained the same irrespective of whether the amplifier was of the grounded cathode, grounded grid, or grounded plate type. Variations in the characteristics of these amplifiers were accommodated by using different values for g'_m and Z_i.

The particular advantage of this technique is that all vacuum tube voltage amplifiers then have stage gain functions of the form

$$A = \pm g'_m Z_m = \frac{E_o}{E_i}$$

where $Z_m =$ mutual impedance of the passive network composed of z_p and the connected load. This reduces the study of vacuum tube voltage amplifiers to an examination of the (1) factors affecting g'_m; (2) factors affecting Z_m.

It should be understood that the preceding equation is the gain function for the *output circuit only*. There is a voltage dividing network in the input circuit, so that

$$A\ (\text{input}) = \frac{E_i}{E_s} = \frac{Z_i}{R_s + Z_i}\quad .$$

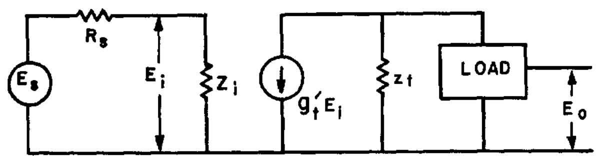
Therefore the *over-all* gain function of the stage would be

$$A\ (\text{over-all}) = A\ (\text{input}) \times A\ (\text{output}) = \frac{E_o}{E_s}$$

$$= \left(\frac{Z_i}{R_s + Z_i}\right) g_m' Z_m$$

It may be necessary on occasion to use this relationship.

There is a serious problem involved here. The use of the over-all gain function makes the amplifier characteristics dependent upon the signal source impedance. This makes the comparison of various amplifier types difficult unless a particular signal source impedance is specified.



Z_L = INPUT IMPEDANCE OF LOAD CIRCUIT

Fig. 8.7. Proposed transistor equivalent circuit.

The problem is overcome in vacuum tube circuits by either of two methods:

(1) The signal source is assumed to have zero internal impedance.

(2) The input circuit is made a part of the output circuit of the driving system and treated as a separate amplifier stage.

The second method is the best and is most generally used. It is more adaptable to the study of transistor amplifiers. Thus it is pretty much general practice to use the gain function of the output circuit only when discussing the response characteristics of amplifiers. The effect of the input circuit is then accommodated by treating it as another stage in cascade with the output circuit.

Exactly the same method of analysis can be applied to transistor voltage amplifiers. An equivalent circuit of the form shown in figure (8.7) can be proposed so that it has exactly the same form as the vacuum tube equivalent circuit. Then it is necessary only to compute the values for the constants of this circuit for the various transistor amplifier configurations. The voltage gain of the output circuit for *any* transistor amplifier is then

$$A = \pm g_t' Z_m$$

where g_t' is the effective mutual transconductance of the transistor amplifier; Z_m is the mutual impedance of the network composed of z_t and the connected load.

You can see that g_t' for a transistor corresponds to g_m for a vacuum tube. Similarly, g_t (without the prime) should correspond to the g_m (without the prime). Here g_m is the mutual transconductance of the vacuum tube; this is exactly equal to the effective transconductance of a nondegenerative, grounded cathode amplifier. This is the reference case for vacuum amplifiers.

A corresponding reference case exists for transistor amplifiers. The cathode is the emitter for vacuum tubes. Thus the transistor analogue of a grounded cathode vacuum tube amplifier would be a grounded

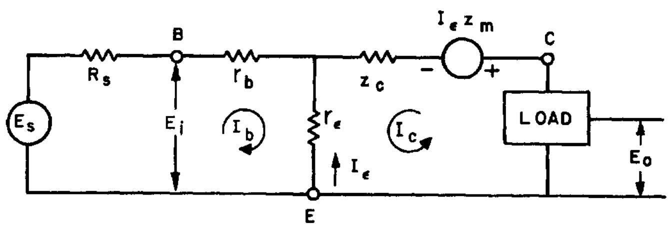


Fig. 8.8. Grounded emitter amplifier (conventional representation).

emitter transistor amplifier. Therefore g_t will signify the actual mutual transconductance of the transistor itself, and this must be numerically equal to the effective transconductance of a nondegenerative grounded emitter amplifier.

In the vacuum tube equivalent circuit, z_p is the output impedance of a nondegenerative grounded cathode amplifier. Thus, by analogy, z_t must be the output impedance of a nondegenerative grounded emitter amplifier.

Here Z_i will always signify the actual input impedance of the amplifier circuit.

On the strength of this discussion you can see that the nondegenerative grounded emitter amplifier is the reference case in the study of transistor amplifiers. This circuit is covered in the next section.

8.5. Grounded Emitter Amplifier, No Feedback

The conventional equivalent circuit of a grounded emitter amplifier is shown in figure (8.8). The voltage loop equation around the input circuit is

$$E_s = I_b(R_s + r_b + r_\epsilon) + I_c r_\epsilon$$

The presence of the $I_c r_\varepsilon$ term in this equation clearly indicates the existence of feedback. In the absence of feedback there is no coupling from the output circuit into the input circuit. Therefore feedback can be prevented in this case if the emitter resistance r_ε is reduced to zero.

The zero feedback condition obtained when r_ε is zero is never reached in practical cases; there are some cases in which it will be approached rather closely. In any case, it will be our *reference circuit* just as the nondegenerative ($Z_k = 0$) grounded cathode amplifier was the reference case for vacuum tube amplifiers.

In the hypothetical zero feedback case, the conventional equivalent circuit of the amplifier has the form shown in figure (8.9a). This is

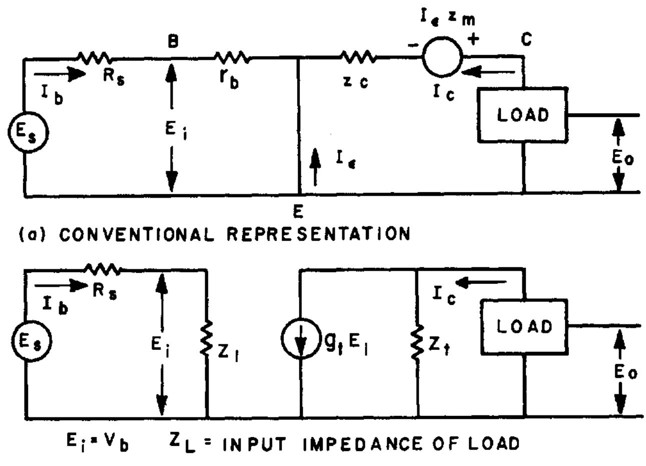


Fig. 8.9. Equivalent circuits for a grounded emitter amplifier without feedback: $r_\varepsilon = 0$.

obtained from figure (8.8) by replacing r_ε with a short circuit. The *proposed* equivalent circuit is given in figure (8.9b). We must now derive formulas for the constants g_t, z_t, and Z_i of the proposed circuit in terms of the constants of the conventional circuit.

The voltage loop equation around the input circuit of the conventional equivalent circuit in figure (8.9a) is

$$V_b = I_b r_b$$

so that the input impedance is

$$Z_i = \frac{V_b}{I_b} = r_b \tag{8.6}$$

Now compute the output impedance z . To do this, disconnect the load circuit from terminals *C-E* of the conventional equivalent circuit; now connect a generator of voltage E_x and zero internal impedance across the same terminals. Then replace the input signal source E_s by a short circuit. With the circuit connected this way the loop equation around the input circuit is $0 = I_b'(R_s + r_b)$, so that $I_b' = 0$. The loop equation around the output circuit is now

$$E_x = I_c'z_c + I_e'z_m$$

However, it is clear from figure (8.9a) that

$$I_e' = -(I_b' + I_c')$$

But, because I_b' is zero, then $I_e' = -I_c'$ and the first loop equation is

$$E_x = I_c'(z_c - z_m)$$

Therefore the output impedance is

$$z_t = \frac{E_x}{I_c'} = z_c - z_m \tag{8.7}$$

It was previously shown that

$$z_c = r_c \frac{\omega_c}{s + \omega_c} \quad \text{and} \quad z_m = r_m \frac{\omega_c}{s + \omega_c}$$

so that the output impedance is

$$z_t = (r_c - r_m)\left(\frac{\omega_c}{s + \omega_c}\right) = r_c\left(1 - \frac{r_m}{r_c}\right)\left(\frac{\omega_c}{s + \omega_c}\right)$$

The current amplification factor of a transistor is

$$\alpha = \frac{r_m + r_b}{r_c + r_b}$$

In nearly all transistors both r_m and r_c are much larger than r_b, so that $\alpha \doteq r_m/r_c$. Therefore the output impedance can be written

$$z_t \doteq r_c(1 - \alpha)\frac{\omega_c}{s + \omega_c} \tag{8.8}$$

This impedance is the parallel combination of the following circuit elements:

$$r_t \doteq r_c(1 - \alpha) \tag{8.9}$$

$$C_t = \frac{C_c}{1 - \alpha} \tag{8.10}$$

All that remains now is to evaluate the mutual transconductance of the transistor. From figure (8.9b) you can see that a current $g_t V_b$ will flow if the terminals C-E are short circuited. That is

$$I_{sc} = g_t V_b$$

or
$$g_t = \frac{I_{sc}}{V_b} \tag{8.11}$$

Therefore to evaluate g it is necessary only to compute the ratio of I_{sc} to V_b in the conventional equivalent circuit of figure (8.9a) when terminals C-E are short circuited. When the terminals are short circuited, the two loop equations are

$$V_b = I_b r_b \tag{8.12}$$
$$0 = I_{sc} z_b + I_\varepsilon z_m$$

However,
$$I_\varepsilon = -(I_b + I_{sc})$$

so that the second loop equation is

$$0 = I_{sc}(z_c - z_m) - I_b z_m = I_{sc} z_t - I_b z_m$$

Therefore
$$I_b = I_{sc} \frac{z_t}{z_m}$$

Substitute this into equation (8.12).

$$V_b = I_{sc} \frac{z_t r_b}{z_m}$$

Therefore the transconductance of the transistor is

$$g_t = \frac{I_{sc}}{V_b} = \frac{z_m}{z_t r_b} = \frac{r_m}{r_t r_b} \tag{8.13}$$

or, in alternative and useful forms,

$$g_t = \frac{1}{r_b} \cdot \frac{\alpha}{1 - \alpha} = \frac{r_m}{r_b(r_c - r_m)} \tag{8.14}$$

Representative values for g_t are easily calculated. For example, a typical junction transistor might have $r_m = 9.75$ megohms, $r_c = 10$ megohms, $r_b = 250$ ohms. Hence

$$g_t = \frac{9.75}{(250)\,(0.25)} = 156{,}000\ \mu\text{mhos}$$

This is an extremely large figure when compared with values of g_m in ordinary vacuum tubes. However, we will see later that transistors are very sensitive to degenerative effects, so that actual values of transconductance will not be quite so large.

A point contact transistor might have the following constants: $r_m = 35{,}000$, $r_c = 20{,}000$ ohms, $r_b = 300$ ohms, so that

$$g_t = \frac{-35}{(300)\,(15)} = -7777\ \mu\text{mhos}$$

This is of the same order of magnitude as the g_m of vacuum tubes.

Any sizable reduction in the base resistance r_b will produce a corresponding increase in transconductance. Low values for r_b are characteristic of new transistor designs. Superficially, it might appear desirable to make α nearly unity because this would make g_t infinite. However, this advantage is offset because the shunting resistance $r_t = r_c(1 - \alpha)$ would become zero.

If Z_m designates the mutual impedance between the g_tV_b generator and the output terminals of the amplifier, the output voltage is $E_o = -g_tV_bZ_m$ and the voltage gain is $A = E_o/V_b = -g_tZ_m$. The presence of the minus sign signifies phase inversion through the amplifier. In some cases g_t may itself be negative so that the gain function is positive. When this is true, there is no phase inversion.

8.6. Grounded Emitter Amplifier

The circuit discussed in the preceding section is a purely hypothetical case in which the emitter resistance r_ε is assumed to be zero. The practical case where r_ε is *not* zero will be covered here.

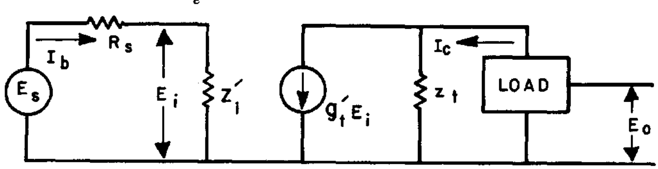


Fig. 8.10. Proposed equivalent circuit for a practical grounded emitter amplifier.

The proposed equivalent circuit of a practical grounded emitter amplifier is shown in figure (8.10). Note that the passive impedance network in the output circuit, which is composed of z_t and the connected load, is exactly the same as that used in the preceding case so that Z_m remains unchanged. The circuit differs from the previous case because a new, or *effective*, transconductance g_t' is specified together with a new value for the input impedance Z_i'.

The voltage gain of the amplifier is clearly $A = -g_t'Z_m$. We need to evaluate only g_t' and Z_i'. Both of these factors can be derived from the voltage loop equations of the conventional tee equivalent circuit previously shown in figure (8.8). We will evaluate g_t' first and then Z_i'.

The voltage loop equation around the second loop of the conventional the equivalent circuit of figure (8.8) is

$$0 = I_c(r_\varepsilon + z_c + Z_L) + I_\varepsilon z_m + I_b r_\varepsilon$$

However, it is clear from the circuit diagram that $I_\varepsilon = -(I_b + I_c)$, so that the preceding loop equation becomes

$$0 = I_c(r_\varepsilon + z_c - z_m + Z_L) + I_b(r_\varepsilon - z_m) \tag{8.15}$$

Thus the loop current I_b is

$$I_b = I_c \frac{r_\varepsilon + z_c - z_m + Z_L}{z_m - r_\varepsilon} \tag{8.16}$$

The voltage loop equation around the input circuit of figure (8.8) is

$$E_i = I_b(r_b + r_\varepsilon) + I_c r_\varepsilon \tag{8.17}$$

Substitute equation (8.16) into (8.17).

$$E_i = I_c \left[r_\varepsilon + \frac{(r_\varepsilon + z_c - z_m + Z_L)(r_b + r_\varepsilon)}{z_m - r_\varepsilon} \right]$$

Obtain the common denominator, expand the numerator, and cancel the $z_m r_\varepsilon$ and r_ε^2 terms. Then solve for I_c, and the result should be

$$I_c = E_i \frac{z_m - r_\varepsilon}{r_b(z_t + Z_L) + r_\varepsilon(r_b + z_c + Z_L)}$$

It is usually possible to assume that z_m and z_c are much larger than r_ε and r_b. Hence the preceding equation can be closely approximated by

$$I_c \doteq E_i \frac{z_m}{r_b(z_t + Z_L) + r_\varepsilon(z_c + Z_L)} \tag{8.18}$$

If the proposed equivalent circuit shown in figure (8.10) is actually equivalent to the conventional tee equivalent, the current I_c must be equal to that given in equation (8.18). In the proposed circuit

$$I_c = g_t' E_i \frac{z_t}{z_t + Z_L}$$

Set this equal to equation (8.18) and solve for g_t'.

$$g_t' = \frac{z_m / z_t}{r_b + r_\varepsilon (z_c + Z_L)/(z_t + Z_L)}$$

Divide numerator and denominator through by r_b and then let

$$g_t = \frac{z_m}{r_b z_t}$$

Hence the effective transconductance of a practical grounded emitter amplifier is

$$g_t' \doteq \frac{g_t}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{8.19}$$

This is the desired result.

The effect of the emitter resistance on the amplifier transconductance can now be illustrated in a general way. Assume that Z_L is a pure resistance of 20,000 ohms. The transistor is a junction type having $r_\varepsilon = 25$ ohms; $r_m = 9.75$ megohms; $r_b = 250$ ohms; $r_c = 10$ megohms. Therefore, at low frequencies where the collector capacitances can be neglected, $g_t = 156{,}000$ μmhos and $g_t' = g_t/(1 + 3.7) = 33{,}190$ μmhos. This illustrates the important degenerative effect present in junction transistors even when the emitter resistance is quite small. This value of transconductance is larger than that obtained with vacuum tubes. The reduction could be minimized by making Z_L of the same order of magnitude as z_c. However, this would require such large collector supply voltages that it is not a practical suggestion.

A slightly different situation prevails with point contact transistors. In these devices r_m is larger than r_c, so that z_t is negative. As a result, the emitter circuit feedback can be either degenerative or regenerative depending upon the relationship between z_t and Z_L. For example, if ω is much less than ω_c and if $R_L = 20{,}000$ ohms, $r_m = 35{,}000$ ohms, $r_c = 20{,}000$ ohms, $r_\varepsilon = 250$ ohms, $r_b = 300$ ohms, then $g_t = -7777$ μmhos, and $g_t' = g_t/(1 + 6.67) = -1014$ μmhos. If R_L had been 30,000 ohms, the effective transconductance would be -2047 μmhos. Or if R_L were 10,000 ohms the effective transconductance would become $+1944$ μmhos. The first two cases are degenerative, while the last one is regenerative.

The input impedance is easily calculated. From equation (8.16) we can write

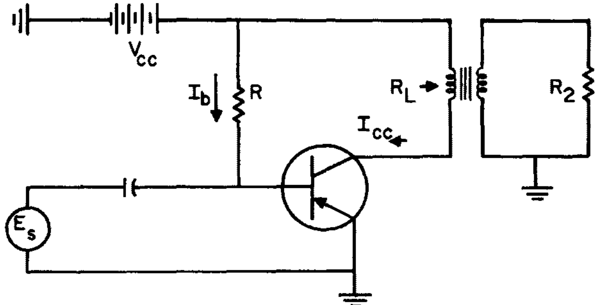
$$I_c = I_b \frac{z_m - r_\varepsilon}{r_\varepsilon + z_c - z_m + Z_L} \tag{8.20}$$

Substitute this into equation (8.17) to obtain

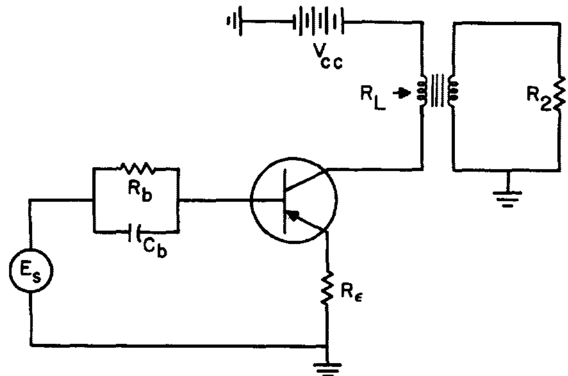
$$E_i = I_b \left[r_b + r_\varepsilon + \frac{r_\varepsilon(z_m - r_\varepsilon)}{r_\varepsilon + z_t + Z_L} \right]$$

where $z_t = z_c - z_m$. Thus after some manipulation, this equation can be used to solve for the input impedance as

$$Z_i' = \frac{E_i}{I_b} = r_b\left[1 + \frac{r_\varepsilon(z_c + Z_L)}{r_b(r_\varepsilon + z_t + Z_L)}\right] \tag{8.21}$$



(a) ONE METHOD OF APPLYING POLARIZING POTENTIALS



(b) ALTERNATIVE BIASING ARRANGEMENT; R_ϵ AND R_b–C_b ACT TO REDUCE I_{cc}

Fig. 8.11. Circuit connections for a grounded emitter amplifier.

Of course, $Z_i = r_b$, so that

$$Z_i' = Z_i\left[1 + \frac{r_\varepsilon(z_c + Z_L)}{r_b(r_\varepsilon + z_t + Z_L)}\right] \tag{8.22}$$

This change in the input impedance of the amplifier can be expressed more simply because the bracketed term is exactly equal to the denominator of the effective transconductance. Therefore

$$\frac{g_t}{g_t'} = 1 + \frac{r_\varepsilon(z_c + Z_L)}{r_b(r_\varepsilon + z_t + Z_L)}$$

and as a result,

$$Z_i' = r_b \frac{g_t}{g_t'} \tag{8.23}$$

It was shown previously that g_t' can be negative at times when point contact or hook collector junction transistors are involved. This would make Z_i' negative, as you can see from equation (8.23), and the resulting circuit is an unstable positive feedback amplifier. Thus, to make sure that Z_i' is positive, g_t' must have the same sign as g_t, or

$$1 + \frac{r_\varepsilon(z_c + Z_L)}{r_b(r_\varepsilon + z_t + Z_L)} > 0$$

Rearrange terms and we find that Z_i' will be positive if

$$\left(1 + \frac{Z_L}{z_c}\right)\left(1 + \frac{r_\varepsilon}{r_b}\right) > \alpha$$

Therefore as long as the current amplification factor of the transistor is unity or less, the input impedance of a grounded emitter amplifier will always be positive.

Two circuit diagrams of grounded emitter amplifiers are shown in figure (8.11) to illustrate some of the biasing arrangements in use. The collector current may be excessive in the circuit of figure (8.11a). Two methods of reducing this current are shown in figure (8.11b). One method requires increased degeneration by increasing the total emitter resistance with an external resistance R_ε. This may be all right with junction transistors, but trouble with regeneration might result in point contact types.

8.7. Grounded Base Amplifier

The conventional and proposed equivalent circuits of a grounded base amplifier with an arbitrary three-terminal load impedance are shown in figure (8.12). If the two circuits are actually equivalent, the same emitter and collector currents will flow in both circuits. As before, this provides the basis for the derivation of the equations for g_t' and Z_i'. You will observe that the passive network in the output circuit remains unchanged, so that the voltage amplification is still $A = -g_t'Z_m$.

The two loop equations for the conventional equivalent circuit in figure (8.12a) are

$$E_i = I_\varepsilon(r_\varepsilon + r_b) + I_c r_b$$
$$0 = I_c(r_b + z_c + Z_L) + I_\varepsilon(r_b + z_m)$$

It is nearly always valid to assume that

$$z_m \gg r_b \quad \text{and} \quad z_c \gg r_b$$

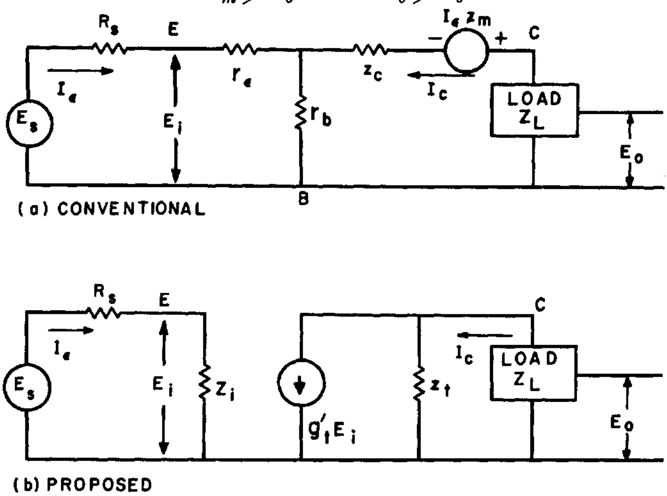


Fig. 8.12. Equivalent circuits for a grounded base amplifier.

so that the second loop equation is approximately

$$0 \doteq I_c(z_c + Z_L) + I_\varepsilon z_m \tag{8.24}$$

Therefore the emitter current is

$$I_\varepsilon \doteq -I_c \frac{z_c + Z_L}{z_m} \tag{8.25}$$

Substitute this into the equation for the first loop of the conventional tee equivalent circuit.

$$E_i \doteq I_c \left[r_b - \frac{(r_\varepsilon + r_b)(z_c + Z_L)}{z_m} \right]$$

Obtain the common denominator, expand the numerator, and collect terms as follows:

$$E_i \doteq -I_c \frac{r_b(z_t + Z_L) + r_\varepsilon(z_c + Z_L)}{z_m}$$

where $(z_c - z_m)$ has been replaced by z_t. Hence the collector current is

$$I_c \doteq -E_i \frac{z_m}{r_b(z_t + Z_L) + r_\varepsilon(z_c + Z_L)}$$

The collector current in the proposed equivalent circuit is

$$I_c \doteq g_t' E_i \frac{z_t}{z_t + Z_L}$$

Set this expression equal to the previous equation for I_c and solve for g_t'. Divide through by $r_b(z_t + Z_L)$. Then because

$$g_t = \frac{z_m}{r_b z_t}$$

the equation for the effective transconductance of the grounded base amplifier becomes

$$g_t' \doteq -\frac{g_t}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{8.26}$$

Except for the minus sign, the effective transconductance of a grounded base amplifier is exactly equal to the effective transconductance of a grounded emitter amplifier.

The voltage gain of the amplifier remains

$$A = -g_t' Z_m$$

or

$$A \doteq +\frac{g_t Z_m}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{8.27}$$

Contrary to the operation of a grounded emitter amplifier, there is no phase inversion through a grounded base amplifier except in special cases where the current amplification is greater than 1, and the relative values of other parameters are such that the effective transconductance becomes negative.

The input impedance is calculated from the two loop equations. Solve equation (8.25) for I_c and substitute the result into the equation for the first loop around the conventional tee equivalent circuit. Then compute the input impedance to be

$$Z_i' = \frac{E_i}{I_\varepsilon} \doteq r_\varepsilon\left(1 + \frac{r_b}{r_\varepsilon}\cdot\frac{z_t + Z_L}{z_c + Z_L}\right) \tag{8.28}$$

The second term inside the main bracket is the reciprocal of the second term in the denominator of the equation for the effective transconductance given in (8.27). Hence the input can also be expressed as

$$Z_i' = r_\varepsilon \frac{g_t}{g_t - g_t'} \tag{8.29}$$

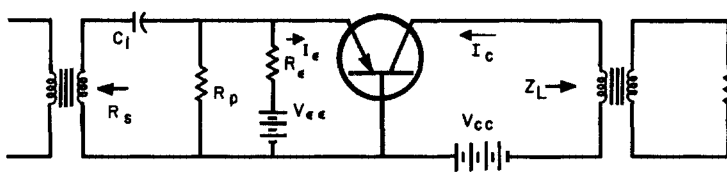
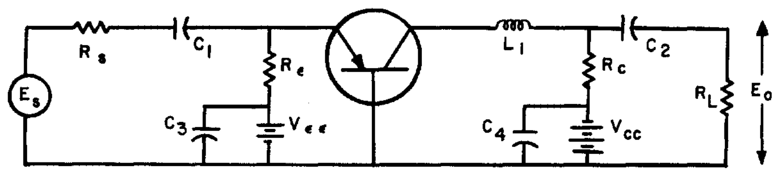


Fig. 8.13. Transformer coupled, grounded base transistor amplifier. (a) Emitter circuit is resistance shunt fed. (b) Collector circuit is transformer fed; could use resistance or choke shunt feed if desired. (c) C_1 is a d-c blocking capacitor. (d) R_ε is the emitter dropping resistor. (e) R_p is a protective resistor for C_1; keeps the voltage across C_1 to safe values if the transistor is removed.



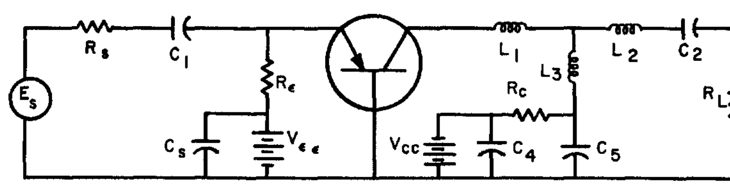


Fig. 8.14. Tuned, bandpass, grounded base amplifiers. In part (a), C_1 and C_2 are blocking capacitors, C_3, C_4, and C_5 are by-pass capacitors, R_ε and R_c are dropping resistors, and L_1, L_2, and L_3 are tuning inductors.

The input impedance will be positive as long as

$$g_t - g_t' > 0$$

From this you can readily prove that

$$\left(1 + \frac{r_\varepsilon}{r_b}\right)\left(1 + \frac{Z_L}{z_c}\right) > \alpha \tag{8.30}$$

for the input impedance to be positive. This is exactly the same inequality as that derived for the grounded emitter amplifier.

Some representative grounded base amplifiers are shown in figures (8.13) and (8.14), together with operating notes of practical interest.

8.8. Grounded Collector Amplifier

The procedures for deriving the equations for g_t' and Z_i' of the grounded collector amplifier are exactly the same as those used for the grounded base connection. The necessary equivalent circuits are given

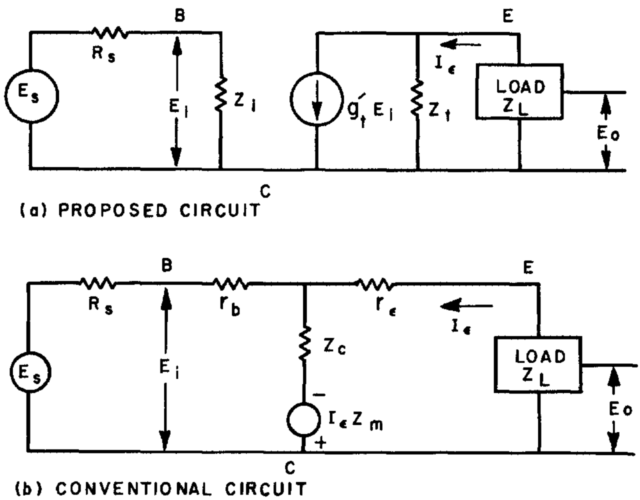


Fig. 8.15. Equivalent circuits for a grounded collector amplifier.

in figure (8.15). It is left as an exercise for the reader to prove that the following equations are correct if the usual approximations are used;

$$g_t' \doteq -\frac{g_t}{\alpha + g_t\left(\dfrac{r_\varepsilon + Z_L}{z_t + Z_L}\right) z_t} \tag{8.31}$$

$$Z_i' \doteq r_b + \frac{z_c(r_\varepsilon + Z_L)}{r_\varepsilon + z_t + Z_L} \tag{8.32}$$

or

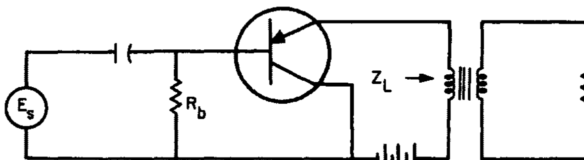
$$Z_i' \doteq -\frac{1}{(1-\alpha)g_t'} \tag{8.33}$$

For Z_i' to be positive,

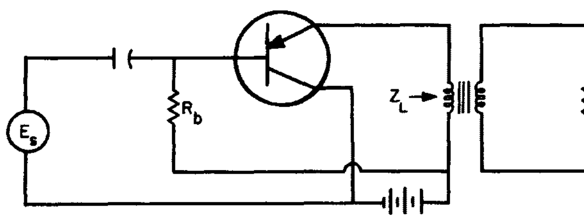
$$\left[1 + \frac{r_\varepsilon + Z_L}{r_b}\left(1 + \frac{r_b}{z_c}\right)\right] > \alpha$$

The voltage gain of a grounded collector amplifier is always less than 1. This is *not* immediately obvious from the amplifier gain function, which is

$$A = + \frac{g_t Z_m}{\alpha + g_t\left(\frac{r_\varepsilon + Z_L}{z_t + Z_L}\right) z_t}$$



(a) LARGER COLLECTOR CURRENT



(b) SMALLER COLLECTOR CURRENT

Fig. 8.16. Grounded collector amplifiers showing biasing arrangements.

However, if we expand the numerator and let $Z_{in} = z_t Z_L/(z_t + Z_L)$ = input impedance of the passive network, then the equation for the voltage gain can be written

$$A = \frac{g_t Z_m}{\alpha + g_t Z_{in} + g_t r_\varepsilon z_t/(z_t + Z_L)}$$

or
$$A = \frac{g_t Z_m}{\alpha + g_t Z_{in}(1 + r_\varepsilon/Z_L)} \tag{8.34}$$

If the load circuit is a *two-terminal network*, then $Z_m = Z_{in}$

and
$$A = \frac{g_t Z_{in}}{\alpha + g_t Z_{in}(1 + r_\varepsilon/Z_L)}$$

From this you can easily see that the voltage amplification will always be less than unity if g_t is positive.

Some representative grounded collector amplifier circuits are given in figure (8.16) to illustrate biasing arrangements.

8.9. Comparison of Amplifier Types

The results of all the preceding derivational work are summarized in table 11; the closest vacuum tube amplifier analogue of each type of transistor amplifier is also given.

TABLE 11

TRANSISTOR AMPLIFIER CHARACTERISTICS

Transistor amplifier	Effective transconductance	Input impedance	Analogous vacuum tube
Grounded emitter $r_\varepsilon = 0$	$g_t = \dfrac{\alpha}{r_b(1-\alpha)}$	r_b	Grounded cathode $Z_k = 0$
Grounded emitter $r_\varepsilon \neq 0$	$\dfrac{g_t}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)}$	$r_b \dfrac{g_t}{g_t'}$	Grounded cathode $Z_k \neq 0$
Grounded base	$\dfrac{-g_t}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)}$	$r_\varepsilon \dfrac{g_t}{g_t - g_t'}$	Grounded grid
Grounded collector	$\dfrac{-g_t}{\alpha + g_t\left(\dfrac{r_\varepsilon + Z_L}{z_t + Z_L}\right) z_t}$	$\dfrac{1}{(1-\alpha)g_t'}$	Grounded plate, or cathode follower

8.10. Resistance Coupled Amplifier, Reference Case

Single stage transistor amplifiers may be operated with pure resistance loads or with a conventional *RC* coupling network. We will discuss the *RC* coupled case here because the pure resistance load is just a special case of *RC* coupling.

The equivalent circuit of a resistance coupled transistor amplifier of any type is shown in figure (8.17). The output capacitance of the transistor is $C_c/(1-\alpha)$; this was proved in equation (8.10). The distributed wiring capacitance is designated as C_w, and the input capacitance of the loading circuit as C_i. Then C_b is the usual d-c blocking capacitor. The type of amplifier used is left unspecified simply by using the general symbol g_t' for the effective transconductance.

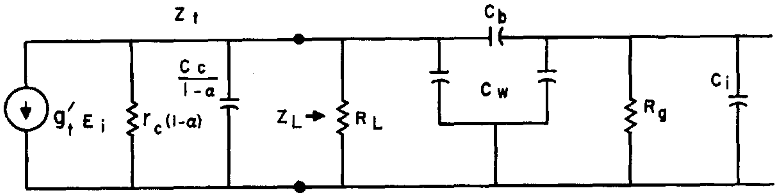


Fig. 8.17. Equivalent circuit of a resistance coupled transistor amplifier.

This circuit is too complex for convenient analysis. So, as in the corresponding vacuum tube amplifier, three separate and special cases are defined as follows:

(1) Low frequency case. The frequency is so low that the reactances of the shunt capacitances are extremely large in comparison with the shunt resistances.

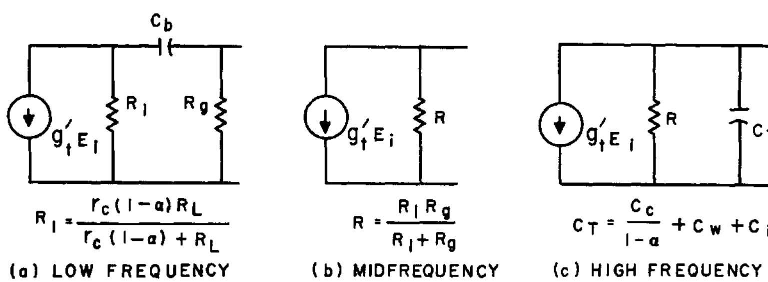


Fig. 8.18. Equivalent circuits for resistance coupled transistor amplifiers.

(2) High frequency case. The frequency is so high that the reactance of the coupling capacitor C_b is virtually zero compared with the resistance of R_g.

(3) Mid-frequency case. The frequency falls midway between the previous extremes, so that all shunt capacitors are virtual open circuits and the coupling capacitor is a virtual short circuit.

The three equivalent circuits corresponding to these three frequency ranges are shown in figure (8.18). As you can see, they have precisely the same forms as the corresponding circuits for vacuum tube amplifiers.

To press the analogy to vacuum tubes a little further, consider the grounded emitter transistor amplifier without feedback, so that $g_t' = g_t$. Therefore the gain functions for the three circuits of figure (8.18) are exactly the same as those obtained for the resistance coupled vacuum tube amplifier in chapter 4, with g_m replaced by g_t. That is,

$$A\,(\text{mid}) = -g_t R = -A_r \tag{8.35}$$

$$A\,(\text{low}) = -A_r \frac{s}{s + \omega_1} \tag{8.36}$$

$$A\,(\text{high}) = -A_r \frac{\omega_2}{s + \omega_2} \tag{8.37}$$

where ω_1 = lower cutoff frequency

$$= \frac{1}{(R_1 + R_g)C_c} \tag{8.38}$$

$$\omega_2 = \text{upper cutoff frequency} = \frac{1}{RC_T} \tag{8.39}$$

Therefore the response characteristics of a resistance coupled grounded emitter amplifier without feedback are exactly the same as those of a nondegenerative grounded cathode vacuum tube amplifier. All data on rise time, sag, and figure of merit are computed from the formulas given in chapter 4.

The high frequency figure of merit of this amplifier is

$$F_a = A_r\omega_2 = \frac{g_t}{C_T} \tag{8.40}$$

This has exactly the same form as the figure of merit of vacuum tube amplifiers.

As in the case of vacuum tubes, we can define a figure of merit for transistors as follows:

$$F_t = \frac{g_t}{C_t}$$

where C_t = transistor capacitance = $C_c/(1 - \alpha)$. Hence the figure of merit of a transistor is

$$F_t = \frac{g_t(1 - \alpha)}{C_c} = \frac{\alpha}{r_b C_c} \tag{8.41}$$

This is a useful criterion for comparing transistors. Representative values for F_t will fall in the region around 500×10^6 radians/sec or more, making them the same order of magnitude as vacuum tube figures of merit.

8.11. Practical Resistance Coupled Amplifiers

The response characteristics just derived are hypothetical because it was assumed that the emitter resistance r_ε was zero. When the emitter resistance is considered, a new set of response characteristics are obtained, just as cathode degeneration alters the response of a grounded cathode vacuum tube amplifier.

For a *practical* grounded emitter amplifier, the general voltage amplification is

$$A' = -\frac{g_t Z_m}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{8.42}$$

In the mid-frequency case, you can see from figure (8.17b) that

$$Z_m = R; \qquad z_c = r_c$$

$$Z_L = R_2 = R_L R_g/(R_L + R_g); \qquad z_t = r_c(1 - \alpha) = r_t.$$

Thus the mid-frequency voltage gain is

$$A_r' = -\frac{A_r}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{r_c + R_2}{r_t + R_2}\right)} \tag{8.43}$$

where $A_r = g_t R$.

In the *high frequency* case, you can show from figure (8.18c) that

$$Z_L = R_2 \frac{\omega_L}{s + \omega_L} \quad \text{where} \quad \omega_\ell = \frac{1}{R_2(C_w + C_i)}$$

$$Z_m = \frac{Z_L z_t}{Z_L + z_t}; \qquad \omega_c = \frac{1}{r_c C_c}$$

$$z_t = z_c(1 - \alpha); \qquad z_c = r_c \frac{\omega_c}{s + \omega_c}$$

Substitute these relationships into the general gain equation given in (8.42). Multiply numerator and denominator through by $r_b(z_t + Z_L)$. Then let

$$g_t = \frac{r_m}{r_b r_c(1 - \alpha)}$$

Finally, multiply numerator and denominator through by $(s + \omega_c) \times (s + \omega_\ell)$. Eventually the final result can be expressed as follows:

$$A' = -\frac{r_m R_2 \omega_\ell \omega_c}{\omega_c \omega_\ell[(r_\varepsilon + r_b)(r_c + R_2) - \alpha r_c r_b] + s[(r_\varepsilon + r_b)(r_c \omega_c + R_2 \omega_\ell) - \alpha r_c r_b \omega_c]}$$

Divide the numerator and denominator through by the first term in the denominator. The result has the form

$$A'\text{ (high)} = -A_r' \frac{1}{1 + s/\omega_H} = -A_r' \frac{\omega_H}{s + \omega_H} \tag{8.44}$$

where
$$A_r' = \frac{r_m R_2}{(r_\varepsilon + r_b)(r_c + R_2) - \alpha r_c r_b} \tag{8.45}$$

$$\omega_H = \frac{\omega_c \omega_\ell [(r_\varepsilon + r_b)(r_c + R_2) - \alpha r_c r_b]}{(r_\varepsilon + r_b)(r_c \omega_c + R_2 \omega_\ell) - \alpha r_c r_b \omega_c} \tag{8.46}$$

$$= \text{upper cutoff frequency}$$

It is fairly easily proven that the reference gain given in equation (8.45) is exactly equal to that computed for the mid-frequency case.

The simplification of the equation for the upper cutoff frequency is a rather involved algebraic manipulation. First, the numerator of the equation for ω_H contains a term of exactly the same form as the denominator of the mid-frequency gain given in (8.45). Therefore ω_H can be expressed

$$\omega_H = \frac{r_m R_2}{A_r'} \cdot \frac{\omega_c \omega_\ell}{(r_\varepsilon + r_b)(r_c \omega_c + R_2 \omega_\ell) - \alpha r_c r_b \omega_c}$$

Divide numerator and denominator through by $\omega_c \omega_\ell$.

$$\omega_H = \frac{r_m R_2}{A_r'} \cdot \frac{1}{(r_\varepsilon + r_b)(r_c/\omega_\ell + R_2/\omega_c) - \alpha r_c r_b/\omega_\ell} \tag{8.47}$$

where
$$\omega_\ell = \frac{1}{R_2(C_w + C_i)}$$

The upper cutoff frequency in the absence of feedback is

$$\omega_2 = \frac{1}{RC_T} = \frac{R_2 + r_t}{R_2 r_t [C_c/(1 - \alpha) + C_w + C_i]}$$

$$= \frac{R_2 + r_t}{R_2/\omega_c + r_t/\omega_\ell}$$

So
$$\frac{R_2}{\omega_c} + \frac{r_t}{\omega_\ell} = \frac{R_2 + r_t}{\omega_2} \quad \text{or} \quad \frac{R_2}{\omega_c} + \frac{r_t}{\omega_\ell} = \frac{R_2 + r_t}{\omega_2} + \frac{\alpha r_c}{\omega_\ell}$$

Substitute this into the second bracket of equation (8.47). Multiply numerator and denominator through by ω_2; then divide through by $r_b(R_2 + r_t)$. The result can then be expressed as

$$\omega_H = \omega_2 \frac{A_r}{A_r'} \left[1 + \frac{r_\varepsilon}{r_b} \left(1 + \alpha \frac{r_c}{r_t + R_2} \cdot \frac{\omega_2}{\omega_\ell} \right) \right]^{-1}$$

Simplify the last term and write

$$\omega_H = \omega_2 \frac{A_r}{A_r'} \frac{1}{1 + \dfrac{r_\varepsilon}{r_b}\left(1 + \dfrac{\alpha}{1-\alpha} \cdot \dfrac{C_w + C_i}{C_T}\right)} \tag{8.48}$$

This will usually be less than ω_2.

The amplifier figure of merit is

$$F_a' = A_r'\omega_H = \frac{F_a}{1 + \dfrac{r_\varepsilon}{r_b}\left(1 + \dfrac{\alpha}{1-\alpha} \cdot \dfrac{C_w + C_i}{C_T}\right)} \tag{8.49}$$

which shows that emitter feedback usually degrades the figure of merit of the amplifier.

A similar analysis of the low frequency case can be made. In this instance, from figure (8.18a) we can show that

$$Z_L = R_2 \left(\frac{s + \omega_g}{s + \omega_x} \right) \quad \text{and} \quad Z_m = R \left(\frac{s}{s + \omega_1} \right)$$

where $\omega_g = \dfrac{1}{R_g C_b}$; $\omega_x = \dfrac{1}{(R_L + R_g)C_b}$; $\omega_1 = \dfrac{1}{(R_1 + R_g)C_b}$

$$R_2 = \frac{R_L R_g}{R_L + R_g}; \quad R_1 = \frac{r_t R_L}{r_t + R_L}; \quad R = \frac{R_1 R_g}{R_1 + R_g}$$

Therefore the low frequency gain function is

$$A'\ (\text{low}) = -A_r \frac{s/(s + \omega_1)}{1 + \dfrac{r_\varepsilon}{r_b} \cdot \dfrac{r_c + R_2(s + \omega_g)/(s + \omega_x)}{r_t + R_2(s + \omega_g)/(s + \omega_x)}} \tag{8.50}$$

After a great deal of algebraic manipulation, this can be written

$$A'\ (\text{low}) = -A_r' \frac{1}{1 + \omega_L/s} = -A_r' \frac{s}{s + \omega_L} \tag{8.51}$$

where A_r' = reference gain given by equation (8.43); ω_L = lower cutoff frequency.

$$\omega_L = \frac{\omega_x(r_b r_t + r_\varepsilon r_c) + \omega_g(r_\varepsilon + r_b)R_2}{(r_b r_t + r_\varepsilon r_c) + (r_\varepsilon + r_b)R_2} \tag{8.52}$$

This can be simplified through the use of equation (8.43) for the reference gain and the equation for the lower cutoff frequency without emitter feedback; that is:

$$\omega_1 = \frac{\omega_x r_t + \omega_g R_2}{r_t + R_2} \tag{8.53}$$

After simplification, the equation for the lower cutoff frequency becomes

$$\omega_L = \omega_1 \frac{A_r'}{A_r}\left[1 + \frac{r_\varepsilon}{r_b}\left(1 - \alpha\frac{\omega_x}{\omega_1}\cdot\frac{R}{R_2}\right)\right] \tag{8.54}$$

Substitute for ω_x, R, ω_1, and R_2 in the last bracketed term and you get

$$\omega_L = \omega_1 \frac{A_r'}{A_r}\left[1 + \frac{r_\varepsilon}{r_b}\left(1 - \frac{\alpha}{1 + R_L/r_t}\right)\right] \tag{8.55}$$

Sample calculations indicate that ω_L will always be less than ω_1. Thus the result is analogous to the grounded cathode vacuum tube amplifier with cathode degeneration.

Exactly the same results will be obtained for the *grounded base* amplifier because it has exactly the same equation for the effective transconductance.

The grounded collector amplifier is quite different. However, the method of analysis is the same; to avoid undue repetition it will not be carried through here. However, the results of the analysis should lead to the following:

$$A_r' = \frac{A_r}{\alpha + \dfrac{r_m}{r_b}\left(\dfrac{r_\varepsilon + R_2}{r_t + R_2}\right)} \tag{8.56}$$

$$\omega_H = \omega_2 \frac{A_r}{A_r'}\left[\frac{1}{\alpha + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{\alpha}{1-\alpha}\cdot\dfrac{C_w + C_i}{C_T}\right)}\right] \tag{8.57}$$

$$\omega_L = \omega_1 \frac{A_r'}{A_r}\left[\alpha + \frac{r_m}{r_b}\left(\frac{r_\varepsilon + R_L}{r_t + R_L}\right)\right] \tag{8.58}$$

$$F_a' = \frac{F_a}{\alpha + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{\alpha}{1-\alpha}\cdot\dfrac{C_w + C_i}{C_T}\right)} \tag{8.59}$$

8.12. Shunt Peaked Amplifier

The high frequency equivalent circuit of a shunt peaked transistor amplifier is shown in figure (8.19). The reactance of the peaking inductance L_b is negligible compared with R_L in the low and mid-frequency regions. Thus the equivalent circuits of the amplifier in these frequency ranges are the same as those of the resistance coupled amplifier.

The type of amplifier, whether grounded emitter, base, or collector is left unspecified in figure (8.19) simply by using the general symbol for

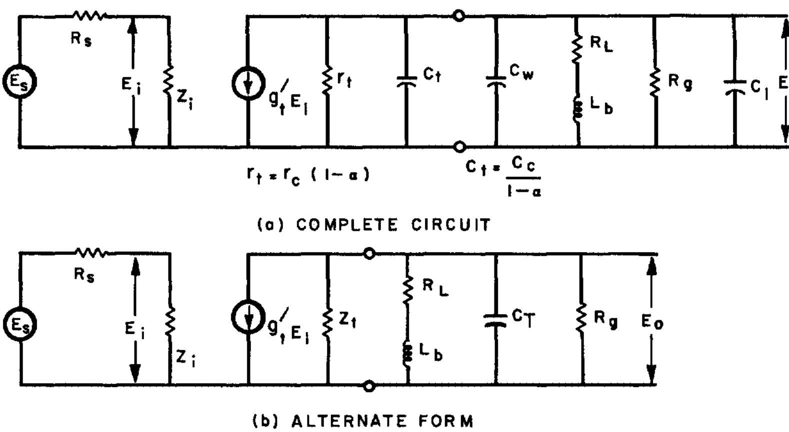


Fig. 8.19. Equivalent circuits of a shunt peaked transistor amplifier of arbitrary type.

the effective transconductance. In any case, because the load is a two-terminal network,

$$Z_m = \frac{z_t Z_L}{z_t + Z_L}$$

$$Z_L = \frac{1}{C_g}\left[\frac{s + \omega_0}{s^2 + (\omega_0 + \omega_g)s + \omega_g \omega_0 R_g/R_2}\right]$$

$$\omega_0 = \frac{R_L}{L_b}; \qquad \omega_g = \frac{1}{R_g(C_w + C_i)}; \qquad C_g = C_w + C_i$$

The voltage gain for a *grounded emitter* amplifier is

$$A' = -\frac{g_t Z_m}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)}$$

$$= -\frac{z_m Z_L}{(r_b + r_\varepsilon)Z_L + r_b z_t + r_\varepsilon z_c} \tag{8.60}$$

Substitute for Z_L, z_t, z_m, and z_c; multiply numerator and denominator through by $(s + \omega_c)$; obtain the common denominator of the gain function.

$$A' = -\frac{\omega_c r_m(s + \omega_0)}{(s+\omega_0)(s+\omega_c)(r_b + r_\varepsilon) + \omega_c r_c C_g[r_b(1-\alpha) + r_\varepsilon][s^2 + (\omega_0 + \omega_g)s + \omega_0\omega_g R_g/R_2]}$$

Multiply all the terms out and then collect coefficients of like powers of s in the denominator. Divide through by the coefficient of s^2. The result is

$$A \text{ (high)} = -K\frac{s + \omega_0}{s^2 + b_1 s + b_0} \tag{8.61}$$

where

$$K = \frac{\omega_c r_m}{(r_b + r_\varepsilon) + (r_b + r_\varepsilon - \alpha r_b)C_g/C_c} \tag{8.62}$$

$$b_1 = \frac{(\omega_0 + \omega_c)(r_b + r_\varepsilon) + (\omega_0 + \omega_g)(r_b + r_\varepsilon - \alpha r_b)C_g/C_c}{(r_b + r_\varepsilon) + (r_b + r_\varepsilon - \alpha r_b)C_g/C_b} \tag{8.63}$$

$$b_0 = \frac{\omega_0\omega_c(r_b + r_\varepsilon) + \omega_0\omega_g(R_g/R_2)(r_b + r_\varepsilon - \alpha r_b)C_g/C_c}{(r_b + r_\varepsilon) + (r_b + r_\varepsilon - \alpha r_b)C_g/C_c} \tag{8.64}$$

Factor ω_0 from equation (8.64). Then, by a rather elaborate algebraic maneuver, rearrange terms and use equation (8.46) to prove that

$$b_0 = \omega_0\omega_H \tag{8.65}$$

where ω_H = upper cutoff frequency of the amplifier when the inductance of the peaking coil is zero.

By a similar and even more involved process, it is possible to prove that

$$b_1 = \omega_0 + \omega_H\left[\frac{R_L}{R_g} + \frac{g_t'}{g_t}\cdot\frac{(1 + r_\varepsilon/r_b)}{(1 + r_t/R_2)}\right] \tag{8.66}$$

This is valid as long as R_g is much greater than R_L. The equation is so complicated that it is desirable to simplify it if possible.

For junction transistors r_t is of the order of a fraction of a megohm; R_2 is nearly equal to R_L, which will usually be only few thousand ohms. Thus r_t will be assumed to be much larger than R_2. Also, for a junction transistor, g_t will ordinarily be many times larger than g_t'. This was proved in a previous section. Thus the second term in equation (8.66) is small. The first term is also small and it will usually be possible to neglect this entire factor, so that

$$b_1 \doteq \omega_0 \tag{8.67}$$

A similar analysis leads to the same result for point contact transistors, though the approximation is not quite so good as it is for a junction transistor.

Now define a *peaking parameter* m as follows:

$$m = \frac{\omega_H L_b}{R_L} = \frac{\omega_H}{\omega_0} \tag{8.68}$$

so that

$$\omega_0 = \frac{\omega_H}{m} \tag{8.69}$$

As a result, the gain function of the shunt peaked transistor amplifier is

$$A = -K\frac{s + \omega_H/m}{s^2 + \omega_H s/m + \omega_H^2/m} \tag{8.70}$$

If you refer back to chapter 4 to the discussion of the shunt peaked vacuum tube amplifier, you will see from equation (4.38) that the gain function of that amplifier and the corresponding transistor amplifier have exactly the same form, within the limitations of the approximations used in both cases. We simply have ω_H for the transistor case where we had ω_2 for the nondegenerative vacuum tube case. Therefore—and this is important—all the design charts for rise time, overshoot, and upper cutoff frequency developed for the shunt peaked amplifier can be used for the shunt peaked transistor amplifier; just replace ω_2 on these charts by ω_H, where ω_H is given by equation (8.46).

Hence the design procedure given in section (4.9) can also be applied here.

The characteristics of the shunt peaked *grounded base* amplifier are exactly the same as those of the grounded emitter amplifier except for the $+$ sign on the gain function.

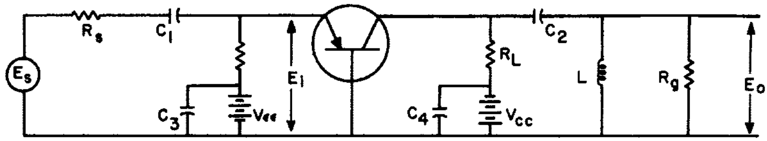
8.13. Single Tuned Amplifier

The circuit diagram of one type of single tuned amplifier is shown in figure (8.20), together with its signal frequency equivalent circuit.

Another, more practical, type of single tuned amplifier was shown in figure (8.14a). Although a grounded base amplifier is shown, the equivalent circuit also applies to the grounded emitter connection.

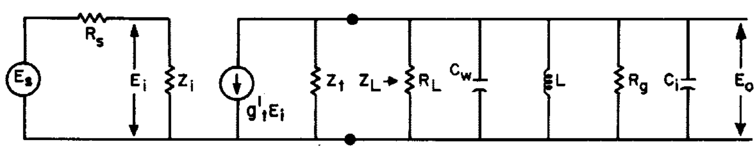
This circuit is like the shunt peaked amplifier in the sense that the mutual impedance of the passive network is equal to the input impedance. Therefore the amplifier gain function is given in equation (8.60) as

$$A' = \pm \frac{z_m Z_L}{Z_L(r_b + r_\varepsilon) + r_b z_t + r_\varepsilon z_c} \tag{8.71}$$



C_1 AND C_2 ARE DC BLOCKING CONDENSERS
C_3 AND C_4 ARE RF BY-PASS CONDENSERS
L IS IDEALLY LOSSLESS, OR LOSSES ARE INCLUDED WITH R_g

(a) CIRCUIT DIAGRAM, GROUNDED BASE AMPLIFIER



(b) EQUIVALENT CIRCUIT FOR EITHER THE GROUNDED BASE OR GROUNDED EMITTER AMPLIFIER

Fig. 8.20. Single tuned transistor amplifier.

From the equivalent circuit of figure (8.20b) it is easily shown that

$$Z_L = \frac{1}{C_L}\left[\frac{s}{s^2 + \omega_\ell s + \omega_r^2}\right] \tag{8.72}$$

where $$C_L = C_w + C_i; \qquad R_2 = \frac{R_L R_g}{R_L + R_g}$$

$$\omega_\ell = \frac{1}{R_2 C_L}; \qquad \omega_r^2 = \frac{1}{LC_L}$$

The transistor parameters z_c, z_m, and z_t are defined in the usual way for the high frequency case.

Substitute all of these expressions into equation (8.71) for the voltage gain. Clear the numerator and denominator of fractions so that the gain equation finally becomes

$$A' = \pm K \frac{s}{s^2 + a_1 s + a_0} \tag{8.73}$$

where

$$K = \frac{\omega_c \omega_\ell r_m R_2}{R_2 \omega_\ell (r_\varepsilon + r_b) + \omega_c (r_c r_\varepsilon + r_t r_b)} \tag{8.74}$$

$$a_1 = \omega_c \omega_\ell \frac{R_2 (r_\varepsilon + r_b) + (r_c r_\varepsilon + r_t r_b)}{R_2 \omega_\ell (r_\varepsilon + r_b) + \omega_c (r_c r_\varepsilon + r_t r_b)} \tag{8.75}$$

$$a_0 = \omega_r^2 \omega_c \frac{r_c r_\varepsilon + r_t r_b}{R_2 \omega_\ell (r_\varepsilon + r_b) + \omega_c (r_c r_\varepsilon + r_t r_b)} \tag{8.76}$$

It is helpful at this point to stop and consider the gain function of a single tuned vacuum tube amplifier. From equation (5.87) we have

$$A \begin{pmatrix} \text{single tuned} \\ \text{vacuum tube} \end{pmatrix} = -F_a \frac{s}{s^2 + Bs + \omega_0^2}$$

where $F_a =$ gain-bandwidth product; $B =$ bandwidth of the single tuned amplifier $=$ upper cutoff frequency of the amplifier when untuned; $\omega_0^2 =$ square of the resonant frequency. Comparison of this gain function with equation (8.73) for the single tuned transistor amplifier shows that $F_a = K =$ transistor amplifier figure of merit; $B = \omega_H = a_1 =$ bandwidth of transistor amplifier; $\omega_0^2 = a_0 =$ resonant frequency squared. By using equations (8.45) and (8.46) for the gain and upper cutoff frequency of a transistor resistance coupled amplifier, you can easily show that this correspondence of terms is valid.

Equation (8.76) for the band center frequency of the tuned transistor amplifier can be simplified somewhat and expressed as follows:

$$\omega_0^2 = a_0 = \omega_r^2 \frac{C_L (r_\varepsilon + r_b - \alpha r_b)}{C_L (r_\varepsilon + r_b - \alpha r_b) + (r_\varepsilon + r_b) C_c} \tag{8.77}$$

As a result, the voltage gain function of a grounded base or grounded emitter transistor amplifier can be written

$$A'(s) = \pm F_a' \frac{s}{s^2 + Bs + \omega_0^2} \tag{8.78}$$

where $F_a' =$ value given by equation (8.49) or (8.74); $B =$ value given by equation (8.46) or (8.75); $\omega_0^2 =$ value given by equation (8.77).

Because the gain function of this amplifier has the same form as that of a corresponding vacuum tube amplifier, the design procedures for both amplifiers are the same. This particular transistor amplifier is not especially practical, particularly in cascaded systems, because the low input impedance of the succeeding stage may well overload the resonant circuit.

8.14. Alpha Cutoff[2]

In all the derivations presented so far it was assumed that the current amplification factor α of the transistor was independent of frequency. This is far from true in practical cases, and the variation in α must be

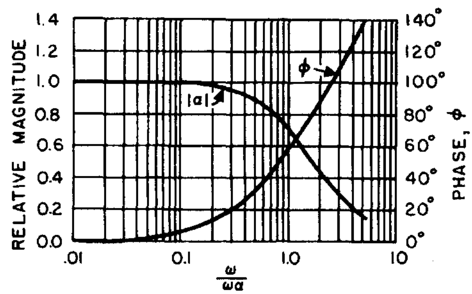


Fig. 8.21. Theoretical variation in magnitude and phase of α as a function of frequency as predicted by equation (8.79).

considered in any wide band amplifier design. Hence the equations derived so far will apply only at those frequencies where α is virtually independent of frequency.

If the nature of the variation of α with frequency can be expressed analytically, our problem in deriving design formulas is made more difficult, but not impossible. The same techniques and principles are used; we just have another frequency dependent term in the original gain function.

The current amplification factor varies both in magnitude and phase as a function of the steady state frequency ω. Typical curves showing these variations are given in figure (8.21) where α_0 designates the low frequency value of the current amplification factor.

[2] R. L. Pritchard, "Frequency Variations of Current-Amplification Factor for Junction Transistors," *Proc. IRE*, vol. 40, November, 1952, pp. 1476–1481; also B. N. Slade, "The Control of Frequency Response and Stability of Point Contact Transistors," *Proc. IRE*, vol. 40, November, 1952, pp. 1382–1384.

The decrease in α as a function of frequency is apparently caused by a number of factors all generally related to the transit time of the minority carrier. Some of these factors may be described as follows:

(1) The flow lines of the minority carriers from emitter to collector are *not* usually straight lines; this causes different charges to travel different distances so that they arrive at the collector slightly out of time phase. This results in a partial neutralization of their effects at the collector.

(2) For other reasons there is a certain amount of dispersion in the travel time of the minority carriers so that some carriers tend to cancel the effects of others.

(3) Both the foregoing factors are influenced to some extent by the base thickness, or spacing between contacts, and the resistivity of the semiconductor.

The variation of α with frequency is usually a complicated function. However, in some cases it can be expressed approximately as

$$\alpha = \alpha_0 \frac{\omega_\alpha}{s + \omega_\alpha} \tag{8.79}$$

where $\alpha_0 =$ low frequency value of α; $\omega_\alpha = \alpha$ cutoff frequency $=$ frequency at which the magnitude of α is equal to $0.707\alpha_0$. Thus the effect of the frequency variation of α can be approximated by substituting equation (8.79) for α into all the high frequency design equations previously derived. The work is routine, though involved, and will not be carried through here.

The effect of α cutoff on the *current gain* of transistor amplifiers is covered in an article by Thomas.[3]

8.15. Current and Power Gain

The *current gain* of an amplifier is easily defined as follows

$$K = \text{current gain} = \frac{I\,(\text{output circuit})}{I\,(\text{input circuit})}$$

The current gain is mainly important when the amplifier has a pure resistance load, so that we shall let $Z_L = R_L$. The current gain is easily evaluated for any amplifier directly from the loop equations written for the tee equivalent circuit. The derivations are purely

[3] D. E. Thomas, "Transistor Amplifier—Cutoff Frequency," *Proc. IRE,* vol. 40, November, 1952, pp. 1481–1483.

routine so only the results are given here for the three basic amplifier configurations.

$$K\,(\text{grounded emitter}) \doteq +\frac{\alpha}{1-\alpha+R_L/r_c} \tag{8.80}$$

$$K\,(\text{grounded base}) \doteq -\frac{\alpha}{1+R_L/r_c} \tag{8.81}$$

$$K\,(\text{grounded collector}) \doteq -\frac{1}{1-\alpha+R_L/r_c} \tag{8.82}$$

Transistors are used as *voltage* amplifiers in cases where they must drive voltage actuated devices such as vacuum tubes, the intensity grid of a cathode ray tube, or the deflection plates of an electrostatically deflected cathode ray tube. In many other cases the transistor should be used as a straight power amplifier, operating under matched load conditions with maximum available power gain. A transistor radio receiver or audio amplifier typify this type of application.

The *power gain* of an amplifier is defined as follows:

$$W = \text{power gain} = \frac{\text{power input to the load}}{\text{power input to the amplifier}} = \frac{P_L}{P_i}$$

The power developed in the load resistance is

$$P_L = E_o^2 R_L$$

where E_o = output voltage developed across R_L; R_L = load resistance. The power input to the amplifier is

$$P_i = E_i^2 R_i = \frac{E_s^2}{R_i}\left(\frac{R_i}{R_i+R_s}\right)^2 = \left(\frac{E_s}{R_i+R_s}\right)^2 R_i$$

where E_i = input voltage to the amplifier; E_s = open circuit voltage of the signal source; R_i = input resistance of the amplifier; R_s = internal resistance of the signal source. Therefore, the power gain of any amplifier is

$$W = \left(\frac{E_o}{E_s}\right)^2 \frac{(R_i+R_s)^2}{R_i R_L}$$

However,
$$\frac{E_o}{E_s} = \frac{E_o}{E_i}\frac{E_i}{E_s} = A\begin{pmatrix}\text{output}\\ \text{circuit}\end{pmatrix} A\begin{pmatrix}\text{input}\\ \text{circuit}\end{pmatrix}$$

For a transistor amplifier,

$$A \text{ (output circuit)} = \pm g_t' R$$
$$= \text{voltage gain of the output circuit}$$

$$A \text{ (input circuit)} = R_i/(R_i + R_s)$$
$$= \text{gain of the input circuit}$$

$$R = r_t R_L/(r_t + R_L)$$

Therefore, the power gain of any transistor amplifier is

$$W = (g_t' R)^2 \, R_i/R_L \tag{8.83}$$

Equation (8.83) can be used to calculate the power gain of any transistor amplifier by substituting the correct values for the effective transconductance, g_t', and input impedance, R_i. This information may be obtained from table 10.

The *available power gain* is obtained when the signal source is matched to the amplifier. This requires that $R_i = R_s$ so that the available power gain is

$$W_a = (g_t' R)^2 \, R_s/R_L \tag{8.84}$$

The condition for *maximum available power gain* (*MAG*) can be computed by maximizing W_a with respect to R_L by the standard method familiar from calculus. However, you must substitute the appropriate values for g_t' and $R_s = R_i$ from table 10 for the particular amplifier under study. The process is very tedious and considerable algebraic maneuvering is required. You can eventually show by this method that the maximum available power gain is obtained under the following conditions:

$$R_L \text{ (grounded emitter)} \doteq r_t \sqrt{\frac{r_b(1-\alpha) + r_\varepsilon}{(r_b + r_\varepsilon)(1-\alpha)}} \tag{8.85}$$

$$R_L \text{ (grounded base)} \doteq r_c \sqrt{\frac{r_b(1-\alpha) + r_\varepsilon}{r_b + r_\varepsilon}} \tag{8.86}$$

$$R_L \text{ (grounded collector)} \doteq r_t \sqrt{\frac{r_b(1-\alpha) + r_\varepsilon}{r_t}} \tag{8.87}$$

These resistances are ordinarily quite large, especially for the grounded emitter and grounded base circuits. The same circuits usually have very low input resistances. Therefore, interstage impedance transformers are required when stages are cascaded. Practical transistor power amplifiers should be designed to approximate this condition of matched impedances so that the maximum available power gain is obtained.

8.16. General Aspects of Cascading

Theoretically, any combination of the various types of transistor amplifiers could be cascaded. This is indicated symbolically in figure (8.22). However, there are a number of practical difficulties that must be understood. In the two-stage cascade shown in the figure, it is clear that the input impedance of stage 2 is a part of the load circuit of stage 1. This makes the evaluation of Z_L and Z_m for stage 1 a fairly complex problem. Also, except for the grounded collector amplifier, the input impedance of transistor amplifiers is quite low, being of the order of only a few hundred ohms in most cases. This would make it difficult to cascade shunt peaked amplifiers or parallel tuned amplifiers. These are

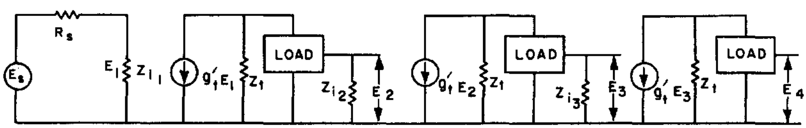


Fig. 8.22. Equivalent circuit of a three stage amplifier.

essentially practical matters and should not deter serious and careful workers. Moreover, if you consider the problem carefully, you will see that the magnitude of the input impedance can be increased, though this results in other penalties on performance.

Except for the factors just noted, the principles of cascading transistor amplifiers are essentially the same as those for vacuum tube amplifiers. Synthesis by factoring can be used to design amplifiers to specific gain functions. Space limitations make it impossible to pursue the subject further here.

PROBLEMS

The following data which are required for the problems, are given for two different transistors.

Parameter	*Point contact*	*Junction type*
r_e	250 ohms	25 ohms
r_b	300 ohms	250 ohms
r_c	20,000 ohms	10 megohms
r_m	15,000 ohms	9.75 megohms
C_c	1 $\mu\mu$f	8 $\mu\mu$f

8.1. Compute the reference voltage, current, and power gain as well as the input impedance of a grounded emitter point contact transistor amplifier if $R_s = 500$ ohms and $R_L = 20{,}000$ ohms.

8.2. Repeat problem (8.1) for a grounded base amplifier.

8.3. Repeat problem (8.1) for a grounded collector amplifier, but use $R_s = 20{,}000$ ohms and $R_L = 10{,}000$ ohms.

8.4. Tabulate the results of problems (8.1) through (8.3) and comment on the comparative values.

8.5. Repeat problem (8.1) for a junction transistor having $R_s = 25$ ohms and $R_L = 200{,}000$ ohms.

8.6. Compute the cutoff frequencies of the junction transistor connected as a resistance coupled, grounded base amplifier. Compute the reference gain and figure of merit. Assume that $R_L = 10{,}000$ ohms, $C_b = 0.01\ \mu$f, $R_g = 50$ K, $C_i = 0$, and $C_w = 2\ \mu\mu$f.

8.7. Derive equation (8.61) for the single tuned amplifier.

8.8. Work out a design procedure for designing resistance coupled transistor amplifiers to have a specific set of cutoff frequencies and reference gain. Work out design formulas for all the parameters of interest.

8.9. Extend the design procedure developed in (8.8) to the case of a shunt peaked transistor amplifier.

8.10. Design a shunt peaked transistor amplifier to have a rise time of 0.2 μsec with not more than 2% overshoot and a gain of not less than 10. The amplifier drives the intensity grid of a cathode ray tube, which has an input capacitance of 10 $\mu\mu$f. The wiring capacitance is estimated to be 2 $\mu\mu$f. This is more of an exploration of design possibilities, rather than a straightforward design of a specific circuit. Discuss some of the problems involved in selecting a transistor for use in this amplifier.

8.11. What problems would you encounter if the design requirements of problem (8.10) had stipulated a rise time of 0.05 μsec?

Chapter 9

NOISE

Communication in the presence of noise is an everyday occurrence for every living person. Nearly all exchanges of information, such as a normal conversation, are carried on in the presence of interference from other conversations, the clanking of machinery, the roar of wind, or any other one or combination of many sources of noise. It is common experience that successful communication can be established only if the strength of the conversation, or *signal*, is sufficiently large to overcome the masking effects of the background noise. Thus the ratio of the signal to the noise will govern the amount of information exchanged.

It is fairly obvious that the problem of information transmission is generally complicated by the presence of noise. In some cases the noise is *pure noise* in the sense that it is purely random and disordered and therefore conveys no intelligence. A classical example of this kind of noise may be found at any cocktail party. There is a continual drone of voices that establishes a mean sound level that must be overcome if information is to be exchanged between two particular people. If the conversation is conducted at a power level equal or less than the general noise power level, the information exchanged will be slight. If suitable vocal power is used, information can be transmitted, because the spoken words constitute a *pattern* in time that can be interpreted by the listener even though it may be partially obliterated by the purely random background noise of the room. Thus information transmission requires establishment of a pattern of sufficient strength to overcome the masking effects of the purely random noise.

Now consider a slightly different situation. Suppose that one person is located in a sound-deadened room, empty except for a microphone. The microphone is connected through an amplifier to a loudspeaker. Regardless of how softly the person might speak, it would be theoretically possible to increase the gain of the amplifier and thereby detect what was said. However, eventually a point is reached where the air pressure variations on the microphone caused by the whispers of the person will be of the same order of magnitude as the impacts of the thermally agitated gas molecules in the air. The pressure variations on the microphone caused by the signal would then be obliterated by the random effects of the air molecules. It is clear that there is an ultimate limit to the useful sensitivity of the amplifier.

The noise in an amplifier might be introduced along with the signal, as in the preceding example, or it might be introduced by the amplifier itself. Ideally, the signal will enter the amplifier alone, without noise. Thus the final limit upon the sensitivity of the amplifier will be set by the noise it introduces by itself. This kind of noise is the primary concern of this chapter.

9.1. Thermal Noise

Nearly all the basic physical laws in the repertory of the engineer describe the large-scale effects or over-all behavior of an aggregation of microscopic particles. These laws apply only to those cases in which the number of discrete particles in the aggregation is so large that statistical averages can be used. They can seldom be used to describe the behavior of a single member of the aggregation.

This kind of situation is analogous to the actuarial tables of the insurance companies. Such tables will predict with considerable accuracy the percentage of deaths in a single year of persons in a certain age group. However, it is impossible to determine when particular individuals in the statistical group will die.

Electrical engineering is based upon laws of this type. It has been established that electric charge is not infinitely divisible. There *is* a smallest unit of electric charge, the charge on an electron. The basic laws of electrical engineering describe the behavior and effects of an aggregation of electronic charges without attempting to specify the actions of a single member of the aggregation. Thus, contrary to Ohm's law, if an attempt is made to apply it to microscopic cases, the *free* electrons in a conductor in the absence of an applied electric field are not at rest, but are undergoing extremely vigorous gymnastics in a purely random manner. The effect is much the same as the Brownian motion of particles and the thermal motion of gas molecules. The relative activity of the electrons is dependent upon the temperature of the conductor.

Electric current is defined as

$$i = dq/dt$$

Thus the random motion of the electrons in a conductor sets up a time-varying charge distribution that causes a purely random current to flow in the conductor. Because it is purely random and is caused by thermal effects, this component of current is called *thermal noise.*

It has been found[1] that the rms noise voltage and current in a conductor can be evaluated by the following formulas:

$$\bar{I}_n = \sqrt{\frac{2}{\pi} kT\beta G} \quad \text{rms amp} \tag{9.1}$$

$$\bar{E}_n = \sqrt{\frac{2}{\pi} kT\beta R} \quad \text{rms volts} \tag{9.2}$$

where $R = 1/G =$ resistance of the conductor; $k =$ Boltzmann's constant $= 1.38 \times 10^{-23}$ joule/°K; $T =$ absolute temperature of the conductor; $\beta =$ noise bandwidth, in radians/sec, of the circuit in which the conductor is located. The noise bandwidth β is not exactly the same thing as the 0.707 bandwidth. This will be explained later.

The thermal noise in the input circuit of an amplifier is amplified along with the signal. Thus the ultimate sensitivity of the amplifier will be governed by the amount of noise introduced at this point. In addition to the input circuit noise, noise is also introduced by the amplifier tube. This is discussed in the next two sections.

9.2. Shot Noise

Thermal noise was shown to arise from a purely random and disordered motion of electrons in a conductor. The noise generated in a vacuum tube is developed in an inherently different manner because the motion of the electrons is unilateral, always from cathode to plate Random motion, in the sense used to describe behavior in a conductor, simply does not occur. However, because the arrival of electrons at the plate of the tube is not thoroughly uniform with respect to time, but occurs in multiples of the electronic charge, a noise component of current appears at the plate. This is called *shot noise.* Like thermal noise, it is caused by the granular character of electric charge.

Shot noise would be produced even if electrons were emitted from the cathode at equally spaced intervals of time. This does not occur, however, because there is a probability distribution of emission times that acts to increase the shot noise over the theoretical minimum. The presence of space charge tends to suppress this effect to some extent.

[1] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, vol. 32, pp. 110, 1928.

It is convenient to account for the presence of the shot noise in the plate circuit by assuming that it is produced by an ideal voltage generator in series with the grid of the tube. This fictitious generator has an open circuit voltage

$$\bar{E}_{sn} = \sqrt{\frac{2}{\pi} kT\beta R_{sn}} \tag{9.3}$$

where R_{sn} = shot noise parameter $= 2.5/g_m$ (9.4) T = ambient temperature °K. The approximation for the shot noise parameter is valid for tubes with oxide coated cathodes.

9.3. Partition Noise and Induced Grid Noise

Ordinary shot noise appears in all vacuum tubes, diodes, triodes, tetrodes, and pentodes alike. However, in the case of pentodes and tetrodes, there are two electrodes that draw current from the electron stream, the plate and the screen. Because the current division between these two electrodes will vary in a purely random fashion, the apparent shot noise in a pentode will be larger than in a triode. This increase in shot noise is identified as *partition noise*; like shot noise it is assumed to be produced by a fictitious voltage generator in series with the grid of the tube. The open circuit voltage of this generator is[2]

$$\bar{E}_{pn} = \sqrt{\frac{2}{\pi} kT\beta R_{pn}} \quad \text{rms volts}$$

where R_{pn} = partition noise parameter $\doteq 20I_d/[g_m(I_b + I_d)]$; g_m = tube transconductance; I_b = plate current through the tube; I_d = screen current of the tube. Thus it is possible to specify an *equivalent noise resistance* R_{eq}, which is merely a fictitious parameter, to be the sum of the shot noise and partition noise parameters. That is,

$$R_{eq}\,(\text{triodes}) = R_{sn} \doteq \frac{2.5}{g_m} \tag{9.4}$$

$$\begin{aligned} R_{eq}\,(\text{pentodes}) &= R_{sn} + R_{pn} \\ &\doteq \frac{2.5}{g_m} + \frac{20I_d}{g_m(I_b + I_d)} \end{aligned} \tag{9.5}$$

The combined effects of shot and partition noise are represented by a

[2] W. A. Harris, "Some Notes on Noise Theory and Its Application to Input Circuit Design," *Electron Tubes*, vol. 2, *RCA Rev.*, Princeton, N. J., p. 286.

generator in series with the grid of the tube and having an open circuit voltage

$$\bar{E}_{np} = \sqrt{\frac{2}{\pi} kT\beta R_{eq}} \tag{9.6}$$

Representative values of the equivalent noise resistance of various vacuum tubes are given in table 12. It is important that the noise resistance of a pentode is considerably larger than that of a triode. The values in this table should not be taken too literally because they will vary considerably from tube to tube; this results from differences in g_m, I_b, and I_d permissible under existing manufacturing specifications.

Another type of shot noise that appears in triodes and pentodes is called *induced grid noise*. It is caused by shot noise fluctuations induced in the grid of the tube as a result of unequal charge distributions on either side of the grid. The induced grid noise is related to the *electronic* input conductance of the tube. This conductance G_T is given approximately by the following equation[3];

$$G_T = \frac{g_m(\omega t)^2}{20} = 1.92 g_m f^2 t^2 \tag{9.7}$$

where $f =$ frequency; $t =$ transit time. Typical approximate values for this *transit time conductance* are given in table 13 for a frequency of 30 mc. Because the value of G_T depends upon the square of frequency, G_T can be approximated at other frequencies by simple proportions using the values tabulated at 30 mc.

TABLE 12

EQUIVALENT NOISE RESISTANCE OF VACUUM TUBES

Tube	*Connection*	R_{eq} *in ohms*
6AC7	pentode	650
6AC7	triode	214
6AG5	pentode	1900
6AG5	triode	380
6AK5	pentode	1900
6AK5	triode	380
6BA6	pentode	3800
6BA6	triode	410
6BJ6	pentode	3800
6BJ6	triode	485
6J4	triode	210
6SK7	pentode	11,600
6SK7	triode	970

[3] *Ibid.*

TABLE 13

TRANSIT TIME CONDUCTANCE OF VACUUM TUBES

Tube	G_T *μmhos at* 30 *mcps*
6AC7	156
6AG5	27
6AK5	12
6BA6	52
6BJ6	26
6J6	18
6SK7	40

The induced grid noise is assumed to be produced by thermal noise in the transit time conductance at a temperature γ times larger than the ambient temperature. Hence the induced grid noise is represented by a generator of short-circuit current

$$I_{ng} = \sqrt{\frac{2}{\pi} k\gamma T\beta G_T} \quad \text{rms amp} \tag{9.8}$$

and internal impedance G_T. For tubes with oxide coated cathodes, γ is usually taken to be approximately 5.

9.4. Noise Bandwidth

The noise bandwidth β appears in all the expressions that have been given for calculating the rms noise voltages and currents, and is defined properly as the width of an idealized, rectangular bandpass characteristic having the same area and peak value as the power gain *vs.* frequency characteristic of the amplifier. The general nature of the noise bandwidth is apparent from figure (9.1).

The necessity for defining the noise bandwidth is understood from the fact that the noise spectrum is uniform from zero to infinite frequency. Because the noise voltage and current developed are proportional to the total noise power, this noise power can be computed from the area under the power gain characteristic of the amplifier. Clearly, the same noise power is developed if the amplifier has the idealized frequency characteristic shown in figure (9.1b).

The preceding definition of noise bandwidth can be expressed mathematically as

$$\beta = \int_0^\infty \frac{W(\omega)}{W_r}\, d\omega \tag{9.9}$$

where $W(\omega)$ = power gain as a function of frequency; W_r = reference value of power gain.

If the input and output resistances of an amplifier are independent of frequency, the power gain can be written in terms of voltages as follows:

$$W(\omega) = \frac{P_o(\omega)}{P_i(\omega)} = \left|\frac{E_o(\omega)}{E_i(\omega)}\right|^2 \frac{R_i}{R_o} = |A(\omega)|^2 \frac{R_i}{R_o}$$

$$W_r = \left|\frac{E_o(\omega_r)}{E_i(\omega_r)}\right|^2 \frac{R_i}{R_o} = A_r^2 \frac{R_i}{R_o}$$

Consequently, the noise bandwidth can be expressed as

$$\beta = \int_0^\infty \left|\frac{A(\omega)}{A_r}\right|^2 d\omega \tag{9.10}$$

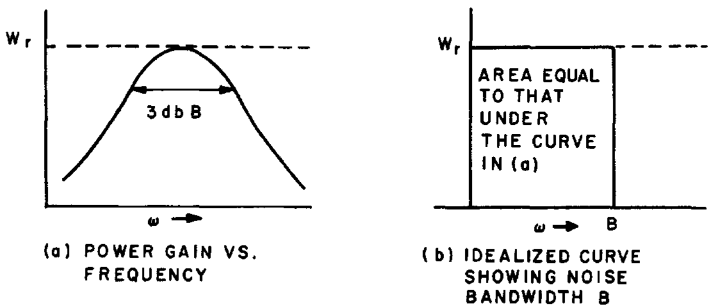


Fig. 9.1. Illustration of the meaning of noise bandwidth.

The integral in equation (9.10) is fairly easy to evaluate for video amplifiers, but bandpass amplifiers are rather difficult to handle. The problem can be simplified somewhat by transforming to the normalized frequency gain function. That is, in the bandpass case, the following gain function will apply to all cascades of synchronous and maximally flat-staggered amplifiers:

$$\left|\frac{A(y)}{A_r}\right| = \left(\frac{1}{\sqrt{1 + y^{2n}}}\right)^m$$

where $n = n$-uple and one n-uple $= 1$ stage; $m =$ number of stages.

Because the frequency y in this equation is normalized, the noise bandwidth is similarly normalized, so that

$$\frac{\beta}{B_n} = \int_0^\infty \left(\frac{1}{1 + y^{2n}}\right)^m dy \tag{9.11}$$

for any synchronous or flat-staggered amplifier, where $B_n =$ bandwidth of the n-uple. A few examples will clarify the use of this equation.

To compute the noise bandwidth of a single high Q amplifier stage, in equation (9.11) let $n = 1$, $m = 1$, and $B_n = 0.707$ bandwidth of the stage.

$$\frac{\beta}{B_n} = \int_0^\infty \frac{1}{1+y^2}\, dy = \left(\tan^{-1} y\right)_0^\infty$$
$$= \frac{\pi}{2} = 1.571$$

Now consider a two-stage synchronous cascade. In this case $n = 1$, $m = 2$, and

$$B_n = \frac{B_m}{\sqrt{2^{1/n} - 1}} = \frac{B_m}{0.643} = 0.707 \text{ stage bandwidth}$$

$$B_m = 0.643 B_n = 0.707 \text{ bandwidth of the cascade}$$

Therefore for two synchronous stages

$$\frac{\beta}{B_n} = \frac{0.643\beta}{B_m} = \int_0^\infty \frac{1}{(1+y^2)^2}\, dy = \frac{\pi}{4}$$

so that

$$\frac{\beta}{B_m} = \frac{\pi}{4(0.643)} = 1.222$$

If the same procedure is followed for a three-stage synchronous cascade, the result is $\beta/B_m = 1.155$. Similar cases are easily worked out. Note that in all cases the result is expressed as a ratio of the noise bandwidth to the over-all 0.707 bandwidth of the amplifier cascade.

Now consider the case of a single staggered pair. In this case, $n = 2$, $m = 1$, and $B_m = B_n$.

$$\frac{\beta}{B_m} = \int_0^\infty \frac{1}{1+y^4}\, dy$$

The integral can be evaluated by expanding the function into partial fractions and then evaluating the integral term by term. This process yields $\beta/B_m = 1.11$.

9.5. Other Sources of Noise

It has been shown that noise is produced in vacuum tubes by the shot effect, partition effect, and induced grid noise. There are other, subsidiary sources of noise also present. They are cathode flicker effect, positive ion noise, secondary emission noise, and microphonics.

At audio frequencies it is observed that the actual noise produced by tubes having oxide coated cathodes is considerably greater than that computed from the theoretical formula for shot effect and induced grid noise. It occurs if the tube is space charge limited or if it is temperature saturated. It is also observed at much lower frequencies for tubes with tungsten filaments. This phenomenon is called *cathode flicker* and it is believed[4] to be caused by the variation in cathode emissivity caused by the arrival or departure of foreign atoms or molecules at the cathode surface. The increased tube noise that results apparently varies with the inverse square of the frequency and with the square of the emission current.

Positive ion noise is produced by the positive ion current in the tube. These ions may be emitted by the cathode or produced by ionization of the residual gases in the tube. The noise produced is small because the positive ion current is small compared with the electronic current.

Secondary emission causes the actual noise of the tube to deviate from the value computed on the basis of the shot effect. However, the contribution is generally small and is often neglected.

Microphonics are noises caused by mechanical vibrations of the tube structure. They usually occur at low frequency and do *not* exhibit the random character of the other noises mentioned thus far. With recent developments in ruggedized construction, this effect has been largely overcome. However, care should be exercised to avoid undue vibration of the tube because the effects of microphonics can be serious.

In addition to the tube noise and the thermal noise in the input circuit, there is also the possibility that noise will be introduced by the signal source driving the amplifier. This noise might be thermal noise in the source conductance plus other noises such as *atmospherics* (static), diathermy machines, intentional interference (jamming), ignition systems, and so on.

9.6. Representation of Noise

The preceding sections have listed and described the various sources of noise in electronic circuits. This is interesting, but it is of little practical use to the circuit designer until the effects of these noises can be expressed in terms of circuit elements that can be combined mathematically and manipulated to yield answers to pressing practical problems.

[4] Valley and Wallman, *Vacuum Tube Amplifiers*, vol. 18, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, p. 588

Noise sources, like any other energy source in an electric circuit, can be treated in many ways, but all methods usually involve a generator of some type and some kind of internal impedance or admittance. Consequently, it is convenient to represent a noise source by an equivalent Thevenin or Norton generator as shown in figure (9.2). This kind of representation is valid for any kind of noise source if particular care is used in the specification of the terms in the equivalent circuit. That is,

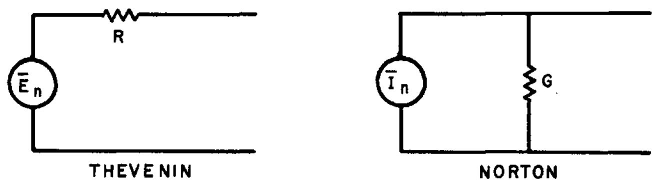


Fig. 9.2. Equivalent circuits of noise sources.

either of these two circuits is an excellent equivalent circuit of a *thermal* noise source where

$$\begin{aligned}\bar{E}_n &= \text{open circuit voltage}\\ &= \sqrt{\frac{2}{\pi}kT\beta R} \quad \text{rms volts}\\ \bar{I}_n &= \text{short circuit current}\\ &= \sqrt{\frac{2}{\pi}kT\beta G} \quad \text{rms amp}\\ R &= 1/G = \text{noise-producing resistance}\end{aligned}$$

Thus the equivalent circuit of a *thermal* noise source is direct.

For *shot* noise sources the equivalent representation is a little less direct. Actually, a two-step transformation is involved:

(1) The shot noise source is replaced by a hypothetical thermal noise source that generates the same amount of noise as the original shot noise source.

(2) The fictitious thermal noise source is then described in terms of the equivalent circuits of figure (9.2).

Thus the same equivalent circuit is used to describe *any* noise source.

The difference between thermal and shot noise representations is simply that the R or G in the equivalent circuit and equations is the actual ohmic resistance of the element in the case of thermal noise, but a fictitious resistance, or parameter, in the case of shot noise.

9.7. Equivalent Noise Representation of an Amplifier

It was shown in the preceding section that any noise source can be represented by an equivalent Norton or Thevenin generator. This technique will now be used in the analysis of an isolated amplifier stage.

Noise is introduced into the amplifier input from the following sources:

(1) The signal generator supplying the desired signal input introduces noise caused by thermal agitation in its internal resistance.

(2) Thermal noise is generated in the ohmic component of the input circuit.

(3) The tube introduces noise from two sources, the equivalent noise resistance in series with the grid and the induced grid noise.

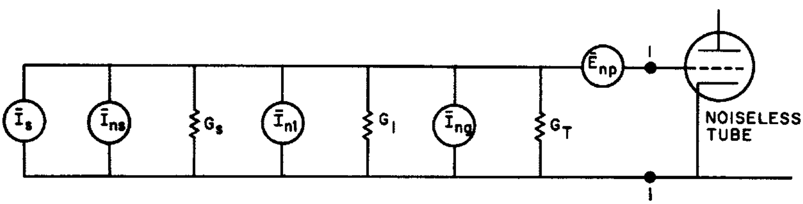


Fig. 9.3. Equivalent circuit of an amplifier with all noise sources referred to the grid circuit.

Thus the equivalent circuit of the amplifier will contain the following noise sources:

Noise source	*Current or voltage*	*Internal G*
Signal source	$\bar{I}_{ns} = \sqrt{\frac{2}{\pi} kT\beta G_s}$	G_s
Input circuit	$\bar{I}_{n1} = \sqrt{\frac{2}{\pi} kT\beta G_1}$	G_1
Induced grid noise	$\bar{I}_{ng} = \sqrt{\frac{2}{\pi} kT\gamma\beta G_T}$	G_T
Shot and partition	$\bar{E}_{np} = \sqrt{\frac{2}{\pi} kT\beta R_{eq}}$	0

If the signal source is included, the complete equivalent circuit of the amplifier appears as shown in figure (9.3). Because all noise sources have been removed from the tube and circuit elements, all the passive components in figure (9.3) are noiseless.

This equivalent circuit can be simplified somewhat and put into a more convenient form by determining the Norton equivalent of the entire circuit to the left of the marked terminals in figure (9.3). This is pretty easy because the equivalent Norton generator supplies a current equal to the short-circuit current that flows when terminals 1–1 are short-circuited. Hence

$$\bar{I}_{sc}^2 = \bar{I}_s^2 + \bar{I}_{ns}^2 + \bar{I}_{n1}^2 + \bar{I}_{ng}^2 + \bar{E}_{np}^2(G_s + G_1 + G_T)^2 \tag{9.12}$$

It is convenient to define the last term in this equation as

$$\bar{I}_{np}^2 = \bar{E}_{np}^2(G_s + G_1 + G_T)^2 \tag{9.13}$$

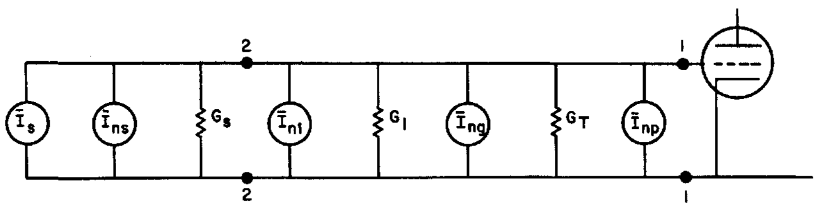


Fig. 9.4. Final equivalent circuit with all noise sources referred to the grid circuit.

Thus the equivalent Norton generator is actually the parallel combination of five separate current generators as follows:

$$\bar{I}_{sc}^2 = \bar{I}_s^2 + \bar{I}_{n1}^2 + \bar{I}_{ns}^2 + \bar{I}_{ng}^2 + \bar{I}_{np}^2 \tag{9.14}$$

The internal admittance remains the sum of the existing conductances. Thus the final equivalent circuit appears as shown in figure (9.4), where the magnitudes of the current generators are

$$\bar{I}_s = \text{rms signal current} \tag{9.15}$$

$$\bar{I}_{ns} = \sqrt{\frac{2}{\pi}\, kT\beta G_s} \tag{9.16}$$

$$\bar{I}_{n1} = \sqrt{\frac{2}{\pi}\, kT\beta G_1} \tag{9.17}$$

$$\bar{I}_{ng} = \sqrt{\frac{2}{\pi}\, kT\gamma\beta G_T} \tag{9.18}$$

$$\bar{I}_{np} = \sqrt{\frac{2}{\pi}\, kT\beta R_{eq}}\ \ (G_s + G_1 + G_T) \tag{9.19}$$

9.8. Noise Figure of an Amplifier Stage

The ultimate purpose behind the preceding formulation of an equivalent circuit is the development of a method of specifying the noise characteristics of a given amplifier. Before proceeding, certain definitions are required.

A very important term is the *available power*. The available power is simply the power that a *matched* generator can supply. For example consider the case of a generator of voltage E and internal resistance R connected to a matched load R. The power delivered to this load, as shown in figure (9.5), is the available power, and it has a value

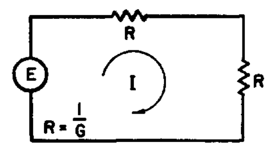


Fig. 9.5. Evaluation of available power.

$$W = \frac{\bar{E}^2}{4R} = \frac{\bar{I}^2}{4G} \tag{9.20}$$

where $\bar{E}$ and $\bar{I}$ are *rms* values. This equation can be used to compute the available power from some noise source, such as the input circuit. In this case,

$$W = \text{available noise power} = \frac{\bar{I}_{n1}^2}{4G_1} = \frac{2kT\beta G_1/\pi}{4G_1}$$

The definition of available noise power now permits another very important factor, the *signal-to-noise ratio*, to be defined. That is

$$S/N = \text{signal-to-noise ratio} = \frac{\text{available signal power}}{\text{available noise power}} \tag{9.21}$$

For example, the signal-to-noise ratio measured at the terminals marked 2–2 in figure (9.4), would be computed as follows:

$$S_2 = \text{available signal power at 2–2}$$
$$= \bar{I}_s^2/4G_s$$
$$N_2 = \text{available noise power at 2–2}$$
$$= \bar{I}_{ns}^2/4G_s$$

Hence the signal-to-noise ratio is

$$\frac{S_2}{N_2} = \left(\frac{S}{N}\right)_2 = \frac{\bar{I}_s^2}{\bar{I}_{ns}^2} \tag{9.22}$$

The specification of the signal-to-noise ratio now permits the definition of a useful factor called the *noise figure*. It is defined in the following way:

$$F = \text{noise figure} = \frac{(S/N) \text{ in ideal case}}{(S/N) \text{ in actual case}} \tag{9.23}$$

In applying this definition to the amplifier circuit of figure (9.4), it is clear that the noise figure of the stage can be expressed as

$$\begin{aligned} F &= \frac{(S/N) \text{ evaluated at terminals 2–2}}{(S/N) \text{ evaluated at terminals 1–1}} \\ &= \frac{(S/N)_2}{(S/N)_1} \end{aligned} \tag{9.24}$$

The noise figure of the amplifier of figure (9.4) can now be evaluated Since $(S/N)_2$ was computed in equation (9.22), it is necessary only to find $(S/N)_1$.

$$\begin{aligned} S_1 &= \text{available signal power at 1–1} \\ &= \frac{\bar{I}_s^2}{4(G_s + G_1 + G_T)} \\ N_1 &= \text{available noise power at 1–1} \\ &= \frac{\bar{I}_{ns}^2 + \bar{I}_{n1}^2 + \bar{I}_{ng}^2 + \bar{I}_{np}^2}{4(G_s + G_1 + G_T)} \end{aligned}$$

Hence the signal-to-noise ratio at terminals 1–1 is

$$\left(\frac{S}{N}\right)_1 = \frac{S_1}{N_1} = \frac{\bar{I}_s^2}{\bar{I}_{ns}^2 + \bar{I}_{n1}^2 + \bar{I}_{ng}^2 + \bar{I}_{np}^2} \tag{9.25}$$

Consequently, the noise figure of the amplifier stage is

$$F = \frac{(S/N)_2}{(S/N)_1} = \frac{\bar{I}_{ns}^2 + \bar{I}_{n1}^2 + \bar{I}_{ng}^2 + \bar{I}_{np}^2}{\bar{I}_{ns}^2} \tag{9.26}$$

The equations for the four noise currents appearing in equation (9.26) for the noise figure were given earlier. If these relationships are substituted into the noise figure equation, the common $2kT\beta/\pi$ term cancels, leaving

$$F = \frac{G_s + G_1 + \gamma G_T + R_{eq}(G_s + G_1 + G_T)^2}{G_s} \tag{9.27}$$

Rearrange terms slightly and write the noise figure as

$$F = 1 + \frac{G_1 + \gamma G_T}{G_s} + \frac{R_{eq}}{G_s}(G_1 + G_s + G_T)^2 \tag{9.28}$$

Although given as a ratio here, the noise figure is customarily expressed in power logits or decibels.

This is an important and useful result for a number of reasons. For one thing, the noise figure of an amplifier stage can be computed from this formula if the various conductances are known. Alternatively, if all but one of the quantities appearing in the formula are known, the remaining unknown can be found.

From the theoretical standpoint, the noise figure equation indicates the manner in which the various terms affect the noise figure, and thereby permits certain recommendations to be made to reduce the noise figure. For example, the formula shows that the noise figure is linearly dependent upon the equivalent noise resistance R_{eq} of the tube. Therefore, to obtain a small noise figure, a tube with a small R_{eq} should be selected. Evidently then, triode amplifiers will have lower noise figures than pentode amplifiers.

9.9. Over-all Noise Figure of Cascaded Stages

It was shown in the preceding section that the noise figure of an isolated amplifier stage could be written

$$F = \frac{\overline{I_{ns}^2} + \overline{I_{n1}^2} + \overline{I_{ng}^2} + \overline{I_{np}^2}}{\overline{I_{ns}^2}}$$

Divide both the numerator and denominator through by $4(G_s + G_1 + G_T)$. Thus the noise figure becomes

$$F = \left[\frac{\overline{I_{ns}^2} + \overline{I_{n1}^2} + \overline{I_{ng}^2} + \overline{I_{np}^2}}{4(G_s + G_1 + G_T)}\right] \Big/ \left[\frac{\overline{I_{ns}^2}}{4(G_s + G_1 + G_T)}\right]$$

The numerator and denominator of this expression now have special meanings, as follows: $N_{1-1}(\text{act})$ = numerator = actual available noise power at terminals 1–1; and $N_{1-1}(\text{min})$ = denominator = minimum available noise power at terminals 1–1. Therefore the noise figure of the amplifier stage can be written in terms of these factors as

$$F = \frac{N_{1-1}\,(\text{act})}{N_{1-1}\,(\text{min})} \tag{9.29}$$

The minimum available power is interpreted simply as the noise power when the amplifier is noiseless and has a noise figure of unity.

Thus it is relatively easy to compute the noise caused by the amplifier itself. That is, in the grid circuit of the amplifier,

$$\begin{pmatrix}\text{noise caused by}\\ \text{the amplifier}\end{pmatrix} = \begin{pmatrix}\text{actual noise}\\ \text{at terminals 1–1}\end{pmatrix} - \begin{pmatrix}\text{minimum noise}\\ \text{terminals 1–1}\end{pmatrix}$$

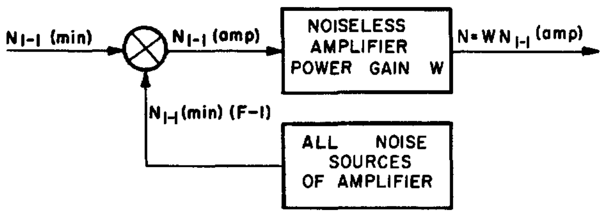


Fig. 9.6. Noise equivalent circuit of an isolated amplifier stage.

Or, in symbolic notation,

$$N_{1-1}(\text{amp}) = N_{1-1}(\text{act}) - N_{1-1}(\text{min})$$

or

$$N_{1-1}(\text{amp}) = N_{1-1}(\text{min})(F - 1) \tag{9.30}$$

This is the equation for the available noise power produced by the

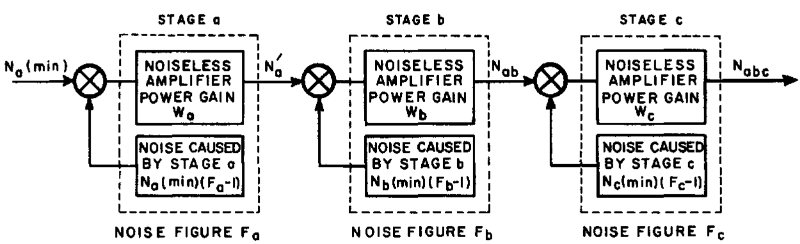


Fig. 9.7. Noise equivalent circuit of three amplifier stages in cascade.

entire amplifier circuit and referred to the input circuit. The results of this brief analysis are best understood from the noise equivalent circuit of the amplifier stage shown in figure (9.6).

This type of equivalent circuit is immensely helpful in computing the over-all noise figure of several amplifier stages connected in cascade. Such a case is illustrated in figure (9.7).

The noise figures of stage *a*, of stages *a* and *b* in cascade, or of the three-stage cascade are easily computed from this equivalent circuit using the noise figure definition given in equation (9.29). It is necessary only to remember that the noise figure is unity in the ideal, or minimum, case. Therefore the noise figure of stage a is

$$F_a = \frac{N'_a\,(\text{act})}{N'_a\,(\text{min})} = \frac{[N_a(\text{min}) + N_a(\text{min})\,(F_a - 1)]W_a}{N_a(\text{min})W_a}$$

In the same way, the over-all noise figure of stages a and b in cascade is

$$F_{ab} = \frac{N_{ab}(\text{act})}{N_{ab}(\text{min})}$$

where

$$\begin{aligned} N_{ab}(\text{act}) &= [N_a + N_b(\text{min})(F_b - 1)]W_b \\ &= N_a(\text{min})F_a W_a W_b + N_b(\text{min})(F_b - 1)W_b \\ N_{ab}(\text{min}) &= N_a(\text{min})W_a W_b \end{aligned}$$

Therefore the over-all noise figure of the two stage cascade is

$$F_{ab} = F_a + \frac{N_b(\text{min})}{N_a(\text{min})} \cdot \frac{(F_b - 1)}{W_a} \tag{9.31}$$

If the amplifier stages are identical so that $N_a(\text{min}) = N_b(\text{min})$, then the over-all noise figure simplifies to the following form:

$$F_{ab} = F_a + \frac{F_b - 1}{W_a} \tag{9.32}$$

This is the equation that is usually given, though the stipulation of identical stages is seldom noted.

A similar method can be applied to the three-stage cascade and the result is

$$F_{abc} = F_a + \frac{F_b - 1}{W_a} + \frac{F_c - 1}{W_a W_b} \tag{9.33}$$

The over-all noise figure of an n-stage cascade is easily computed in this same way.

The equations for the over-all noise figures of cascaded amplifiers are important; they show that the over-all noise figure can never be less than the noise figure of the first stage. Moreover, if the power gains of the stages are comparatively large, or if the noise figures of the separate stages are close to unity, the over-all noise figure of the cascade is determined primarily by the first stage of the amplifier. This is a design factor of paramount importance; it shows that the input stage of a high gain amplifier should usually be designed for minimum noise figure rather than maximum gain. A method of reducing the noise figure of the first amplifier stage will be discussed in section (9.10).

Some difficulty is occasionally met in the evaluation of the noise figure of the second stage of the amplifier cascade because there may be some doubt regarding the way in which the plate load conductance of the first stage should be treated. That is, should the plate load conductance $G_L = 1/R_L$ be considered a part of the source conductance G_s, or should it be included with the input circuit loss G_1? It is obvious that the physical situation is unaffected by the method of treatment. The noise figure is governed by physical factors, and the same result should be obtained regardless of the method of treatment. However, the two methods of treating G_L will lead to different values for F_b, but this will be exactly counterbalanced by a change in W_a, so that F_{ab} is unchanged.

9.10. Available Mid-Band Power Gain

When computing the over-all noise figure of an amplifier cascade, it was shown in the preceding section that the result is partially governed by the power gains of the amplifiers. A method of computing this power gain is developed here.

The *available power gain* of an amplifier is defined as

$$W = \frac{\text{maximum power delivered to the load resistance}}{\text{available power from the signal source}}$$

$$= \frac{P_L\,(\text{max})}{P_s\,(\text{max})}$$

If the signal source supplies a signal current $\bar{I}_s$ to a conductance G_s, the available signal power at the amplifier input is

$$P_s\,(\text{max}) = \frac{I_s^2}{4G_s}$$

This current flows through the input impedance Z_i of the amplifier and produces a grid voltage $\bar{E}_g$. Because

$$Z_i = \frac{1}{G_s + G_1 + G_T}$$

then
$$\bar{E}_g = \bar{I}_s Z_i = \bar{I}_s \frac{1}{G_s + G_1 + G_T} \quad \text{rms volts}$$

The output voltage across the amplifier load is then

$$\bar{E}_o = A\bar{E}_g$$

where A = voltage amplification. Therefore, substituting for the grid voltage,

$$\bar{E}_o = A\bar{I}_s \frac{1}{G_s + G_1 + G_T}$$

The power output is then

$$P_L = \bar{E}_o^2(G_L + G_g) = A^2\bar{I}_s^2 \frac{G_L + G_g}{(G_s + G_1 + G_T)^2}$$

It was shown in chapter 4 that the voltage amplification of a grounded cathode amplifier in the reference case was

$$A_r = g_m R = \frac{g_m}{G_p + G_L + G_g}$$

where $$G_p = \frac{1}{r_p}; \qquad G_L = \frac{1}{R_L}; \qquad G_g = \frac{1}{R_g}$$

Thus the power gain of the stage is

$$W = g_m^2 \frac{4(G_L + G_g)G_s}{(G_p + G_L + G_g)^2(G_s + G_1 + G_T)^2}$$

The power output and the power gain will be a maximum when $G_p = G_L + G_g$. Hence

$$W = \frac{P_L\,(\max)}{P_s\,(\max)} = g_m^2 \frac{G_s r_p}{(G_s + G_1 + G_T)^2}$$

or $$W = \mu g_m \frac{G_s}{(G_s + G_1 + G_T)^2} \tag{9.34}$$

9.11. Optimum Source Conductance

It has been shown that the noise figure for an isolated amplifier stage has the form:

$$F = 1 + \frac{G_1 + \gamma G_T}{G_s} + \frac{R_{eq}}{G_s}(G_1 + G_s + G_T)^2$$

This equation clearly indicates that the noise figure is a relatively complex function of the signal source conductance G_s. This fact suggests that there might be some optimum value of G_s that would make the noise figure a minimum. This proposition can be investigated by differentiating F with respect to G_s, setting the result equal to zero, and solving for G_s. This is the standard minimizing procedure, and yields

$$\frac{\partial F}{\partial G_s} = 0 = -\frac{G_1 + \gamma G_T}{G_s\,(\text{opt})^2}$$
$$+\frac{R_{eq}}{G_s\,(\text{opt})^2}\{2G_s\,(\text{opt})\,[G_1 + G_s\,(\text{opt}) + G_T] - [G_1 + G_s\,(\text{opt}) + G_T]^2\}$$

Solve this equation for G_s (opt) and the result is

$$G_s\,(\text{opt}) = \sqrt{\frac{G_1 + \gamma G_T}{R_{eq}} + (G_1 + G_T)^2} \tag{9.35}$$

This value of source conductance will produce a minimum noise figure for a grounded cathode amplifier stage.

In most cases the source is connected to the amplifier input through some sort of impedance transforming device. As a result, it is possible to use the impedance transforming properties of this circuit to adjust the source conductance to the optimum value.

9.12. Transistor Noise[5]

Noise in bulk semiconductors is larger than that predicted by the thermal noise formula given earlier. Moreover, unlike thermal noise the spectrum is not constant but varies inversely with frequency according to a relationship that can be approximated by K/f^n, where f = frequency; and n = constant having a value of about 1.1 or 1.2. This $1/f$ law apparently holds rather exactly up to frequencies of the order of about 50 kc, but at higher frequencies, departures from this variation are common.

The *excess noise* in semiconductors, which is the noise power above that predicted by the thermal noise formula, has certain special characteristics:[5]

(1) The excess noise is not especially dependent upon temperature.

(2) The excess noise increases approximately as the square of the average current.

[5] R. L. Petritz, "On the Theory of Noise in P-N Junctions and Related Devices," *Proc. IRE,* vol. 40, November, 1952, pp. 1440–1456 (contains an extensive bibliography); E. Keonjian, J. S. Schaffner, "An Experimental Investigation of Transistor Noise," *Proc. IRE,* vol. 40, November, 1952, pp. 1456–1460; H. C. Montgomery, "Transistor Noise in Circuit Applications," *Proc. IRE,* vol. 40, November, 1952, pp. 1461–1471; also H. C. Montgomery, "Electrical Noise in Semiconductors," *Bell System Tech. J.,* vol. 31, September, 1952, pp. 950–975.

It has been observed that point contact transistors are *noisier* than junction transistors by a considerable margin.

The noise of a transistor is presently represented by adding two noise voltage sources to the transistor equivalent circuit as shown in figure (9.8). The two generators have rms voltages of $\bar{E}_{n\varepsilon}$ and $\bar{E}_{nc}$; one is located in series with the emitter, while the other is in series with the collector. The remaining elements of the transistor equivalent circuit are assumed to be noiseless.

If the two equivalent noise generators were truly independent sources, their outputs would be uncorrelated and could be added directly by mean square addition. However, they *do* appear to be correlated, and the addition of their contributions to the total noise must be made with due attention to this correlation. Addition with attention to correlation is designated by a circled plus sign, $\oplus$.

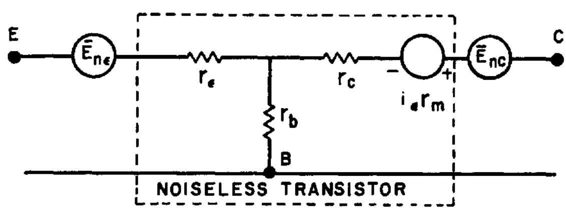


Fig. 9.8. Transistor equivalent circuit showing fictitious equivalent noise sources $E_{n\varepsilon}$ and E_{nc}.

The mechanism of noise production in transistors will not be discussed here. It appears that this subject is not throughly understood at the present time. However, it has been theorized (see Montgomery articles) that variations in the concentration of the minority carrier (electrons in *p*-material, holes in *n*-material) are distributed throughout the material, causing fluctuations in the conductivity of the material that modulate the average or bias current and produce a noise voltage. The activity of the minority carrier concentrations is believed to be modified by some unknown local influence. Noise correlation measurements support this viewpoint as well as the fact that the noise is dependent upon the average current.

A discussion of this subject is beyond the scope of this text and interested readers should consult the footnote references for more details.

9.13. Noise Figures of Transistor Amplifiers

The noise figure of a circuit was defined previously as $F =$ noise figure, or

$$F = \frac{(S/N) \text{ in the ideal noiseless case}}{(S/N) \text{ in the actual case}}$$

where $$S/N = \text{signal-to-noise ratio} = \frac{\text{available signal power}}{\text{available noise power}}$$

In some special cases, such as the one to be discussed in this section, the available signal power is the same in both the noiseless and actual cases. Thus the noise figure can be expressed as

$$F = \frac{N\,(\text{act})}{N\,(\text{min})} = \frac{\text{actual available noise power}}{\text{available noise power in the ideal case}}$$

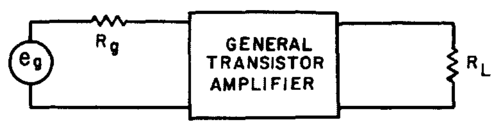


Fig. 9.9. General transistor amplifier.

For example, consider the case of a general transistor amplifier supplied by a signal source E_g of internal impedance R_g and having a connected load R_L. The general circuit connections are shown in figure (9.9). The total noise measured across the load will be contributed by two sources: (1) thermal noise in R_g; (2) noise from the transistor amplifier. Clearly then, if the transistor is noiseless, the available noise power at the output will be the thermal noise in R_g multiplied by square of the voltage amplification of the amplifier. That is,

$$N\,(\text{min}) = \frac{2kT\beta R_g/\pi}{4R_L} A^2 \tag{9.36}$$

Under the same circumstances, the available signal power is

$$S\,(\text{min}) = \frac{E_g^2}{4R_L} A^2 \tag{9.37}$$

Hence $$\left(\frac{S}{N}\right)_{\text{min}} = \frac{E_g^2}{2kT\beta R_g/\pi} \tag{9.38}$$

Under actual conditions the noise power measured across R_L will be greater than the minimum value because of the noise introduced by the transistor amplifier. Thus if a new parameter is defined as

$$\bar{E}_{nc}^2 = \text{mean square noise voltage across } R_L$$

then the actual available noise power is

$$N\,(\text{act}) = \frac{\bar{E}_{nc}^2}{4R_L} \tag{9.39}$$

and the corresponding available signal power is the same as before as

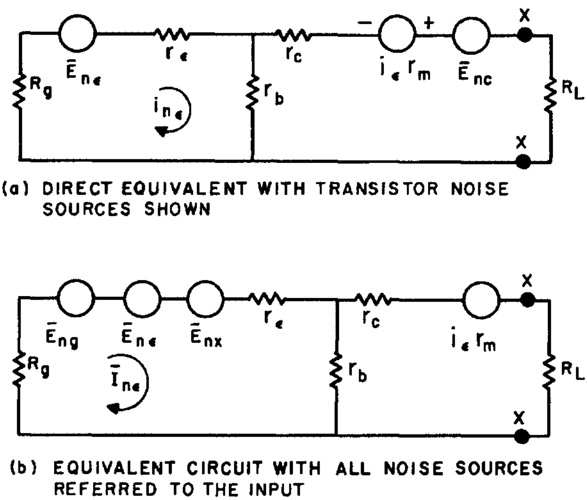


Fig. 9.10. Grounded base transistor amplifier; noise equivalent circuits.

given in equation (9.37). Hence

$$\left(\frac{S}{N}\right)_{\text{act}} = \frac{\bar{E}_g^2}{\bar{E}_{nc}^2} A^2 \tag{9.40}$$

Consequently, the noise figure of the transistor amplifier is

$$F = \frac{(S/N)_{\text{min}}}{(S/N)_{\text{act}}} = \frac{\bar{E}_{nc}^2}{2kT\beta R_g A^2/\pi} \tag{9.41}$$

This equation is perfectly general and can be used to compute the noise figure of any transistor amplifier. It is necessary only to evaluate $\bar{E}_{nc}$ and A for each amplifier configuration. For example, consider the case of a grounded base amplifier. The circuit diagram with equivalent noise source is shown in figure (9.10a). As in the case of vacuum tube amplifiers it is convenient to refer all noise sources to the input circuit. There $\bar{E}_{n\varepsilon}$ is already in the input circuit, and the noisy resistor R_g has been replaced by a noiseless resistor R_g in series with a noise voltage source $\bar{E}_{ng} = \sqrt{2kT\beta R_g/\pi}$. This is shown in figure (9.10b).

The noise generator $\bar{E}_{nc}$ in the collector circuit has been replaced by another noise source $\bar{E}_{nx}$ in the emitter circuit of figure (9.10b). If the two generators $\bar{E}_{nc}$ and $\bar{E}_{nx}$ are to be actually equivalent, they must produce the same open circuit voltage at terminals X-X. The open circuit voltage produced by $\bar{E}_{nc}$ is easily computed from figure (9.10a) to be

$$\bar{E}_{oc} = \bar{E}_{nc}$$

The open circuit voltage produced by $\bar{E}_{nx}$ is computed from figure (9.10b) by replacing $\bar{E}_{ng}$ and $\bar{E}_{n\varepsilon}$ by their internal impedances, short circuits in this case. Then open the collector circuit and write the circuit equations. Under these conditions

$$\bar{E}_{oc} = \bar{I}_{n\varepsilon}(r_b + r_m) \tag{9.42}$$

and

$$\bar{I}_{n\varepsilon} = \frac{\bar{E}_{nx}}{R_g + r_b + r_\varepsilon} \tag{9.43}$$

Therefore

$$\bar{E}_{oc} = \bar{E}_{nx}\frac{r_b + r_m}{R_g + r_b + r_\varepsilon} \tag{9.44}$$

Equate the two expressions for $\bar{E}_{oc}$. Solve for $\bar{E}_{nx}$.

$$\bar{E}_{nx} = \bar{E}_{nc}\frac{R_g + r_b + r_\varepsilon}{r_b + r_m} \tag{9.45}$$

Thus the total noise voltage in the output will be

$$\bar{E}_{nc} = \sqrt{\bar{E}_{ng}^2 + (\bar{E}_{n\varepsilon}^2 \oplus \bar{E}_{nx}^2)}\, A \tag{9.46}$$

where $A =$ voltage amplification of the amplifier. If this value for $\bar{E}_{nc}$ and the corresponding expressions for $\bar{E}_{ng}$, $\bar{E}_{n\varepsilon}$, and $\bar{E}_{nx}$ are substituted into the general noise figure formula, the result for the grounded base connection is

$$F_{gb} = 1 + \frac{\pi}{2kT\beta R_g}\left[\bar{E}_{n\varepsilon}^2 \oplus \bar{E}_{nc}^2\left(\frac{R_g + r_b + r_\varepsilon}{r_b + r_m}\right)^2\right] \tag{9.47}$$

A similar analysis applied to the other two types of transistor amplifiers yields

$$F_{g\varepsilon} = 1 + \frac{\pi}{2kT\beta R_g}\left[\bar{E}^2_{n\varepsilon}\left(\frac{R_g + r_b + r_m}{r_m - r_\varepsilon}\right)^2 \oplus \bar{E}^2_{nc}\left(\frac{R_g + r_\varepsilon + r_b}{r_m - r_\varepsilon}\right)^2\right] \tag{9.48}$$

$$F_{gc} = 1 + \frac{\pi}{2kT\beta R_g}\left[\bar{E}^2_{n\varepsilon}\left(\frac{R_g + r_\varepsilon + r_b}{r_c}\right)^2 \oplus \bar{E}^2_{nc}\left(\frac{R_g + r_b}{r_c}\right)^2\right] \tag{9.49}$$

where $\oplus$ indicates that addition is made with attention to any correlation that might exist between $\bar{E}_{n\varepsilon}$ and $\bar{E}_{nc}$.

The fact that the noise figure is independent of R_L is important. Since F does depend upon R_g, a minimum noise figure can be obtained by using a specific value for R_g. The proper value for R_g can be found by the usual minimizing procedure. After using the usual transistor approximations, the results of this calculation are:

(1) for the grounded base and grounded emitter amplifiers

$$R_g\,(\text{opt}) \doteq \sqrt{(r_\varepsilon + r_b)^2 \oplus \frac{\bar{E}^2_{n\varepsilon}}{\bar{E}^2_{nc}}(r_m + r_b)^2} \tag{9.50}$$

(2) for the grounded collector amplifier

$$R_g\,(\text{opt}) \doteq \sqrt{r_b^2 \oplus \frac{\bar{E}^2_{n}}{\bar{E}^2_{nc}}\,r_c^2} \tag{9.51}$$

The minimum in the noise figure caused by variations in R_g is very broad, so that the dependence of the noise figure on R_g is not critical.

PROBLEMS

9.1. Compute the rms noise voltage and current in a 10,000 ohm resistor at a temperature of 20°C. The noise voltage is measured with an instrument having a noise bandwidth of 5 mcps.

9.2. Compute R_{eq} for a pentode connected 6SJ7.

9.3. Compute R_{eq} for a 6J5.

9.4. Compute the rms induced grid noise current in a 6AC7 and 6AK5 at 60 mcps, at room temperature, and a noise bandwidth of 5 mcps.

9.5. Evaluate the noise bandwidth of

(a) A two-stage connection of synchronous single tuned amplifiers having an over-all bandwidth of 3 mcps.

(b) A staggered pair adjusted to give maximal flatness and a bandwidth of 3 mcps.

9.6. Determine the noise figure of a pentode connected 6AK5 amplifier tube with a single tuned input circuit. The following data apply: $C_{in} = 4.0$ $\mu\mu$f; $g_m = 4500$ μmhos; $R_L = 2000$ ohms; $C_w = 2.0$ $\mu\mu$f; $r_p = 400{,}000$ ohms; $f_o = 30$ kc; $C_o = 3.0$ $\mu\mu$f; $R_s = 450$ ohms. The bandwidth of the input circuit must be at least 6 mcps.

9.7. Compute the optimum source conductance for the amplifier of problem (9.6) and then determine the new noise figure. Comment upon the improvement.

9.8. Calculate the available mid-band power gain of the amplifier of problem (9.6).

9.9. Compute the over-all noise figure of a two-stage 30 mcps amplifier using 6AK5 tubes for which the values given in problem (9.6) apply. The bandwidths of the two single tuned coupling circuits are not less than 6 mcps.

9.10. Repeat problem (9.9), using the optimum source conductance for the input circuit.

9.11. Repeat problem (9.6) for a triode connected 6AK5.

9.12. Repeat problem (9.7) for a triode connected 6AK5.

9.13. Repeat problem (9.9) for a triode connected 6AK5.

9.14. Summarize in words your conclusions as a result of the computations made in problems (9.6), (9.7), (9.9), (9.11), (9.12), and (9.13).

Chapter 10

NEGATIVE RESISTANCES AND CLASS A OSCILLATORS

An *oscillator* is a circuit that converts d-c power into a-c or signal power. Amplifiers accomplish the same effect, but the energy conversion process is under the control of some input or excitation signal. Ordinarily, no excitation or input is required in the operation of an oscillator.

Oscillator circuits are found in all radio, radar, and television transmitters and receivers as well as in carrier telephony systems. Indeed, there are only a few systems that do not require the use of oscillators.

Whenever an analysis is made of an electronic circuit capable of producing a sustained oscillation, it is found that the circuit exhibits a *negative resistance* as a part of its static or dynamic characteristics. It may not be possible to prove that this is exclusively true in every instance. However, whenever a sustained oscillation occurs, a negative resistance invariably appears. Thus basic oscillator theory can be presented from this standpoint in a way that will apply to all electronic devices, regardless of whether they are gas or vacuum tubes, transistors, or feedback circuits.

Oscillators are classified in many different ways and each system has its own merits. In this book the oscillator is assumed to operate either in class A or in the switching mode. Only class A oscillators are covered in this chapter.

10.1. Theory of Negative Resistance Oscillators

The essential components of a negative resistance oscillator are shown in figure (10.1). The principal parts are (1) a negative resistance circuit or device, and (2) some sort of connected load circuit. The negative resistance component is represented by an equivalent circuit consisting of a variable resistor R, which can vary both in magnitude and sign, in parallel with a current source $i(t)$. This current source supplies the initial excitation necessary to produce the oscillation and can be disconnected when oscillation commences. Although it is assumed to be constant in this discussion, it is an easy matter to prove that it can have *any* time dependence, just so long as it actually exists momentarily. In an actual circuit this current source might represent a current impulse caused by thermal agitation or some other circuit unbalance of transient character.

The integrodifferential equation for the circuit of figure (10.1) is

$$C\frac{de}{dt} + \left(\frac{1}{R_L} + \frac{1}{R}\right) e + \frac{1}{L}\int e\, dt = i(t) \tag{10.1}$$

Assume that the capacitor is initially uncharged and that there is no initial current through the coil. Using the methods given in chapter 2, it is a simple matter to write the Laplace transform of equation (10.1)

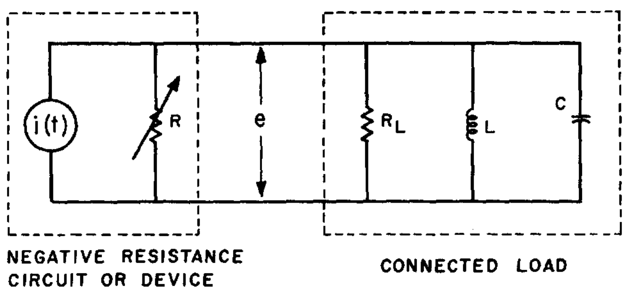


Fig. 10.1. Essential components of a negative resistance oscillator.

and solve the result for the transform response voltage $E(s)$. Thus

$$E(s) = \frac{I(s)}{C} \cdot \frac{1}{s^2 + (R_L + R)s/R_LRC + 1/LC} \tag{10.2}$$

so that the characteristic equation is

$$s^2 + \frac{R_L + R}{R_LR} \cdot \frac{s}{C} + \frac{1}{LC} = 0 \tag{10.3}$$

The roots of this equation are the poles of the response transform, and these roots are computed from the quadratic formula to be

$$s_1 = -\frac{1}{2C} \cdot \frac{R_L + R}{R_LR} + j\sqrt{\frac{1}{LC} - \left(\frac{1}{2C}\right)^2\left(\frac{R_L + R}{R_LR}\right)^2} \tag{10.4}$$

$$s_2 = -\frac{1}{2C} \cdot \frac{R_L + R}{R_LR} - j\sqrt{\frac{1}{LC} - \left(\frac{1}{2C}\right)^2\left(\frac{R_L + R}{R_LR}\right)^2} \tag{10.5}$$

The poles of the response transforms will be complex conjugates and the time response will be oscillatory only if the quantity under the radical is positive. This is a necessary condition that must be fulfilled if oscillations of any type are to be produced.

There is another important condition. If the real part of the pole is positive, the amplitude of the oscillation will increase as a function of time. If the real part is negative, the amplitude of the oscillation will decay with time. A constant amplitude results when the real part of the pole is zero. The explanation of this was given in chapter 2.

With the foregoing in mind you can see that an oscillation is produced only if

$$\left(\frac{1}{2C}\right)^2\left(\frac{R_L+R}{R_LR}\right)^2<\frac{1}{LC}$$

This inequality must be true at all times. However, when the oscillation first starts out, it is small and must build up in amplitude. So the real part of the poles must be positive, so that

$$-\frac{R_L+R}{R_LR}>0$$

This condition can be achieved only if R is negative. That is, if $R=-|R|$, the inequality can be maintained if $R_L>|R|$. During this buildup period the oscillator frequency

$$\omega_0=\sqrt{\frac{1}{LC}-\left(\frac{1}{2C}\right)^2\left(\frac{R_L-|R|}{R_L|R|}\right)^2} \tag{10.6}$$

will vary because $|R|$ will change as a result of nonlinearity in the negative resistance circuit.

Eventually, to obtain an oscillation of constant amplitude, the real part of the pole must vanish. This will occur only if

$$R_L=|R| \tag{10.7}$$

When this transpires, the oscillator frequency becomes constant at

$$\omega_0=\sqrt{\frac{1}{LC}} \tag{10.8}$$

This value is governed by the constants of the load circuit if the negative resistance device or circuit is free from inductive and capacitive components.

The discussion can now be summarized by listing the essential components of oscillators together with their functions:

(1) Negative resistance. This reduces the real part of the poles of the response transform to zero for steady state operation.

(2) A nonlinear circuit element. Either R_L or $|R|$ must be variable so that the oscillation can build up to some specified value and then level off.

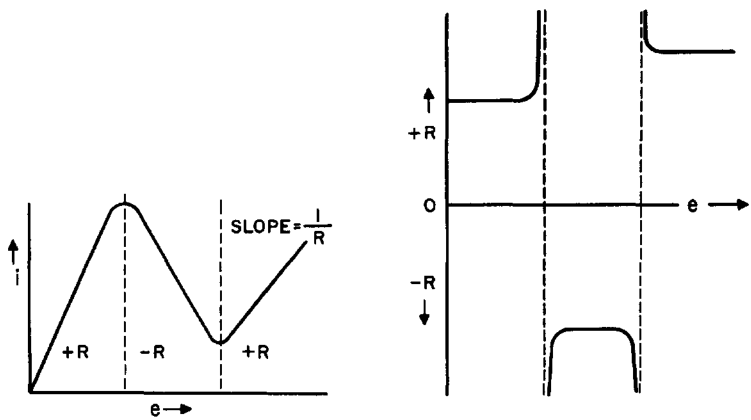
(3) Frequency controller. This function is usually supplied by the load circuit and acts to keep the oscillator frequency constant under normal conditions.

(4) Initiation of oscillation. Usually thermal agitation or a circuit unbalance will provide the transient fluctuation in current necessary to institute the oscillation.

It seems clear that much of the discussion of oscillators will revolve about methods of producing negative resistance. The nature of the negative resistance characteristic so produced is also of considerable interest, and is treated in the next two sections.

10.2. Types of Negative Resistance Characteristics

It is found that negative resistance circuits and devices generally exhibit negative resistance characteristics only over a confined range of



(a) CURRENT-VOLTAGE CHARACTERISTIC (b) RESISTANCE VS. VOLTAGE

Fig. 10.2. Characteristics of a short-circuit stable negative resistance device or circuit.

the complete characteristic. The resistance measured at a pair of terminals will generally be positive over a part of the characteristic, then negative, and then positive again. There seems to be two different types of such characteristics as shown in figures (10.2) and (10.3); they are identified as follows: (1) short-circuit stable; (2) open-circuit stable. The origin of this terminology will be apparent later.

At first glance it might appear that there is little difference between the two types of characteristics. However, from figure (c) of both illustrations, in making the transition from $+R$ to $-R$,

(1) R goes through infinity in the short-circuit stable case.

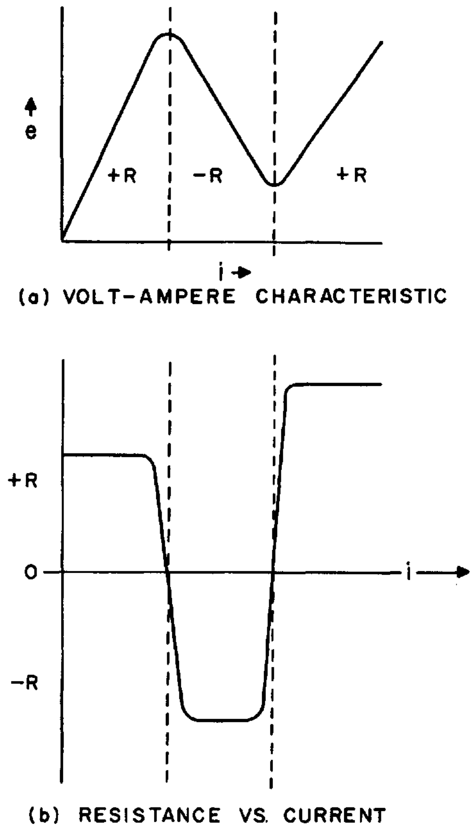
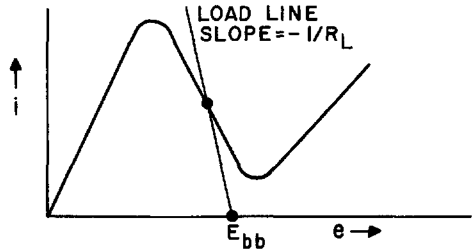


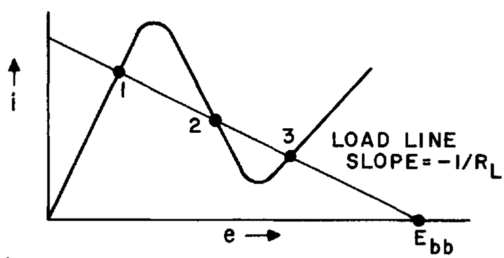
Fig. 10.3. Characteristics of an open-circuit stable negative resistance device or circuit.

(2) R goes through zero in the open-circuit stable case.

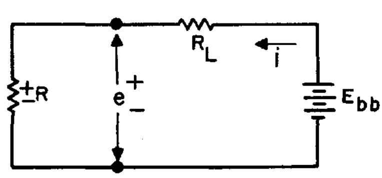
The quiescent operating point of a circuit composed of a series combination of negative resistance, load resistance, and power supply is easily deduced. For example, consider the case of a load resistance R_L, connected in series with a voltage source E_{bb} and a short circuit stable negative resistance device. The operating point of this combination is obtained by drawing the load line for E_{bb} and R_L on the current-voltage characteristic of the negative resistance device and determining the point or points of intersection of the two characteristics. This is illustrated in figure (10.4a) for the case when R_L is less than the



(a) LOAD RESISTANCE LESS THAN THE MAGNITUDE OF THE NEGATIVE RESISTANCE



(b) LOAD RESISTANCE GREATER THAN THE MAGNITUDE OF THE NEGATIVE RESISTANCE



(c) CIRCUIT DIAGRAM

Fig. 10.4. Different operating conditions shown by load line construction on a short-circuit stable characteristic.

magnitude of the negative resistance. One possible operating point results.

When the load resistance is larger than the magnitude of the negative resistance, there are three possible operating points as shown in figure (10.4b). The question of the relative *stability* or *instability* of these operating points is discussed in the next section.

10.3. Stable and Unstable Operating Points

The circuit diagram of a negative resistance oscillator showing d-c connections is given in figure (10.5a). The negative resistance characteristic of the device is assumed to be of the short-circuit stable type as shown in figure (10.5b). Two different load lines are shown corresponding to the two possible relationships between the load resistance and the magnitude of the negative resistance. The variational equivalent circuit is given in figure (10.5c), in which $i(t)$ is an assumed current

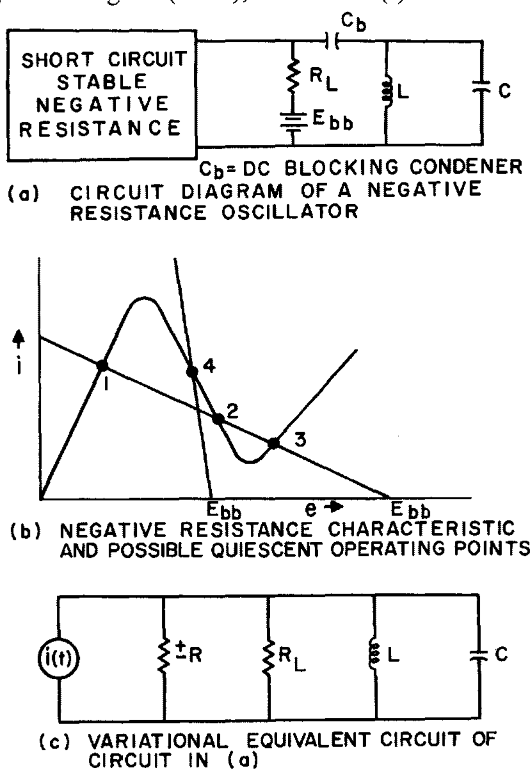


Fig. 10.5. Circuits and quiescent operating points in a short-circuit stable negative resistance circuit.

generator caused by thermal agitation in the circuit. The circuit is of the same form as that in figure (10.1), so that the poles of the response transform are

$$s_{1,2} = -\frac{1}{2C}\cdot\frac{R_L + R}{R_L R} \pm j\sqrt{\frac{1}{LC} - \left(\frac{1}{2C}\right)^2\left(\frac{R_L + R}{R_L R}\right)^2}$$

Now suppose that the load resistance R_L is less than the magnitude of the negative resistance, so that the quiescent operating point is at (4) in figure (10.5b). Under the assumed conditions, the poles of the response transform will fall in the left half of the complex s plane, so that any transient disturbance of any character will eventually die out. Clearly, this is a *stable* operating point.

Now suppose that the circuit is adjusted so that R_L is greater than the magnitude of the negative resistance and that the Q point is at (1) or (3) in figure (10.5b). In this case the resistance of the device is positive and the poles of the response transform fall in the left half of the s plane. The system is again stable.

Under the same operating conditions, assume that the quiescent operation is at point (2) in figure (10.5b). In this case the magnitude of the load resistance is larger than that of the negative resistance, and the poles of the transform response fall in the right half plane. Hence *any* slight and momentary disturbance in the system will institute a transient that will build up exponentially indefinitely until limited by the system. Thus point (2) is an *unstable* operating point.

If operation at point (2) is initially assumed, any momentary current increase or decrease will cause the circuit to become unstable and switch over to operation at one of the two stable points. A positive current impulse will cause the circuit to switch over to operation at point (1); a negative current impulse will cause operation to switch over to point (3).

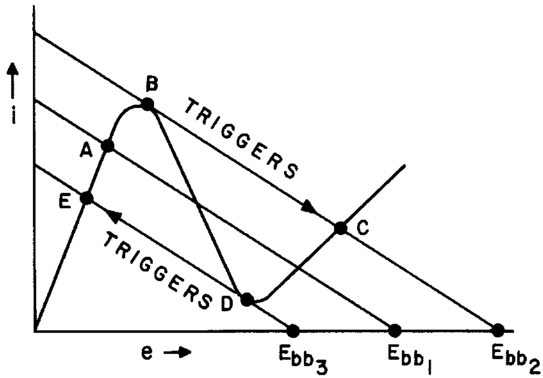
Evidently, a circuit of this type will be stable when the load resistance is zero, so that it is short circuited. Thus it is called a *short-circuit stable* negative resistance characteristic. Instability can be produced in a short-circuit stable negative resistance circuit when the load resistance is greater than the magnitude of the negative resistance.

Exactly converse conclusions regarding the relative sizes of R_L and the negative resistance are obtained by the same analysis if applied to the open-circuit stable case.

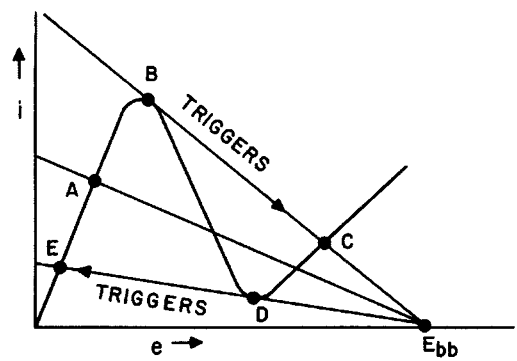
10.4. Methods of Triggering

The case of a series connection of load resistance R_L, power supply, and short-circuit stable negative resistance device was treated in the preceding section. It was shown that there are three possible operating points when R_L is greater than the magnitude of the negative resistance. The operating points in the positive resistance region proved to be stable, while the one in the negative resistance region was unstable. In the latter case, any small change in circuit current or voltage causes operation to switch from the unstable point to one of the stable points. This sudden change in operating point is called *triggering*, or the circuit is said to have been *triggered.*

In any practical case, under quiescent conditions such a circuit will



(a) TRIGGERING BY VARIATIONS IN SUPPLY VOLTAGE



(b) TRIGGERING BY VARIATIONS IN LOAD RESISTANCE

Fig. 10.6. Methods of triggering a short-circuit stable negative resistance circuit.

always be in one of its stable states. This is an obvious conclusion because the fluctuation in *e* or *i* necessary to trigger the circuit from the unstable point will always be produced by some random phenomenon such as thermal agitation.

Assume such a circuit operating at point *A* under quiescent conditions as shown in figure (10.6a). Now suppose that the load line is translated parallel to itself by changing the power supply voltage from E_{bb_1} to E_{bb_2}. This causes the operating point to move from A to B. Point B falls in the negative resistance region and is an unstable operating point. Hence the circuit triggers over to operation at point C. If the power supply voltage is now decreased to E_{bb_3}, the operating point moves to point D. This is also an unstable point and triggering to point E occurs.

Figure (10.6b) shows how triggering can be accomplished by varying the *slope* of the load line by changing the value of the load resistance R_L.

It is also possible to produce triggering by varying the size and shape of the current-voltage characteristic.

Similar results can be deduced in the same way for the open-circuit stable characteristic.

10.5. Oscillator Classification

When a negative resistance circuit has the inherent capability of *self-triggering*, it is an *oscillator* in the strictest sense of the word. Under

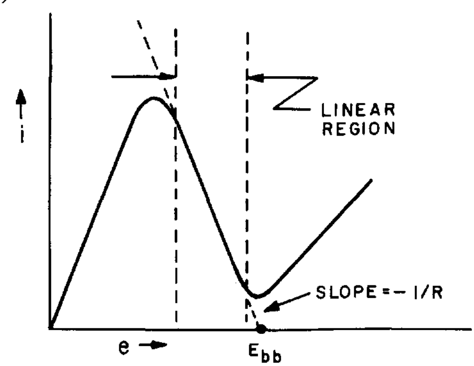


Fig. 10.7. Operating conditions in a linear class A oscillator.

these conditions it is capable of converting the d-c power from the power supply into some repeating waveform or variational signal without the assistance of any externally applied signal. In such cases, the operating point falls into three distinct regions, and a switch type of equivalent circuit is required to describe circuit behavior. Such circuits are covered in Part III, which is devoted to operation in the switching mode.

The discussion in this chapter will be confined to those circuits operating in a nonswitching or class A manner. Operation is at all times in the negative resistance region, so that the device can be represented by a simple equivalent circuit composed of a negative resistance and an equivalent intercept voltage source. Under steady state conditions operation will occur as shown in figure (10.7) so that the intercept voltage is approximately equal to E_{bb}, the power supply voltage. This simplifies oscillator analysis somewhat, because the corresponding equivalent circuit is fundamentally the same circuit as that used in the elementary discussion given in section (10.1).

As long as operation is restricted to the specified region the circuit will be linear. Considerable nonlinear distortion results when operation shifts to the switching mode.

The *linear* oscillators treated in this chapter can be classified in many ways, as follows:

(1) According to the electronic device used
   (a) Vacuum tube oscillators
   (b) Transistor oscillators
   (c) Gas tube oscillators
   (d) Thermistor oscillators
(2) According to function
   (a) Power oscillators (covered in chapter 12)
   (b) Frequency controlling oscillators
(3) According to waveform generated
   (a) Sinusodial
   (b) Square wave
   (c) Sawtooth
   (d) Trapezoidal, and so on
(4) According to the nature of the frequency determining element
   (a) *RLC* resonant circuit
   (b) *RC* circuit
   (c) Piezoelectric
   (d) Magnetostriction
(5) According to the location of the negative resistance
   (a) In the d-c or static characteristic
   (b) In the dynamic characteristic

The last method of classification is preferred here and is explored in some detail in later sections.

10.6. Classification of Negative Resistance Devices

It was stated in the preceding section that oscillators can be classified according to the location of the negative resistance. Thus two main oscillator types can be defined as follows:

(1) *Negative resistance oscillators.* The negative resistance is a part

of the static characteristics of the device or circuit. Such a characteristic arises in two ways:

(a) As an intrinsic property of the electronic device (*dynatron type*).

(b) As a result of particular external circuit connections (*transitron type*).[1]

(2) *Feedback oscillators.* Specific external circuit connections are required to produce a negative resistance in the dynamic characteristics of the complete circuit.

Another group of circuits are called *transit time* oscillators, but these are more appropriately treated in ultrahigh frequency books. Their operation is governed by factors entirely unrelated to the discussion presented here.

The actual devices and circuits that fit into this classification are given in the next two sections.

10.7. Negative Resistance in the d-c Characteristic

Thermistors, arc discharge tubes, and certain magnetrons have a negative resistance as an important part of their static characteristics. Similarly, over a confined range of operating potentials, vacuum tube tetrodes exhibit a negative resistance region caused by secondary emission effects. Therefore all these characteristics can be used with an appropriate tuned circuit to produce dynatron oscillations. The necessary design relationships can be derived directly from the results given in section (10.1).

The tetrode characteristic has little practical use. The magnitude of the negative resistance characteristic is so large and so variable that unreasonably large values of R_L and E_{bb} are required. Also, the region of operation is small and quite nonlinear, so that considerable nonlinear distortion results, and the amplitude of the output is small.

Dynatron oscillations in magnetrons are mainly of academic interest because such operation does not utilize the particular attributes of the magnetron to advantage.

The arc type of oscillator is not of general importance today as a signal source. However, inadvertent oscillation in gas discharges is often a serious practical problem.

[1] The term *transitron oscillator* currently refers to a specific circuit. However, the technique of using external circuit connections to produce a d-c negative resistance is representative of a group of circuits. Thus the term is used in a general sense here, though such terminology is not standard.

Thermistor oscillators have a wide area of potential use. This is especially true in the production of low frequency oscillations. In this type of circuit the thermal inertia of the thermistor replaces the electrical inductance of the tuned circuit. Because this inertia is ordinarily quite large, the *effective* inductance of the thermistor can be extremely large, and low frequency oscillations result.

When special external connections are made about some devices they exhibit a negative resistance in their d-c characteristics. For example, transistors can exhibit a negative resistance at a pair of terminals under certain conditions if the current amplification factor α is greater than 1. In most cases it is necessary to add some resistance to the base; if the

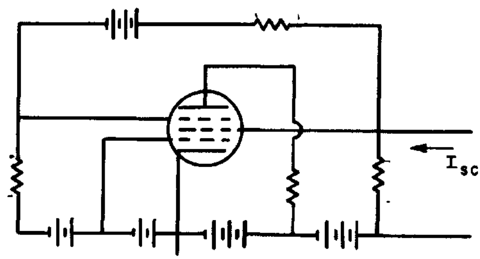
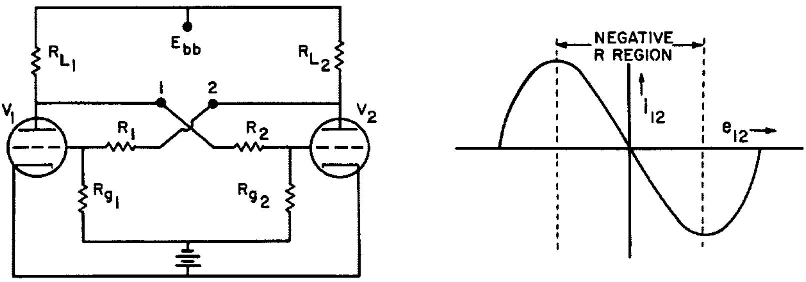


Fig. 10.8. Negative transconductance circuit.

output terminals are short circuited the input resistance will be negative over a part of the characteristic. Similarly, a negative resistance region will appear in the output characteristic if the input terminals are short circuited. Oscillation in the *transitron* sense is then possible if a tuned circuit of the proper characteristics is connected across the pair of terminals exhibiting the negative resistance.

A successful vacuum tube circuit is the original transitron oscillator shown schematically in figure (10.8). This is also called a *negative transconductance oscillator*. It consists of a pentode with the screen and suppressor coupled together and the screen more positive than the plate. The combination of electrode potentials is such that a virtual cathode is formed about the suppressor, and this supplies electrons to both the screen and the plate. If the screen voltage should rise, this increase is coupled over to the suppressor. The suppressor is then more positive than before and it allows more electrons to go to the plate and fewer to the screen. Thus the screen current is reduced; this is evidence of a negative resistance characteristic appearing in the screen circuit at the terminals marked in figure (10.8). Connection of an appropriate tuned circuit at these terminals will result in sustained oscillations.

Another vacuum tube circuit that exhibits a negative resistance as a part of its static characteristics in the transitron sense is shown in figure (10.9). This is the well-known *Eccles-Jordan* circuit, and the negative resistance appears at the marked terminals. This circuit has not been widely used for linear oscillators, but it is extensively used for oscillators operating in the switching mode.



(a) ECCLES-JORDAN CIRCUIT (b) CURRENT-VOLTAGE CHARACTERISTIC

Fig. 10.9. Eccles-Jordan negative resistance circuit.

10.8. The Method of Isoclines

In section (10.1) it was shown that a negative resistance oscillator necessarily requires a nonlinear circuit element, otherwise an oscillation could not build up in amplitude and then level off at a fixed value. It was shown that this required complex conjugate poles in the right half of the s plane during the build up period. As the oscillation builds up, these poles must move to the left until they are located on the imaginary axis when steady state operation is achieved. The system is clearly nonlinear during the period in which the poles must move. The linear circuit analysis presented in section (10.1) is useful only during the steady state period when the pole locations are fixed. When more information is required about the buildup period, nonlinear differential equations must be solved. One practical method, called the *method of isoclines*,[2] is outlined here.

[2] Ph. leCorbeiller, "The Nonlinear Theory of the Maintenance of Oscillations," *Proc. IRE*, vol. 23, 1935, pp. 361–378; Wm. A. Edson, *Vacuum Tube Oscillators*, John Wiley & Sons, Inc., New York, 1953, pp. 43–51; Wm. B. Wrigley, "A Note on Moving Poles in Nonlinear Oscillating System," *Proc. IRE*, vol. 41, 1953, pp. 774–777.

The method of isoclines is basically graphical in nature and suffers from the disadvantages characteristic of such methods. However, it *will* work in practical cases, and this is a strong recommendation. It holds a great deal of promise in the analysis of negative resistance transistor oscillators.

The essential circuit configuration of a negative resistance oscillator is shown in figure (10.10). The node equation for the circuit is

$$i_n + Ge + i + C\frac{de}{dt} = 0 \tag{10.9}$$

The current through the negative resistance device is some complicated

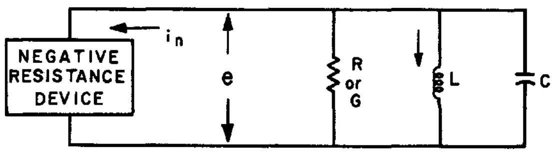


Fig. 10.10. Negative resistance oscillator circuit.

function of the voltage e. That is,

$$i_n = f(e) \tag{10.10}$$

Typical graphical representations of $f(e)$ were shown earlier in figures (10.2a) and (10.3a). The network differential equation now becomes

$$f(e) + Ge + i + C\frac{de}{dt} = 0 \tag{10.11}$$

It is desirable to eliminate the time variable. This can be done with the aid of the following derivative identity:

$$\frac{de}{dt} = \frac{de}{di}\cdot\frac{di}{dt} \tag{10.12}$$

This can be simplified somewhat because

$$e = L\frac{di}{dt} \tag{10.13}$$

and the current derivative is

$$\frac{di}{dt} = \frac{e}{L} \quad \text{so that} \quad \frac{de}{dt} = \frac{e}{L}\cdot\frac{de}{di} \tag{10.14}$$

and the circuit differential equation in (10.11) reduces to

$$f(e) + Ge + i + \frac{C}{L}e\frac{de}{di} = 0 \tag{10.15}$$

The time variable has been eliminated from the equation.

Define the following parameter:

$$Z_c = \sqrt{\frac{L}{C}} \tag{10.16}$$

so that the differential equation in (10.15) is

$$f(e) + Ge + i + \frac{1}{Z_c^2} e \frac{de}{dt} = 0 \tag{10.17}$$

Now transform the variable in this equation by defining a new variable such that

$$e = Z_c u \tag{10.18}$$

Because Z_c is a constant of the circuit having the dimensions of resistance, u has the dimensions of current. Substitute equation (10.18) into (10.17) to obtain

$$[f(Z_c u) + GZ_c u] + i + u\frac{du}{di} = 0 \tag{10.19}$$

Because both G and Z_c are constants of the tuned circuit, the bracketed term in this equation is a function of the variable u. That is,

$$F(u) = f(Z_c u) + GZ_c u \tag{10.20}$$

or

$$F(u) = f(e) + Ge \tag{10.21}$$

Hence equation (10.19) becomes

$$F(u) + i + u\frac{du}{di} = 0 \tag{10.22}$$

or finally,

$$\frac{di}{du} = -\frac{u}{i + F(u)} \tag{10.23}$$

Although the reason for this manipulation is probably not clear now, it should be shortly. For the moment it is sufficient to say that equation (10.23) has great importance in the graphical construction described in the next paragraph.

The function $-F(u)$ is to be plotted on a system of coordinates having i as the ordinate and u as the abscissa. This is accomplished as follows:

(1) Carefully plot the current-voltage characteristic of the negative resistance device. A representative case is shown in figure (10.11a). This curve is $f(e)$.

(2) Locate the approximate center of the negative resistance region.

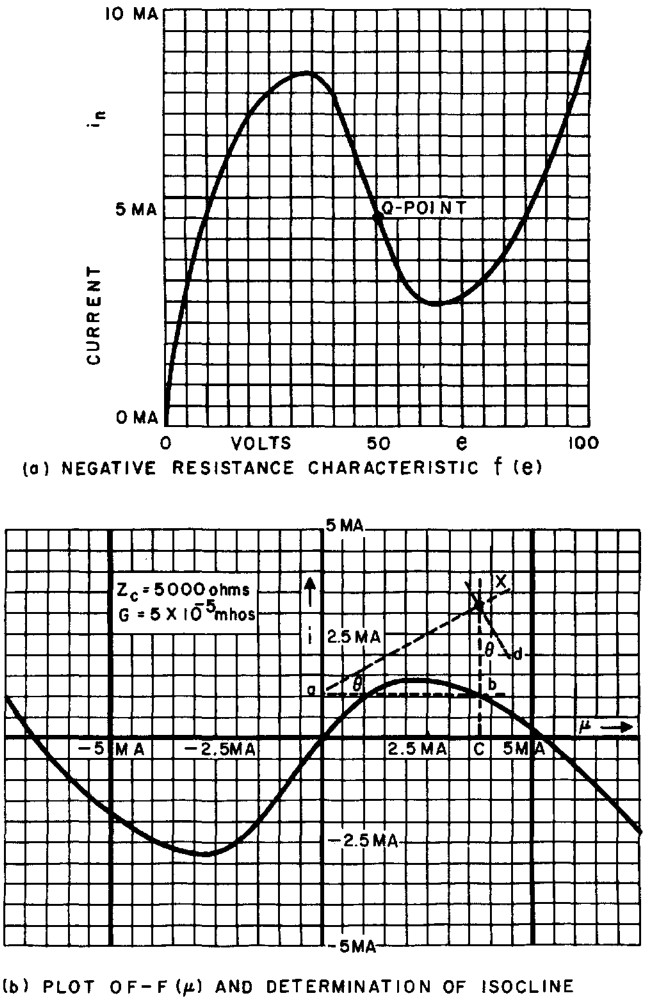


Fig. 10.11. Graphical constructions required in the preliminary steps of the method of isoclines.

This locates the Q point. Measure all voltage increments from this point.

(3) Assume a series of values for Δe. For each such value assumed,

(a) Compute $\Delta u = \Delta e/Z_c$.

(b) Determine $\Delta f(e) = \Delta i_n$ from the curve drawn in (1).

(c) Calculate $G\Delta e$.

(d) Compute $\Delta F(u) = \Delta f(e) + G\Delta e$

(e) Plot $-F(u)$ against u as shown in figure (10.11b).

Use the same scale calibration for both the ordinate and the abscissa.

Some additional graphical construction is now required, as follows:

(1) Draw a vertical line, as at c in figure (10.11b). This intersects the $-F(u)$ characteristic at a point marked as b.

(2) From b project horizontally to the vertical axis and locate the point a.

(3) Take *any* point x on the original vertical line and draw in the line a–x.

(4) Construct the perpendicular to a–x at point x as shown by the line d–x. Thus d–x is tangent to an arc drawn from a as center and through x. Then d–x is an *isocline.*

From the preceding construction, which is shown in figure (10.11b), it is clear that

$$a\text{–}b = u; \qquad c\text{–}b = -F(u)$$

$$c\text{–}x = i; \qquad b\text{–}x = c\text{–}x - c\text{–}b = i + F(u)$$

The slope of the isocline d–x is

$$(\text{slope of isocline}) = -\frac{1}{\tan\theta}$$

but

$$\tan\theta = \frac{b\text{–}x}{a\text{–}b}$$

so

$$(\text{slope of isocline}) = -\frac{a\text{–}b}{b\text{–}x} = -\frac{u}{i + F(u)} \tag{10.24}$$

Clearly, the slope of the isocline is equal to di/du as given by equation (10.23) and (10.24). In other words, the isocline is the slope of the i vs. u characteristic at any point x.

An isocline diagram is constructed by simply drawing a number of vertical lines for arbitrary values of u. Then locate points b and a for each such line. From each point a use a compass to draw a series of short arcs intersecting the corresponding vertical line. These arcs are all isoclines, and the over-all result is an isocline diagram as shown in figure (10.12).

The isocline diagram is used to construct a *cyclogram*, as described in the next section, and the cyclogram is useful in analyzing and understanding the behavior and characteristics of the oscillator during both the buildup and steady state periods.

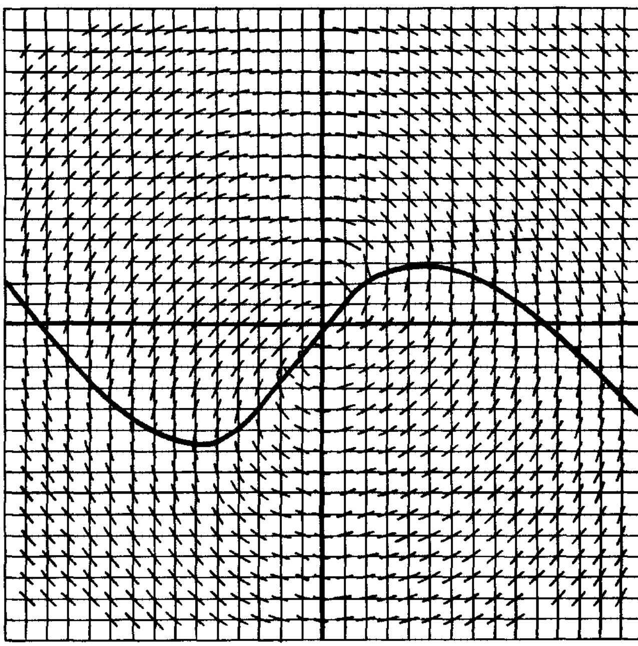


Fig. 10.12. Isocline diagram drawn from the plot of $-F(\mu)$ in figure (10.11).

10.9. Cyclograms

The entire performance of an oscillator starting from any assumed initial conditions, through buildup and into the steady state can be shown by curves called *cyclograms*. Cyclograms are plots of the oscillator current and voltage as a function of time. They are plotted on isocline diagrams and appear as shown by a typical example in figure (10.13).

The direction of rotation on this curve as time increases can be determined from the expression

$$L\frac{di}{dt} = e = Z_c u$$

In terms of increments this is

$$\Delta t = \frac{L}{Z_c} \cdot \frac{\Delta i}{u}$$

Thus a positive increment of time requires either a positive value for u and a positive increment in i, or a negative value for u and a decrement in i. Therefore increasing time corresponds to counterclockwise rotation on the cyclogram.

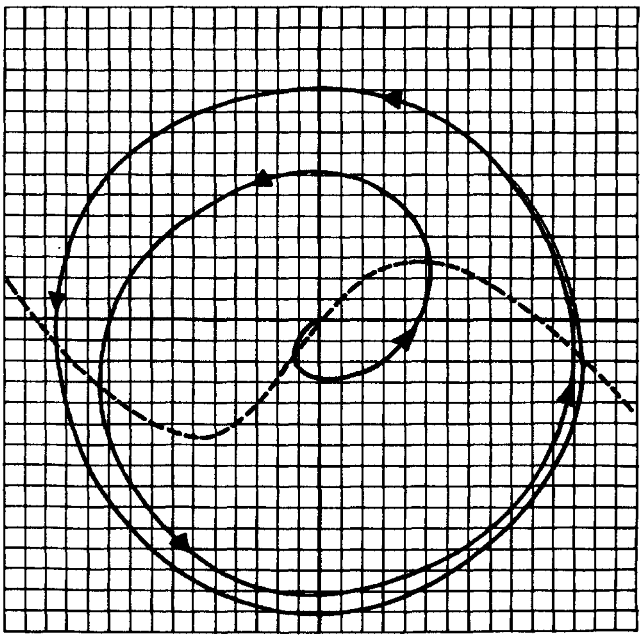


Fig. 10.13. Cyclogram drawn for zero initial conditions from the isocline diagram of figure (10.12).

The cyclogram is constructed rather simply by the following method:

(1) Take any assumed or specified starting point on the isocline diagram.

(2) From this point, which represents the initial conditions, draw a continuous curve in the counterclockwise sense, always staying parallel to the isoclines.

(3) Eventually the curve closes upon itself and the steady state is thereby attained.

A different path of buildup will result from different assumed starting points, but steady state operation will always follow the same curve.

The use of the cyclogram in understanding circuit operation is apparent. The number of cycles required for buildup from any set of initial conditions is readily determined. The waveforms of current and voltage can be determined by calculating the time rate of traverse of the cyclogram. This is worked out in the next paragraph.

The cyclogram is a curve of i against u, where $u = e/Z_c$ as a function of time. An angle ϕ can be defined as

$$\phi = \tan^{-1}\frac{i}{u}$$

and the rate of rotation around the cyclogram is the time rate of change of this angle. That is

$$\omega = \frac{d\phi}{dt} = \frac{u(di/dt) - i(du/dt)}{u^2 + i^2} \tag{10.25}$$

However, from equation (10.14),

$$\frac{di}{dt} = \frac{e}{L} = \frac{uZ_c}{L} = \frac{u}{\sqrt{LC}} \tag{10.26}$$

Also, the original differential equation (10.11) was of the form

$$[f(e) + Ge] + i + C\frac{de}{dt} = 0$$

or

$$F(u) + i + Z_cC\frac{du}{dt} = 0$$

Therefore

$$\frac{du}{dt} = -\frac{i + F(u)}{CZ_c} = \frac{i + F(u)}{\sqrt{LC}} \tag{10.27}$$

Now define the parameter $\omega_0 =$ undamped natural frequency $= 1/\sqrt{LC}$, so that equations (10.26) and (10.27) become

$$\frac{di}{dt} = \omega_0 u; \qquad \frac{du}{dt} = -\omega_0[i + F(u)]$$

Substitution of these relationships into equation (10.25) eventually gives

$$\omega = \omega_0\left[1 + \frac{iF(u)}{u^2 + i^2}\right] \tag{10.28}$$

In high Q cases the second term in equation (10.28) is negligible and $\omega = \omega_0$. The cyclogram becomes nearly circular and the rate of traverse becomes constant at a frequency equal to the undamped natural frequency of the resonant circuit. The oscillator then generates a sinusoidal waveform.

When the second term in equation (10.28) is not negligible, the rate of traverse around the cyclogram is not constant. Moreover, the cyclogram itself is distorted from a circular shape. As a result a very nonsinusoidal waveform is generated. In any case, equation (10.28) can be used to calibrate the rate of rotation around the cyclogram.

10.10. Negative Resistance Produced by Feedback

Feedback oscillators are probably the most common of all class A oscillator types. The basic physical arrangement is shown in figure

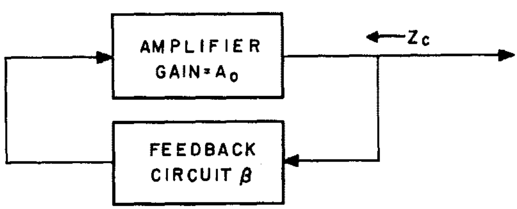


Fig. 10.14. Feedback circuit.

(10.14) where it is understood that the amplifier can use either vacuum tubes or transistors. It was shown in chapter 7 that the output impedance of such a circuit is

$$Z_{out} = Z_c = \frac{Z_o}{1 - \beta A_o} = \frac{R_o \pm jX_o}{1 - \beta A_o}$$

where A_o = open loop voltage gain of the amplifier; β = voltage transfer function of the feedback circuit; Z_o = output impedance of the open loop amplifier. It is clear from this equation, which was derived by using variational currents and voltages, that the output impedance of the feedback amplifier can be made to have a negative resistance component by using positive feedback in an amount sufficient to make βA_o larger than 1. It is also clear that the output impedance will contain reactive terms if the output impedance of the open loop amplifier is reactive or if βA_o is a complex number. Should this occur, the frequency of the oscillations will be determined in part by the reactance component of the output impedance.

10.11. Representative Feedback Oscillator Circuits

Every feedback oscillator, regardless of type, involves the essential components: (1) amplifier; (2) amplitude limiter or controller; (3) feedback circuit; (4) frequency controller.

Either vacuum tube or transistor amplifiers can be used and because of the nonlinearities inherent in such components, they frequently provide amplitude limiting service as well, In other cases some other nonlinear circuit element may be used for limiter service.

The functions of frequency control and feedback are often provided by the same circuit. This is especially true in most vacuum tube oscillators.

In this book feedback oscillators are first classified according to whether the amplifier uses a vacuum tube or transistor. Then, within each of these two groups, the circuits are further subdivided according to the nature of the frequency controlling network. Representative circuits of various types are shown in figures (10.15), (10.16), and (10.17). In the interests of simplicity, all d-c connections have been omitted and only essential a-c components are shown. The proper d-c connections are easily deduced because the same general considerations apply here and for amplifiers of the same type.

In a grounded cathode vacuum tube amplifier with resistance load, there is an inherent 180° phase shift of a sinusoidal signal in its transmission through the amplifier. To make a single tube amplifier of this type operate as an oscillator, it is necessary to produce another 180° of phase shift in the feedback circuit, bringing the total up to 360°. Actually, the phase shift in the amplifier is seldom exactly 180° and the feedback circuit must make up the deficit between whatever it is and the required 360°.

The phase shift problem may be somewhat less involved when certain types of grounded base transistor amplifiers are used, because there is no phase inversion through such circuits. Thus no severe phase shifting requirements are imposed upon the feedback circuit, because no phase shift is required.

It would appear from figure (10.15) that there is a multiplicity of types of vacuum tube oscillators with resonant circuit frequency controllers. However, it is a relatively simple matter to show that all such circuits are basically either Hartley or Colpitts oscillators. For example, consider the tuned plate, tuned grid (TPTG) oscillator of figure (10.15c). For the feedback circuit to produce the required phase

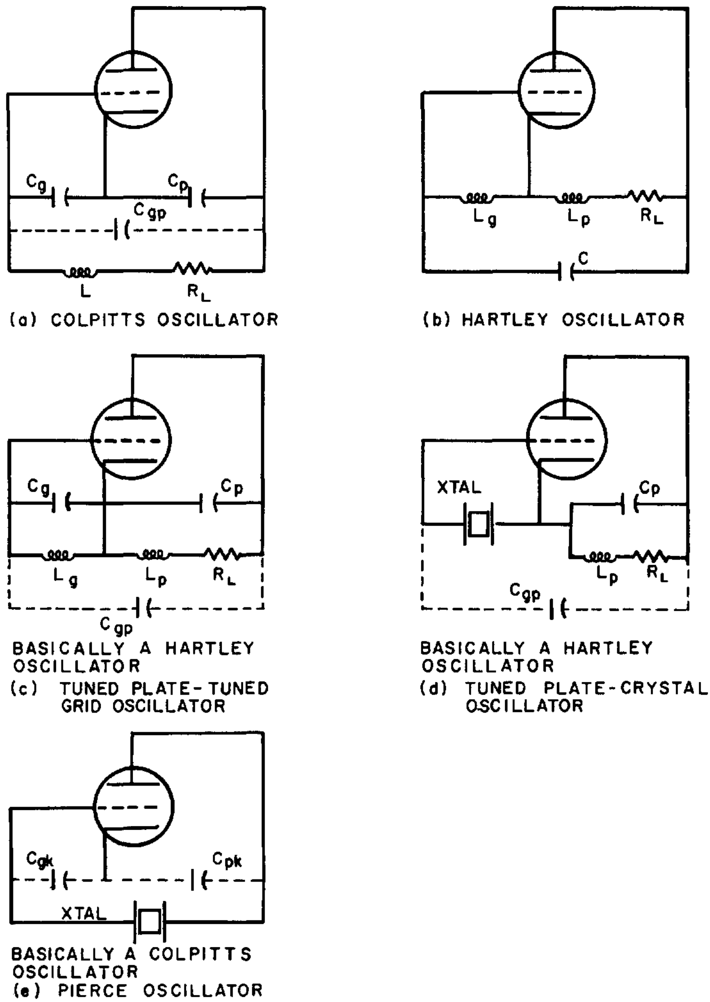


Fig. 10.15. Vacuum tube oscillations with resonant circuit frequency control. A-c paths only; all d-c connections omitted.

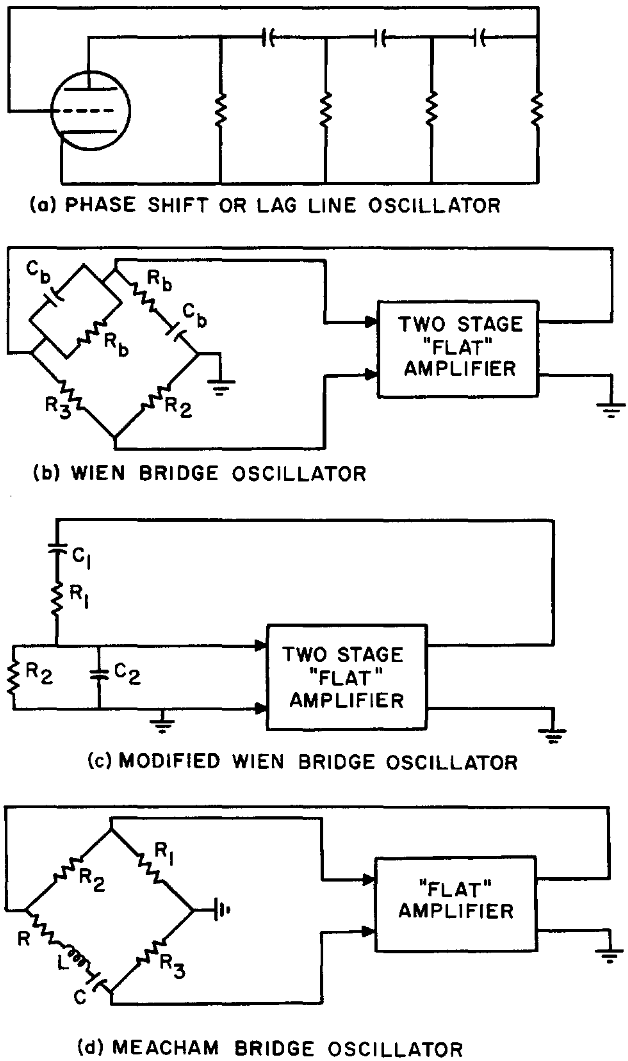
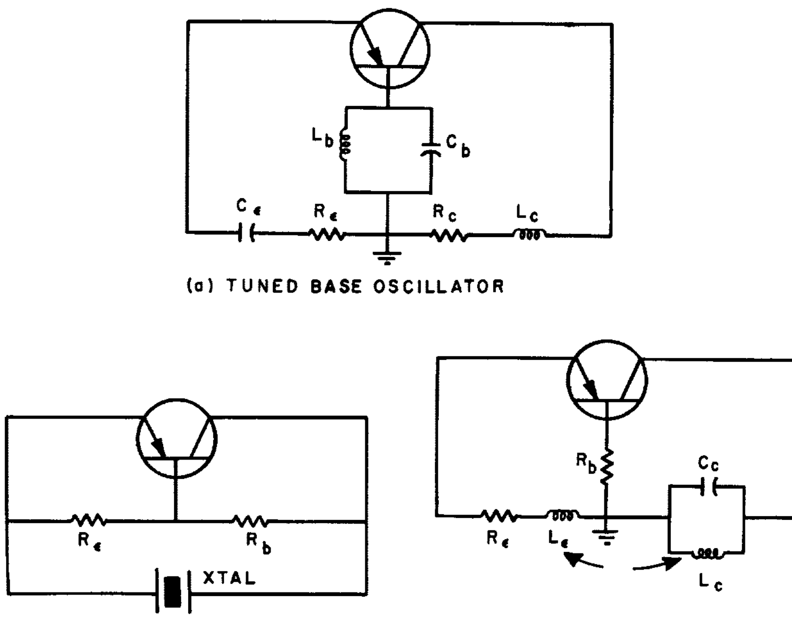


Fig. 10.16. *RC* tuned and bridge type vacuum tube oscillators.

shift, the reactances on both sides of the tap must be of the same type, but opposite to the reactance in parallel with the tapped element. Thus the parallel tuned circuits must appear as effective inductances, thereby requiring the resonant frequencies of these circuits to be higher than the oscillator frequency. Clearly, when these circuits are replaced by inductances the circuit diagram reduces to that of a Hartley oscillator.

The same simplification can be applied to the crystal oscillator of figure (10.15d). By replacing the crystal with a parallel tuned circuit, the equivalent circuit of a crystal in this case, the circuit is obviously



(a) TUNED BASE OSCILLATOR

(b) CRYSTAL OSCILLATOR

(c) TUNED COLLECTOR OSCILLATOR

Fig. 10.17. Representative transistor oscillators; all d-c paths omitted.

the same as that of a TPTG oscillator, and is therefore inherently a Hartley oscillator. A similar analysis will show that the Pierce oscillator is fundamentally a Colpitts circuit.

There are many ways of obtaining the necessary phase shift required for positive feedback without using resonant circuits. Four representative vacuum tube circuits using *RC* circuits for frequency control, were shown in figure (10.16).

10.12. Analysis of Feedback Oscillators

The analysis of feedback oscillators and the derivation of suitable design formulas is a straightforward application of standard techniques and ideas covered in previous chapters. Two of these points are important, as follows:

(1) From chapter 2. The response of a circuit will be oscillatory if the poles of the transfer function are complex conjugates. The oscillation will be sustained with constant amplitude if the real parts of these poles are zero.

(2) From chapter 7. The voltage amplification, or transfer function, of a feedback amplifier with voltage feedback is

$$A_c = \frac{A_o}{1 - \beta A_o}$$

The poles of this function are clearly the roots of the equation $\beta A_o - 1 = 0$.

With these facts in mind the proper procedure for analyzing feedback oscillators can be summarized as follows:

(1) Draw the complete class A equivalent circuit:

(a) Replace bypass capacitors by short circuits.

(b) Replace radio frequency chokes with open circuits.

(c) Show tube or transistor capacitances.

(d) Draw the circuit as an open loop amplifier. That is, if the amplifier uses vacuum tubes, the equivalent plate generator is $g_m E_g$ and the output is E_g.

(2) Compute the over-all amplification of the open loop circuit as a function of the frequency s. This will equal the feedback factor βA_o.

(3) Write the equation $\beta A_o - 1 = 0$. Make any valid approximations.

(4) The final form of this equation should appear as a polynomial in s with unity as the coefficient of the highest power of s.

(5) Compute the roots of this equation.

(a) For oscillations to occur there must be a pair of complex conjugate roots with zero real parts. These two conditions will yield two design equations.

(b) If factoring is difficult, the Routh-Hurwitz or Nyquist criterion will prove helpful.

This procedure is best illustrated by examples. Some relatively simple cases are treated in the next few sections.

10.13. Analysis of a Colpitts Oscillator

Consider the vacuum tube Colpitts oscillator shown in figure (10.18a). The circuit has been redrawn in figure (10.18b) with only the significant

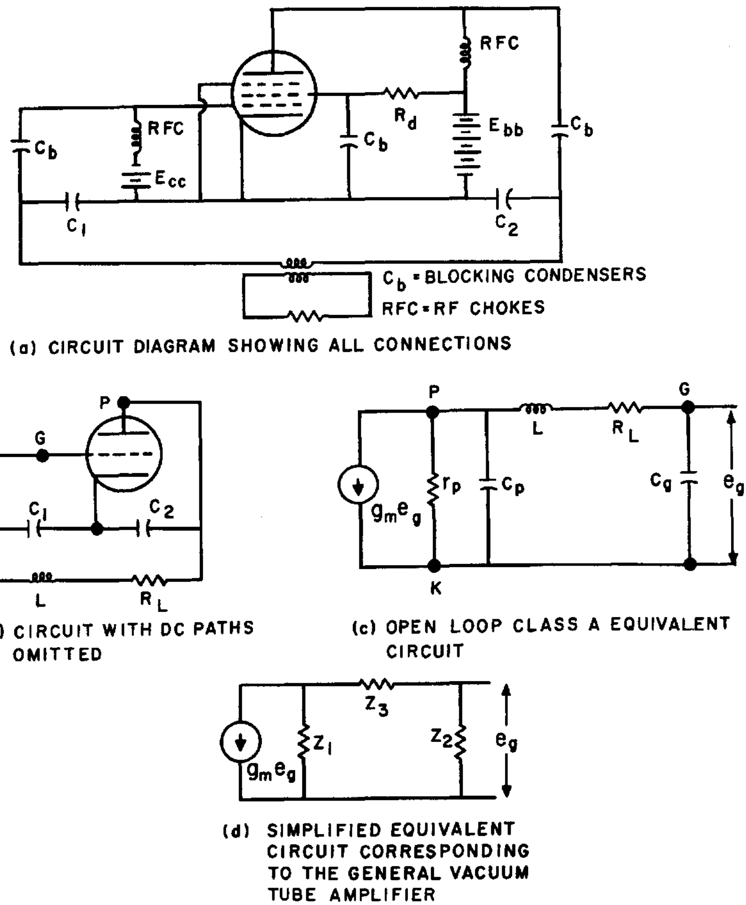


Fig. 10.18. Analysis of a Colpitts oscillator circuit.

signal frequency elements shown. This circuit results from part (1) of the procedure given in the preceding section. Here R_L is used to represent the losses in the oscillator resonant circuit and the power delivered to the load. The class A equivalent circuit is drawn as directed in figure (10.18c) and the external capacitances C_1 and C_2 have been combined with the tube interelectrode capacitances so that $C_g = C_1 + C_{gk}$, and $C_p = C_2 + C_{pk}$. It is assumed that the effect of C_{gp} is negligible at the operating frequency ω_0 of the oscillator. This is not a necessary assumption because the effect of C_{gp} is easily determined. However, it does simplify the problem, and the assumption is usually valid at low frequencies when pentodes are used.

The final open loop equivalent of the amplifier is shown in figure (10.18d) where it has been arranged in the form of the general amplifier treated in chapter 4. In this case, the components of the unloaded pi section are

$$Z_1 = \frac{r_p}{s(r_pC_p) + 1}; \qquad Z_2 = \frac{1}{sC_g}; \qquad Z_3 = sL + R_L$$

The system amplification is

$$\beta A_o = \frac{E_o}{E_i} = \frac{E_g}{E_g} = 1 = -g_m Z_m \tag{10.29}$$

However, it was shown in chapter 4 that

$$Z_m = \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3}$$

so that the system amplification of the equivalent circuit is

$$\beta A_o = 1 = -g_m \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3} \tag{10.30}$$

If the expressions for the three impedances are substituted into this relationship, the result can be expressed in the following form after some standard algebraic manipulation:

$$\beta A_o = 1$$

$$= -\frac{\dfrac{\mu}{r_pC_pLC_g}}{s^3 + \left(\dfrac{R_m}{L} + \dfrac{1}{r_pC_p}\right)s^2 + \left(\dfrac{1}{LC_p} + \dfrac{1}{LC_g} + \dfrac{R_L}{LC_pr_p}\right)s + \dfrac{1}{r_pC_pLC_g}} \tag{10.31}$$

Subtract 1 from both sides of the equation to obtain the characteristic equation $\beta A_o - 1 = 0$; cross-multiply, collect terms, and express the result as:

$$s^3 + (R_L/L + 1/r_pC_p)s^2 + (1/LC_p + 1/LC_g + R_L/r_pC_pL)s + (\mu + 1)/r_pC_pLC_g = 0 \tag{10.32}$$

This equation has the general cubic form

$$s^3 + a_2 s^2 + a_1 s + a_0 = 0$$

Because a cubic equation is involved, the roots can be computed only with some difficulty. Fortunately, the Routh-Hurwitz criterion (see chapter 7) resolves this problem.

Form the Routh-Hurwitz determinant as follows:

$$\begin{array}{c|l|l|}
s^3 & 1 & a_1 \\
s^2 & a_2 & a_0 \\
s^1 & (a_2 a_1 - a_0)/a_2 & 0 \\
s^0 & a_0 & 0
\end{array}$$

During the buildup there must be a pair of right half plane poles, so there must be two sign changes in the first column of the determinant. This requires that

$$a_2 a_1 < a_0 \tag{10.33}$$

during buildup.

In the steady state the third row of the determinant must vanish, so that

$$a_2 a_1 = a_0 \tag{10.34}$$

This is the *condition for oscillation* expressed in terms of the coefficients of the cubic characteristic equation. The frequency of oscillation is readily found from the coefficients of the last nonvanishing row of the determinant. That is, the roots of $a_2 s^2 + a_0 = 0$ or $s^2 + a_0/a_2 = 0$ give the oscillator frequency in the steady state. These roots are

$$s_{1,2} = \pm j\sqrt{\frac{a_0}{a_2}} = \pm j\omega_0$$

and the *oscillator frequency* is

$$\omega_0 = \sqrt{\frac{a_0}{a_2}} \tag{10.35}$$

Substitute the values for the coefficients into equations (10.34) and (10.35) so that the oscillator frequency in the steady state is

$$\omega_0 = \sqrt{\frac{\mu + 1}{LC_g + R_L C_g r_p C_p}} \tag{10.36}$$

and the condition for oscillation is

$$\frac{C_g}{C_p}\left(1 + \frac{C_p}{C_g} + \frac{R_L}{r_p}\right) = \frac{\mu + 1}{R_L r_p C_p/L + 1} \tag{10.37}$$

In certain special cases it is valid to assume that

$$\mu \gg 1; \qquad r_p \gg R_L; \qquad \frac{R_L r_p C_p}{L} \ll 1$$

When these conditions exist, the preceding equations assume the following approximate forms:

$$\omega_0 \doteq \sqrt{\frac{\mu}{LC_g}} = \sqrt{\frac{1}{LC_p}} \tag{10.38}$$

Condition for oscillation is

$$\frac{C_g}{C_p} \doteq \mu \tag{10.39}$$

The analytical process only yields two equations. Hence all the oscillator variables except two must be known before this analysis is especially useful.

The technique and procedure just illustrated are perfectly general and can be followed for any given circuit. The required degree of familiarity can be achieved only by experience.

10.14. Wien Bridge Oscillator

One of the most popular circuits using *RC* circuits for frequency control is the Wien bridge oscillator shown in figure (10.19a). The method of analyzing this circuit and deriving the equations for the frequency and conditions for oscillation is exactly the same as that followed in the preceding section. The analysis given here is a further illustration of the technique.

The open loop class A equivalent circuit of the oscillator is shown in figure (10.19b). All coupling capacitors C_c were assumed virtual short circuits at the signal frequencies of interest. The first amplifier stage is assumed to have an amplitude response that is *flat* over the entire frequency range of the oscillator so that

$$A_1 = A_{r1} = -g_{m_1} R \tag{10.40}$$

where

$$R = \frac{1}{1/r_{p1} + 1/R_{L1} + 1/R_g} \tag{10.41}$$

The calculation of the voltage gain of the second stage is somewhat involved because the connected load circuit is quite complicated.

Because cathode degeneration is involved, the gain of the amplifier in general terms is

$$A_2 = -g'_{m_2} Z_m$$

where
$$g'_{m_2} = g_{m_2} \frac{1}{1 + R_k(\mu + 1)/(r_p + Z_L)}$$

The tube is a pentode, so it is nearly always valid to assume that r_p is

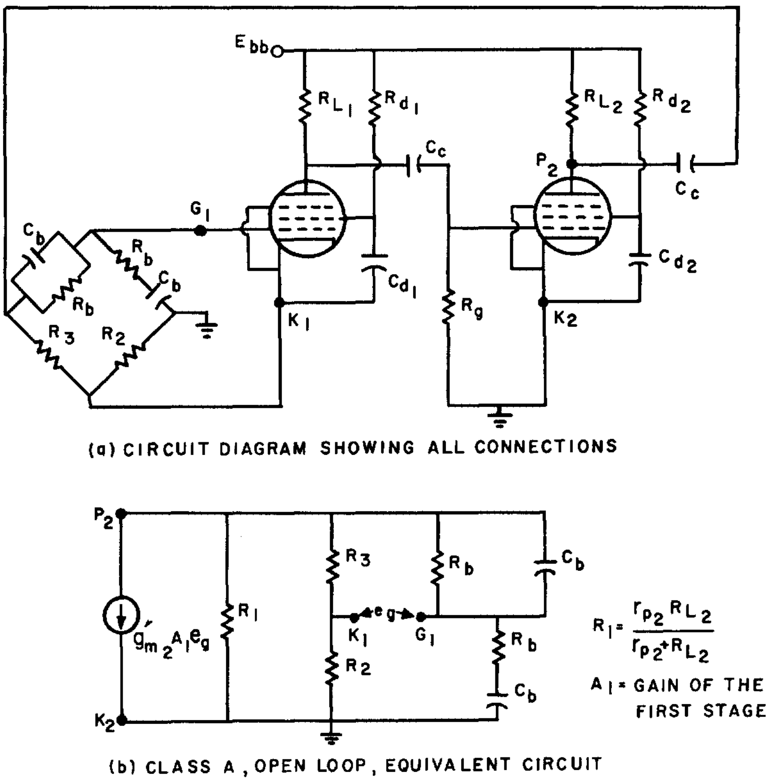


Fig. 10.19. Circuit diagrams for the Wien bridge oscillator.

much larger than Z_L. Thus the effective transconductance is approximately

$$g'_{m_2} \doteq \frac{g_{m_2}}{1 + R_k(\mu + 1)/r_p} \tag{10.42}$$

The values of Z_m and the over-all gain are evaluated only with considerable labor. However, the complete result can eventually be written

$$A_2 = -g_{m_2} R' \frac{s^2 + a_1 s + a_0}{s^2 + b_1 s + b_0} \tag{10.43}$$

where

$$R' = \frac{R_b R_x}{R_b + R_x} \cdot \frac{R_3}{R_2 + R_x} \tag{10.44}$$

$$R_x = \frac{R_{L_2}(R_2 + R_3)}{R_{L_2} + R_2 + R_3} \tag{10.45}$$

$$\omega_b = \frac{1}{R_b C_b} \tag{10.46}$$

$$a_1 = \omega_b(2 - R_2/R_3) \tag{10.47}$$

$$a_0 = \omega_b^2 \tag{10.48}$$

$$b_1 = \omega_b \frac{3R_b + R_x}{R_b + R_x} \tag{10.49}$$

$$b_0 = \omega_b^2 \frac{R_b}{R_b + R_x} \tag{10.50}$$

The product of the two stage gains is equal to the feedback factor. That is, $\beta A_0 = A_1 A_2$. Then, for oscillations to occur, $\beta A_0 - 1 = 0$. The values for A_1 and A_2 have been computed and can be substituted into this relationship. Once the common denominator is obtained and then cancelled into the zero, the result can finally be written

$$s^2 + c_1 s + c_0 = 0 \tag{10.51}$$

where

$$c_1 = \omega_b \left[\frac{A_{r_1} A_{r_2}}{A_{r_1} A_{r_2} - 1} \left(2 - \frac{R_2}{R_3}\right) - \frac{1}{A_{r_1} A_{r_2} - 1} \cdot \frac{3R_b + R_x}{R_b + R_x} \right]$$

$$c_0 = \omega_b^2 \left(\frac{A_{r_2} A_{r_1}}{A_{r_1} A_{r_2} - 1} - \frac{1}{A_{r_1} A_{r_2} - 1} \cdot \frac{1}{R_b + R_x} \right)$$

$$A_{r_1} = g_{m_1} R; \qquad A_{r_2} = g_{m_2} R'$$

It is generally true that the product of A_{r_1} and A_{r_2} is much larger than 1. Therefore the preceding coefficients simplify to

$$c_1 = \omega_b \left(2 - \frac{R_2}{R_3} - \frac{1}{A_{r_1} A_{r_2}} \cdot \frac{3R_b + R_x}{R_b + R_x} \right) \tag{10.52}$$

$$c_0 = \omega_b^2 \tag{10.53}$$

The roots of equation (10.51), which are the poles of the closed loop transfer function, are $s_{1,2} = -c_1/2 \pm j\sqrt{c_o - (c_1/2)^2}$. Evidently a sustained oscillation will be produced only if c_1 is zero. If it is zero, the frequency of oscillation is

$$\omega_0 = \omega_b = \frac{1}{R_b C_b} \tag{10.54}$$

For c_1 to be zero,

$$\frac{R_2}{R_3} + \frac{1}{A_{r_1} A_{r_2}} \cdot \frac{3R_b + R_x}{R_b + R_x} = 2 \tag{10.55}$$

This is the condition for sustained oscillation. If the circuit is designed so that

$$\frac{1}{A_{r_1} A_{r_2}} \cdot \frac{3R_b + R_x}{R_b + R_x} \ll 1$$

and this will generally be the case, the condition for oscillation becomes

$$\frac{R_2}{R_3} = 2 \tag{10.56}$$

From this last relationship it is evident that either R_2 or R_3 must be variable. This is necessary so that the ratio can exceed 2 during buildup, but will maintain the equality under all conditions thereafter. Clearly, should the amplitude of the oscillation momentarily increase because of a change somewhere in the circuit, the amplitude will be stabilized if the equality in (10.56) is violated and R_2/R_3 becomes less than 2. This would require either that (1) R_2 decrease, or (2) R_3 increase by the right amount to make the amplitude constant. Thus either R_2 or R_3 could be variable, but their temperature coefficients of resistance would necessarily be opposite.

The Wien bridge oscillator is popular with designers for wide range tunable oscillators. One of the most important reasons for this is because the oscillator frequency depends upon the reciprocal of the tuning capacitance, rather than upon the square root of the reciprocal. Thus a given change in capacitance will produce a wider range of oscillator frequencies. Also, changes in frequency band are easily made by changing the tuning resistance, and the tuning capacitance is used for continuous frequency variation within one band. This makes a 10 to 1 change in frequency readily available within one band. The oscillator output is quite free from harmonic distortion because of the selective characteristic of the Wien bridge. This is another important practical advantage of the circuit.

10.15. Transistor Crystal Oscillator

The technique used in analyzing vacuum tube feedback oscillators also applies to transistor feedback oscillators without modification in principle. A simple example should serve to illustrate the point.

Consider the transistor crystal oscillator shown in figure (10.17b). From this diagram construct the class A open loop equivalent circuit. This is shown in figure (10.20). The crystal has been replaced by a series resonant circuit. You can see from the equivalent circuit that the crystal *must* operate in the series mode; in the parallel mode there would be practically no output voltage and the feedback would be insufficient to sustain oscillation.

The voltage gain of this amplifier is the βA_o of the oscillator, just as

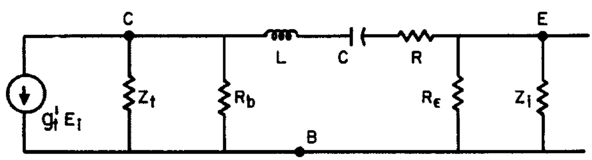


Fig. 10.20. Open loop equivalent circuit of the transistor crystal oscillator of figure (10.17b).

it was for vacuum tube oscillators. For a constant amplitude oscillation to be produced, βA_o must be equal to 1.

The voltage gain of the amplifier is given by the general gain equation

$$\beta A_0 = -g_t' Z_m = -g_t' \frac{Z_1 Z_2^2}{Z_1 + Z_2 + Z_3} \tag{10.57}$$

Terms Z_1, Z_2, and Z_3 are the usual three impedances of the general unloaded pi network associated with amplifiers. For a grounded base amplifier it was shown in chapter 8 that

$$g_t' = -\frac{g_t}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{10.58}$$

$$Z_i = r_\varepsilon \left(1 + \frac{r_b}{r_\varepsilon} \cdot \frac{z_t + Z_L}{z_c + Z_L}\right) \tag{10.59}$$

where z_c and z_t have the usual meaning associated with them as derived in chapter 8. For the purpose of this discussion we shall also assume that the oscillator frequency is below alpha cutoff by a sufficient margin to assume that α is a constant. Here Z_L is the input impedance of the passive load circuit.

The voltage gain of the open loop circuit can now be written

$$\beta A_o = + \frac{g_t Z_m}{1 + \dfrac{r_\varepsilon}{r_b}\left(\dfrac{z_c + Z_L}{z_t + Z_L}\right)} \tag{10.60}$$

We immediately run into trouble at this point because Z_L is partially governed by Z_i, and Z_i is partially governed by Z_L. As long as this situation prevails, the circuit is virtually insoluble. Fortunately, there is a way out of this dilemma. Because the over-all gain of the amplifier is only unity, the input impedance Z_L of the load circuit is nearly always much less than either z_c or z_t. As a result, the equations for the gain and input impedance of the amplifier are closely approximated by

$$\beta A_o = + \frac{g_t Z_m}{1 + r_\varepsilon z_c / r_b z_t} = + \frac{g_t Z_m}{1 + r_\varepsilon / [r_b(1 - \alpha)]} \tag{10.61}$$

$$Z_i = r_\varepsilon \left(1 + \frac{r_b z_t}{r_\varepsilon z_c}\right) = r_\varepsilon + r_b(1 - \alpha) = R_i \tag{10.62}$$

The transistor transconductance is

$$g_t = \frac{z_m}{r_b z_t} = \frac{\alpha}{r_b(1 - \alpha)} \tag{10.63}$$

so that the voltage gain is

$$\beta A_o = \frac{\alpha Z_m}{r_\varepsilon + r_b(1 - \alpha)} = \frac{\alpha Z_m}{R_i} \tag{10.64}$$

The three impedances of the unloaded pi section are computed from the equivalent circuit given in figure (10.20). The results may be expressed in the following form:

$$Z_1 = \frac{R_b z_t}{R_b + z_t} = r_t \frac{\omega_c}{s + \omega_b} \tag{10.65}$$

$$Z_2 = \frac{R_\varepsilon R_i}{R_\varepsilon + R_i} = R_2 \tag{10.66}$$

$$Z_3 = \frac{L}{s}(s^2 + Bs + \omega_x^2) \tag{10.67}$$

where
$$r_t = r_c(1 - \alpha) \tag{10.68}$$

$$\omega_b = \omega_c(1 + r_t/R_b) \tag{10.69}$$

$$\omega_x^2 = (\text{crystal resonant frequency})^2 = \frac{1}{LC} \tag{10.70}$$

$$B = \frac{R}{L} \tag{10.71}$$

The Q of crystals is so high that the losses can be virtually ignored, so that

$$Z_3 = \frac{L}{s}(s^2 + \omega_x^2) \tag{10.72}$$

Substitute the three equations for the impedances into the gain equation. Set βA_o equal to 1; multiply numerator and denominator through by $s(s + \omega_b)/L$. The result is

$$\beta A_o = 1 = \frac{\alpha R_2 \omega_c r_t s / R_i L}{(s + \omega_b)(s^2 + \omega_x^2) + s(s + \omega_b)R_2/L + s\omega_c r_t/L} \tag{10.73}$$

Expand the denominator and then collect coefficients of like powers of s. Cross multiply and set up the equation $\beta A_o - 1 = 0$. This leads to

$$s^3 + a_2 s^2 + a_1 s + a_0 = 0 \tag{10.74}$$

where
$$a_2 = \omega_b + \frac{R_2}{L} \tag{10.75}$$

$$a_1 = \omega_x^2 + \frac{1}{LC_c} \cdot \frac{R_2}{r_c}\left(1 + \frac{r_t}{R_b}\right) + \frac{1 - \alpha}{LC_c}\left(1 - \alpha\frac{R_2}{R_i}\right) \tag{10.76}$$

$$a_0 = \omega_b \omega_x^2 \tag{10.77}$$

Because a cubic equation is involved, you can show from the Routh-Hurwitz determinant that the oscillator frequency in the steady state will be

$$\omega_{osc}^2 = \frac{a_0}{a_2} \tag{10.78}$$

and the condition for sustained oscillation will be

$$a_1 = \frac{a_0}{a_2} \tag{10.79}$$

Substitute the equations for the coefficients into these two expressions. After some algebraic manipulation we can show that

$$\omega_{osc}^2 = \omega_x^2 \frac{\omega_b L}{R_2 + \omega_b L} \tag{10.80}$$

where $L =$ effective inductance of the crystal; $\omega_x =$ series resonant frequency of the crystal; ω_b is given by equation (10.69); R_2 is given by equation (10.66). The condition for oscillation is

$$\frac{r_c}{R_b} + \frac{r_c}{R_2} + \frac{r_m}{R_i} = \frac{1}{\alpha - 1}\left\{1 + \frac{\omega_{osc}}{\omega_c[1 - r_c(\alpha - 1)/R_b]}\right\} \tag{10.81}$$

The condition for oscillation just given sheds some light on the circuit. Thus, if α is about 2 and r_c is much larger than R_b, the condition for oscillation can be met only if

$$\omega_{osc} < \omega_c\left[1 - \frac{r_c}{R_b}(\alpha - 1)\right] \tag{10.82}$$

If α is less than 1 and r_c is much larger than R_b, oscillations will not result under any circumstances. Therefore the circuit can be used only with transistors having a current amplification factor greater than 1.

10.16. Negative Grid UHF Oscillators

Oscillators designed for use as ultrahigh frequencies (UHF) are of two distinctly different types: (1) transit time oscillators and (2) negative grid oscillators. Both of these circuit types are covered in detail in other books treating just the subject of UHF techniques. However, despite the fact that UHF negative grid oscillators differ remarkably in appearance from lower frequency oscillators, it is relatively simple to show that the same physical processes are involved in both cases.

For an oscillator to operate in conventional manner at ultrahigh frequencies, two main precautions must be observed:

(1) The transit time of the electrons through the interelectrode space must be small compared to the period of the oscillation.

(2) Because the oscillator frequency will be given approximately by $\sqrt{1/LC}$, where the L and C are the constants in the resonant tank circuit, it is clear that very small values of L and C are required for UHF operation. To avoid tube selectivity effects, the total L and C should be affected only slightly by the lead inductances and interelectrode capacitances of the tube.

The physical appearance of the UHF negative grid oscillator tube is largely governed by the necessity of reducing transit time, lead inductance, and interelectrode capacitance without reducing the power handling capacity of the tube and load circuit to the vanishing point.

Because the values required for L and C are small, ordinary coils and capacitors would be so small that their power handling capacity would be negligible. Hence resonant transmission lines are used in place of conventional tuned circuits. The variation of input impedance of a lossless line as a function of the electrical length is shown in figure (10.21), where it is assumed that the line is short circuited. Evidently

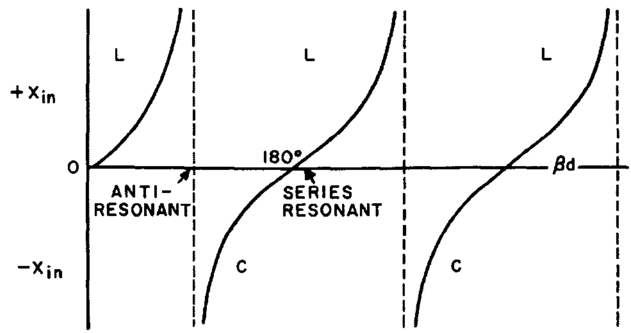


Fig. 10.21. Reactance characteristics of a short-circuited lossless transmission line.

Z_{in} can assume the nature of an inductance, a capacitance, or a series or parallel resonant circuit, depending upon the length βd of the line. Thus short-circuited transmission lines can be used in place of ordinary tuned circuits at ultrahigh frequencies. The greater physical size of the line permits a greater power handling capacity.

A typical UHF negative grid oscillator is shown in figure (10.22a). Because resonant transmission lines are used in both the plate and grid circuits, this is clearly a tuned plate, tuned grid (TPTG) oscillator. Thus the equivalent circuit can be drawn as shown in figure (10.22b) with the lines replaced by parallel tuned circuits. The interelectrode capacitances C_{gk} and C_{pk} have been combined with the capacitances of the equivalent tuned circuit, while the grid-to-plate capacitance is shown. For a resonant circuit to be formed, the parallel tuned circuits must appear as effective inductances, which means that the transmission lines must be somewhat less than a quarter wavelength ($\beta d = 90°$) long. The circuit is clearly that of a Hartley oscillator as shown in figure (10.22c).

The same procedure can be followed for any of the UHF negative grid oscillators.

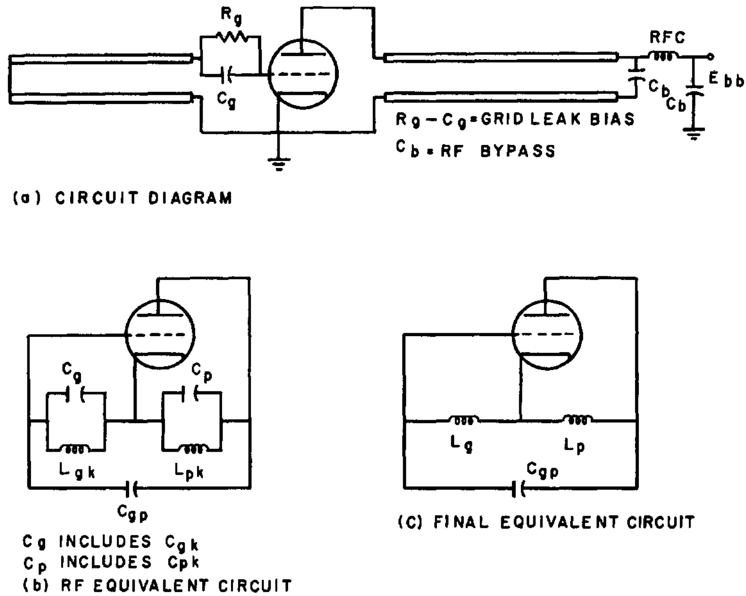


Fig. 10.22. Tuned plate, tuned grid UHF oscillator.

10.17. Parasitic Oscillations

A *parasitic* oscillation is an undesired or spurious oscillation in an electronic circuit. They often occur in amplifiers and oscillators as a result of a wide variety of causes. Parasitics are undesirable and should be eliminated because they can lead to reduced component life arising from overloading, loss in efficiency, reduced stability, and so on. Also, under some conditions the parasitic may radiate and thereby become objectionable as well as illegal. Fortunately, through the use of good design judgment, with emphasis on careful physical layout, the possibility of parasitic oscillation can be reduced; the possibility cannot be entirely removed, and when parasitics exist after the equipment has been constructed, their elimination can be one of the most trying experiences the circuit designer can undergo.

The most common kind of parasitic oscillation is of the tuned plate, tuned grid variety. For example, consider the case of a tuned vacuum tube voltage amplifier such as that shown in figure (10.23a). Only essential a-c components are shown. A simple rearrangement of the circuit makes it appear as shown in figure (10.23b) and this is the same as the circuit diagram of a tuned plate, tuned grid oscillator. Under certain circuit conditions, oscillations result. This can be a serious problem in intermediate frequency and radio frequency amplifiers.

Essentially the same problem occurs in transformer coupled push-pull amplifiers, a common circuit when tubes are operated in class AB_2 or B_2 (see chapter 12). A typical circuit diagram with d-c connections omitted is given in figure (10.23c). In this case the transformer induc-

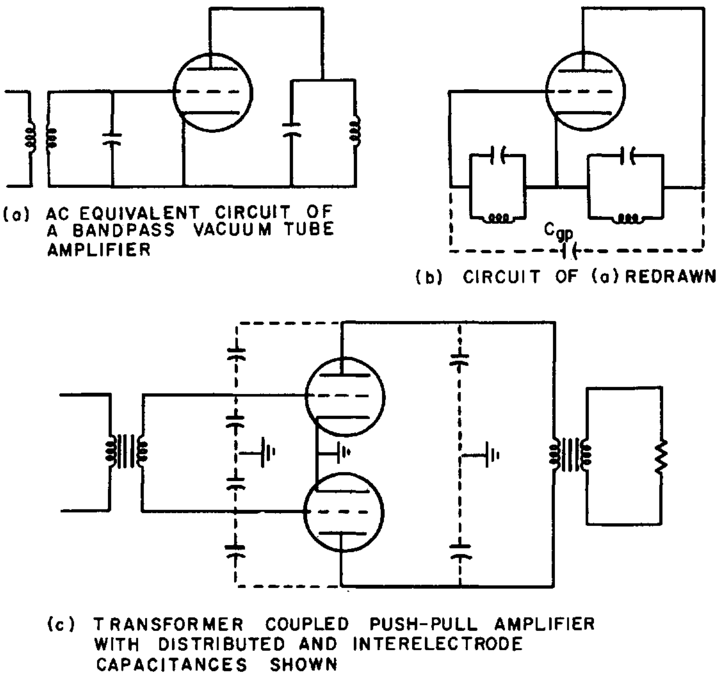


Fig. 10.23. Conditions leading to tuned plate, tuned grid oscillation in amplifiers.

tances form parallel resonant circuits with wiring and tube capacitances. Each of the transformer center taps is grounded for alternating current, so that the class A equivalent circuit for either tube is the same as that of a tuned grid, tuned plate oscillator.

This type of parasitic can usually be eliminated by damping one of the tuned circuits. The damping is preferably applied in the grid circuit and usually takes the form of small resistances in series with the grid leads.

From the preceding discussion you can see that it is generally undesirable to use choke shunt feed in both the input and output circuits of amplifiers and oscillators. The chokes can form resonant circuits with various capacitances, and the conditions necessary for TPTG oscillation at some haphazard frequency could easily result. If choke shunt feed *must* be used in both circuits, the possibility of parasitic oscillation can be reduced by using chokes with values of inductance different by a factor of about 100.

Resonant circuits can appear in other, possibly unexpected, ways. Neutralizing and bypass capacitors can become series resonant at some frequency. Above this frequency they are inductive and may form parallel resonant circuits with various stray capacitors. Also, if the leads between tubes are too long, they may form resonant transmission lines, which will result in UHF oscillations of the TPTG variety.

The conditions necessary for oscillation may be produced by positive feedback in many ways. The *waveguide effect* in high gain, bandpass amplifiers is one such possibility. The amplifier chassis operates as a waveguide cutoff attenuator, and the attenuation may not be greater than the amplifier gain. Feedback through the power supply impedance is a common cause of oscillation in high gain multistage amplifiers.

Feedback can also be introduced in high frequency circuits if multiple signal frequency ground connections are used. A *single* signal frequency ground should be used if possible.

Electromagnetic coupling and feedback can result if components are poorly positioned with respect to one another or are crowded together. Tuning capacitors should generally be mounted with the rotors grounded. Contacts in nonsoldered joints, switch contacts, and coil taps should be clean. These are only some of the many causes of parasitics.

PROBLEMS

10.1. The losses in the tank circuit and the power delivered to the load by a Colpitts oscillator can be represented by a 6000 ohm resistor in parallel with the tank coil. It is known that $L = 20\ \mu$h, $C_g = C_p = 500\ \mu\mu$f, and $r_p = 30{,}000$ ohms. Compute the frequency of operation of the oscillator and the value of tube transconductance required to sustain oscillations. Note that this is *not* the same as the circuit covered in the text. Use any justifiable approximations.

10.2. A Colpitts oscillator is to be designed using a 6C5 triode vacuum tube for which $C_{gk} = 3.0\ \mu\mu$f; $r_p = 10{,}000$ ohms; $C_{pk} = 11.0\ \mu\mu$f; $g_m = 2000\ \mu$mhos; $C_{gp} = 2.0\ \mu\mu$f; $\mu = 20$. The oscillator is to operate at 10 mcps when fully loaded. The most convenient available tank coil has an inductance of 10 μh and a Q of 100 at 10 mcps. From stability considerations, the loaded Q of the tank circuit is limited to a minimum value of 20; that is, when the circuit Q is 20, the oscillation should be barely self-sustaining. Assume that all energy losses can be represented by a resistor in series with the coil. Use any justifiable approximations.

(a) Compute the necessary values of grid and plate capacitance that must be added.

(b) What is the maximum value of resistance that can be coupled into the tank circuit?

10.3. A three-stage resistance coupled amplifier with all stages identical receives power from a common power supply of R_0 ohms internal impedance. Assume operation in the mid-frequency range of all three amplifiers. If $R_L = 25{,}000$ ohms, $r_p = 500{,}000$ ohms, $R_g = 500{,}000$ ohms, $g_m = 1800$ μmhos, compute the largest possible value for R_0 that will *not* cause the amplifier to oscillate.

10.4. In the RC oscillator of figure (10.16c) assume that $R_1 = R_2 =$ 10,000 ohms; $C_1 = C_2$. In the first amplifier stage, $R_L = 30{,}000$ ohms and $g_m = 1000$ μmhos. In the second stage amplifier, R_L is unknown but $g_m = 1000$ μmhos. Neglect the effects of r_p and R_g. Assuming mid-band operation of the amplifier stages,

(a) Determine the oscillator frequency as a function of C_1.

(b) Compute the smallest possible value for R_L of the second stage for which oscillation is possible.

10.5. Three identical resistance coupled amplifiers, using pentodes, are connected in cascade and supplied from a power supply of zero internal impedance. The output from the last stage is connected directly into the input of the first stage. Compute the oscillator frequency and the value of tube g_m required to sustain the oscillation. Neglect the effects of interelectrode capacitance.

10.6. A crystal controlled transistor oscillator is shown in figure (10.17b). Does the crystal operate in the parallel or series resonant mode? Explain your answer.

10.7. Derive equations for the condition for oscillation and the frequency of oscillation of the transistor amplifier of figure (10.17b).

Chapter 11

NONLINEAR CLASS A CIRCUITS

It is an extremely fortunate fact that many electronic circuits are linear. Because of this, the equivalent circuits developed in chapter 1 become powerful tools in the design and analysis of electronic circuits.

All the circuits used in electronics are not linear. A large number of circuits are fundamentally dependent upon the *nonlinear* characteristics of electronic devices for their operation.

Despite the fact that nonlinearity in the characteristics of electronic devices may have important applications, it also causes considerable difficulty. It makes the analytical treatment developed in the preceding chapters inapplicable directly. Also, many class A circuits are desirably linear, but because of other requirements are forced to operate over a nonlinear part of the characteristics. In such cases the electronic device causes *nonlinear distortion*; this is a decided disadvantage.

In summary, the nonlinear properties of electronic devices may be of practical use or the source of great difficulty. In any case, the direct analytical treatment of the preceding chapters must be somewhat revised. This is the primary purpose of this chapter. A secondary purpose is to illustrate the application of the ideas to practical problems of interest. Although the discussion is primarily concerned with vacuum tubes, the techniques are general and can be applied to any nonlinear device.

11.1. Dynamic Transfer Characteristics, Triodes and Pentodes

The behavior of vacuum tubes is generally represented by the static plate characteristic curves; they are useful in making numerical calculations, but are not especially adapted to the analysis and description of circuit behavior. Instead, it is found that the *dynamic transfer characteristic* is more easily used. The construction of this curve was discussed in section (3.3) of chapter 3.

The circuit diagram of a class A vacuum triode circuit is shown in figure (11.1). Following the method outlined in chapter 3, the load line is plotted on the static characteristics of the tube as shown in figure (11.2).

The dynamic transfer characteristic is the graph of plate current i_b as a function of the grid voltage e_c for specified values of load resistance R_L and plate power supply voltage E_{bb}. It is computed from figure (11.2) by arbitrarily assuming a series of operating points, a, b, c, d, e, Q, f, g, and so on, along the load line. Each such point corresponds to a particular value of plate current and grid voltage, and these data are plotted to obtain figure (11.3). This is the dynamic transfer characteristic of the circuit.

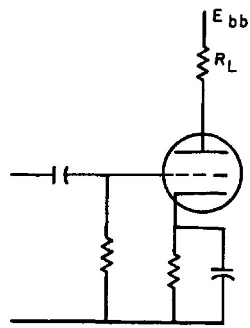


Fig. 11.1. Class A triode amplifier.

Figure (11.3) illustrates an important point: the transfer characteristic of a triode with resistance load is nonlinear. A definite curvature can be seen in the figure. This means that if a given grid signal is impressed on the tube, the resulting plate current will not be an exact replica of the input. This effect is shown in figure (11.4) for the case of a

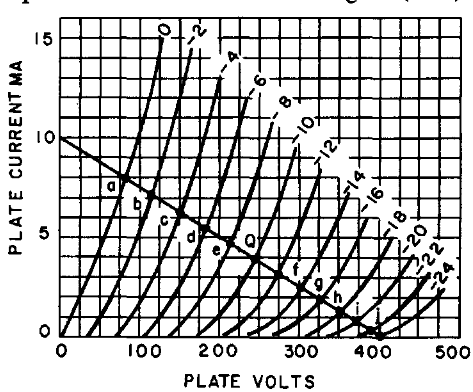


Fig. 11.2. Load line on triode characteristics.

sinusoidal grid signal. The resulting plate current is definitely *not* sinusoidal because of the obvious flattening of the bottom of the waveform. The output is different from the input so the output is distorted. The distortion is caused by the nonlinearity of the tube, so it is called *nonlinear distortion.*

Consideration of figures (11.3) and (11.4) shows that the effects of nonlinearity in the tube characteristics will be important in two cases:

(1) When the signal amplitude is so large that it causes the operating point to swing over a large part of the complete transfer characteristic.

(2) When the transfer characteristic is extremely nonlinear so that even small signal excursions about the Q point cause the operating point to traverse a nonlinear part of the curve.

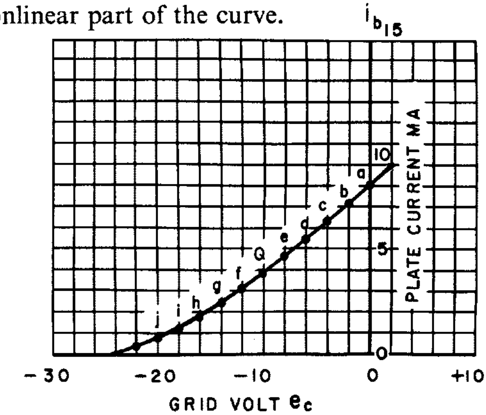


Fig. 11.3. Dynamic transfer characteristic constructed from figure (11.2).

In the linear class A circuits discussed in previous chapters it was assumed that neither of these two conditions existed. The devices used

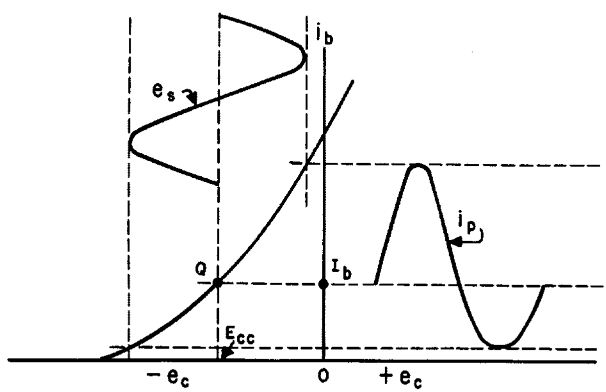


Fig. 11.4. Effect of nonlinearity on the signal transmission through a triode.

were selected from those exhibiting only a small over-all curvature, and the signal amplitudes were kept small. As a result, the part of the transfer characteristic traversed by the operating point could be closely approximated by a straight line segment and constant values of g_m and r_p could be assumed.

It is found that the dynamic transfer characteristic of a triode can be closely approximated by a parabolic curve. The mathematical expression for a parabolic function contains second degree, but no higher degree, terms. It will be shown later that this type of nonlinearity causes a purely sinusoidal grid voltage to produce a plate current signal having two components, one at the original signal frequency and called the fundamental, and one at twice the signal frequency, called the *second harmonic.* The distortion caused by a parabolic characteristic is called *second harmonic distortion.*

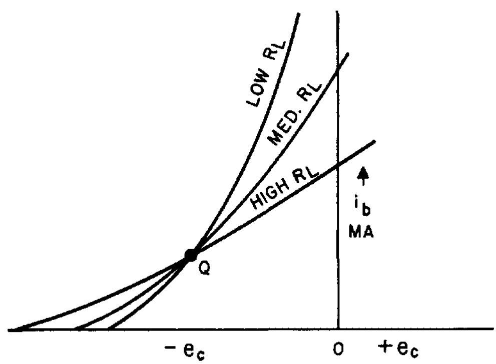


Fig. 11.5. Effect of variations in load resistance on the transfer characteristic of a triode.

The effect of variations in load resistance on the second harmonic distortion is important. The dynamic transfer characteristics for a triode with three different values of load resistance and the same Q point are shown in figure (11.5). You can see that increasing the load resistance has the effect of reducing the curvature in the transfer characteristic. It is reasonable to conclude that this will also reduce the second harmonic distortion. This proves to be true.

The dynamic transfer characteristic of a pentode is obtained by the same method as that outlined for triodes. The resulting curve, which is shown in figure (11.6), differs from the triode characteristic. In this case it is necessary to use at least a cubic function to describe analytically the transfer function. The application of a pure, single frequency, sinusoid to the input causes the plate current to have at least three components: (1) fundamental = original frequency term; (2) second harmonic = twice the original frequency; (3) third harmonic = three times the fundamental frequency. Hence, in a pentode, both second and third harmonic terms must be considered.

You can see from figure (11.6) that increasing the load resistance causes a general increase in the curvature of the characteristic at the

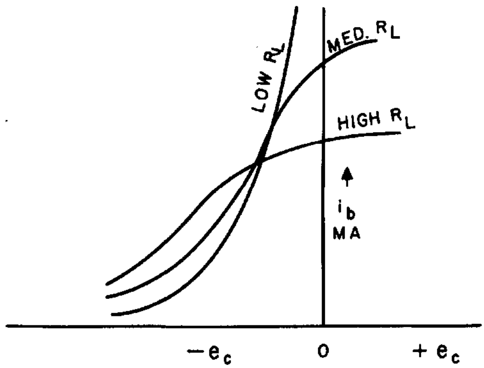


Fig. 11.6. Effect of variations in load resistance on the dynamic transfer characteristic of a pentode.

high current end. This results in increased third harmonic distortion. It is difficult to anticipate the variation in second harmonic distortion caused by load resistance variations, but it will be investigated in detail later.

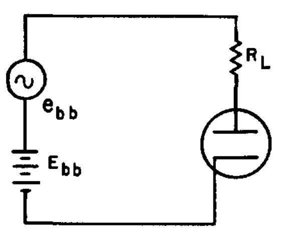


Fig. 11.7. Class A diode circuit.

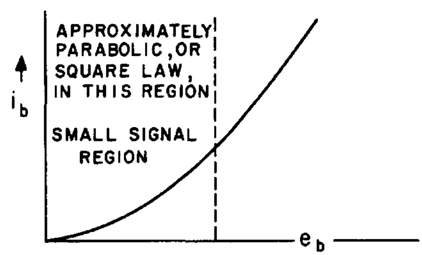


Fig. 11.8. Diode characteristic showing the usual region for class A operation.

11.2. Nonlinearity in Diodes

Diodes, either vacuum or semiconductor, exhibit nonlinearities, but the dynamic transfer characteristics must be evaluated by a slightly different method than that followed in section (11.1).

A class A diode circuit with resistive load appears in figure (11.7). The order of connection of generators, diode, and load is irrelevant to the discussion. The signal voltage e_{bb} is a variation in the supply voltage. When it is desirable for the diode to be nonlinear, both components of the plate supply voltage are adjusted to restrict tube operation to the

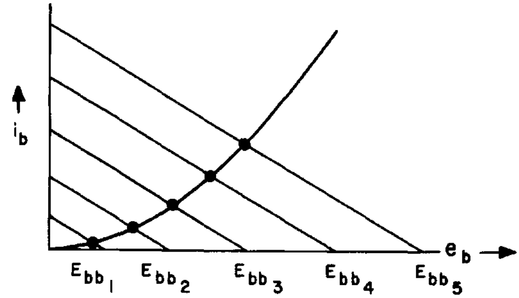


Fig. 11.9. Effect of signal voltage variations on the load line and operating point of a diode.

low current region where the nonlinearity is most pronounced. Figure (11.8) illustrates a *small signal* operating condition.

Because the diode has only a single characteristic curve and the load

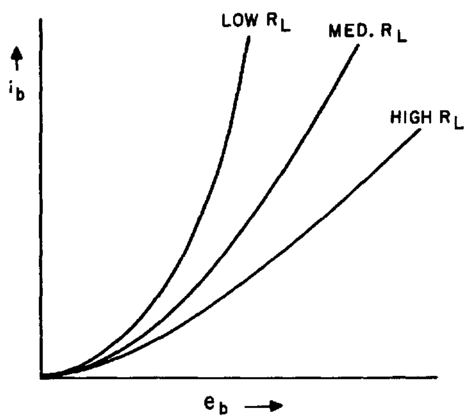


Fig. 11.10. Dynamic transfer characteristics of a diode with resistance load.

resistance is fixed, variations in the signal voltage cause the load line to be translated parallel to itself and laterally as shown in figure (11.9). In this figure the parabolic, or *square law*, part of the curve has been expanded so that only the region of interest in small signal operation is shown. The operating point moves along the tube characteristic as the signal voltage causes the effective power supply voltage to vary about the quiescent value of E_{bb}. Thus the dynamic transfer characteristic is simply obtained by plotting the total plate supply voltage and the corresponding plate current as shown in figure (11.10) for three values of load resistance.

The results obtained for the diode are similar to those found for the triode because (1) the curve is essentially parabolic, following an approximate square law, and (2) the amount of curvature is decreased by increasing the load resistance. Hence increasing the load resistance will decrease the amount of second harmonic distortion generated by the diode.

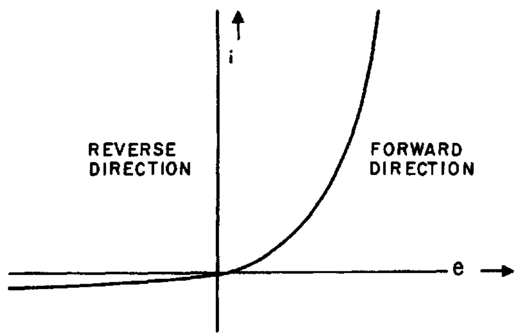


Fig. 11.11. Varistor characteristic.

A semiconductor diode, or varistor, presents fundamentally the same problem and results, though the static characteristic differs in some respects from that of a diode. A varistor characteristic is shown in figure (11.11).

11.3. Power Series Treatment of Nonlinearity

We have seen that the transfer characteristics of electronic components are always nonlinear. It is important to compute the effects produced by this nonlinearity.

Graphical methods can be used to solve any problem, either linear or nonlinear; unfortunately, such solutions do not lead to the formulation of specific laws or rules of behavior. Instead, each graphical solution is simply a single solution for the particular problem at hand. This is not desirable because it does not permit proper correlation of the behavior of various nonlinear circuits. Hence, while analytical methods cannot be used exclusively in nonlinear circuit solutions, and because graphical methods do not possess the desired degree of generality, a combination of the two methods will be used.

It can be shown that nearly any smooth, physically determined curve, can be represented by a power series.[1] For example, $f(x)$, a function of x, can be expanded into a power series about some point b as follows:

$$f(x) = a_0 + a_1(x - b) + a_2(x - b)^2 + a_3(x - b^3 + \ldots \quad (11.1)$$

where

$$a_0 = f(b)$$
$$= \text{value of the function when } x = b$$
$$a_1 = \frac{df(b)}{dx}$$
$$= \text{value of first derivative when } x = b$$
$$a_2 = \frac{d^2f(b)}{dx^2}\frac{1}{2}$$
$$= \text{value of second derivative when } x = b$$

and so on.

For the transfer characteristic of a vacuum tube, the signal component of the plate current should be expressed in terms of a power series involving the input signal voltage. Thus, by analogy to equation (11.1),

$$i_p = a_0 + a_1e_s + a_2e_s^2 + a_3e_s^3 + \ldots \quad (11.2)$$

From the definition of the terms i_p and e_s as signal components you can see that when the input signal is zero, the resulting output signal component must also be zero. Hence $a_0 = 0$ and the power series for i_p becomes

$$i_p = a_1e_s + a_2e_s^2 + a_3e_s^3 + \ldots \quad (11.3)$$

Now specialize the problem to the case in which the signal input is a pure, single frequency sinusoid. Let $e_s = E_s \cos \omega t$. The signal component of plate current is then

$$i_p = a_1E_s \cos \omega t + a_2(E_s \cos \omega t)^2 + a_3(E_s \cos \omega t)^3 + \ldots$$

The equation can be put into a more convenient form by using the following trigonometric identities:

$$\cos^2 \omega t = \tfrac{1}{2}(1 + \cos 2\omega t)$$
$$\cos^3 \omega t = \tfrac{1}{4}(3 \cos \omega t + \cos 3\omega t)$$
$$\cos^4 \omega t = \tfrac{1}{4}(2 \cos 2\omega t + \tfrac{1}{2} \cos 4\omega t + \tfrac{3}{2})$$

[1] See for example, Phillip Franklin, *Methods of Advanced Calculus*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1944, pp. 18–21.

and so on. Substituting into the equation for i_p and collecting like terms yields a series for the plate current of the following form:

$$i_p = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \ldots \quad (11.4)$$

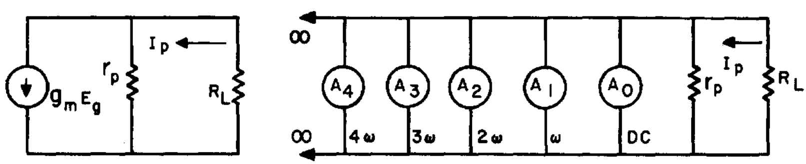
This is the equation for the output current flowing in response to a sinusoidal input. Three things are clearly evident:

(1) A d-c term A_0 is produced.

(2) The original frequency term $\cos \omega t$ appears in the output. This is the *fundamental.*

(3) An indefinite number of higher frequency terms, each integrally related to the fundamental, are produced. These terms are the *harmonics.*

Only a comparatively minor alteration is required to adapt the linear equivalent circuit for use in the nonlinear case. The required change is apparent from equation (11.4). In place of the single funda-



(a) EQUIVALENT PLATE CIRCUIT OF A LINEAR TRIODE OR PENTODE

(b) EQUIVALENT PLATE CIRCUIT OF A NON-LINEAR TRIODE OR PENTODE

Fig. 11.12. Class A equivalent circuits of triodes and pentodes.

mental frequency current source in the linear equivalent circuit, it is necessary to add an infinite array of generators to include the d-c component, fundamental, and harmonics that are generated in the nonlinear case. This is illustrated in figure (11.12). According to the superposition theorem, the behavior of the circuit to each individual current source can be computed separately and the complete circuit solution obtained by summing individual solutions.

Equation (11.4) is useful in specifying the *effects* produced by nonlinearity. Unfortunately, in the form in which it is presented, the equation is impractical because the coefficients, A_0, A_1, A_2, and so on, are all unknown. It is true that they could be expressed in terms of the coefficients, a_1, a_2, a_3, and so on, of the original power series. This is not helpful, however, because it is extremely difficult to measure derivatives of orders higher than the first. A more practical and highly useful method is given in the next section.

11.4. Evaluation of the Coefficients in the Harmonic Series

If the signal voltage applied to a nonlinear circuit element is sinusoidal, the resulting current in the element will consist of a d-c term, a fundamental, and an infinite series of harmonics. This was proved in the preceding section. Thus, if $e_s = E_s \cos \omega t$, the corresponding current flow is

$$i_p = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \dots$$

where the coefficients are as yet undetermined.

The practical numerical evaluation of the coefficients in the preceding series requires a little technical insight and some prior knowledge of the properties of the nonlinear circuit element involved. That is, to solve for the coefficients in any practical case, the series must be made *finite*, rather than infinite, so that the number of unknown constants is finite.

In *most* cases involving vacuum and semiconductor diodes and vacuum triodes, reasonable accuracy results if only the first three terms in the series are retained. For tetrodes and pentodes it is usually necessary to retain at least the first four terms. Hence a fairly general evaluation of the constants can be made if the first five terms in the series are retained. Thus it is assumed that the signal component of current can be closely approximated by the following finite series:

$$i_p \doteq A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + A_4 \cos 4\omega t \quad (11.5)$$

The equation for the signal current now contains five unknowns; it is necessary to write five independent equations for i_p to solve for the constants. In effect, five *different* points must be selected on the curve of i_p vs. ωt and then the values of the five constants are adjusted to make the computed curve pass through these five points on the actual curve. It is assumed that the deviation of the computed curve from the actual curve between these points will be negligibly small. If it is not, more points must be selected and an additional term included in the series for each added point. Thus the procedure used to evaluate the coefficients in the series involves merely the rather arbitrary selection of five values of ωt. For each ωt the actual value of i_p is determined. This data is substituted into equation (11.5). The five equations thereby produced are solved simultaneously for the five unknown coefficients.

Figure (11.13) shows that the signal component of current is symmetrical about a line as shown. Because of this symmetrical characteristic, all five of the points should be chosen in one of the two half cycles, otherwise the points will not be *different*. It is suggested that i_p be evaluated at angles of 0, 60, 90, 120, and 180 degrees. There is nothing significant about this choice of points. Other convenient values might be used with equal success, depending upon the regions of interest.

From this point on it is difficult to discuss the procedure in general

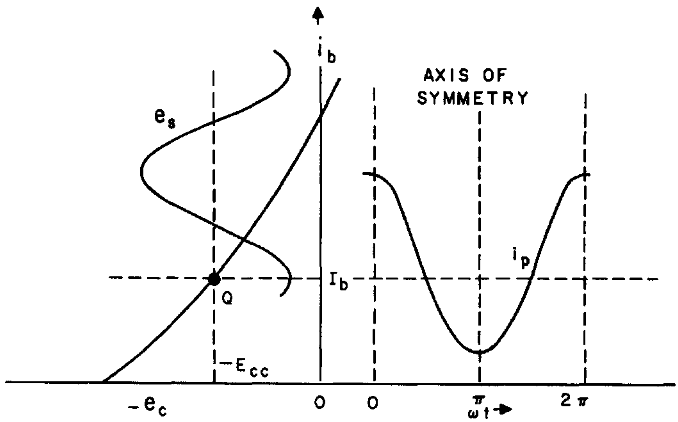


Fig. 11.13. Symmetry in the signal component of plate current.

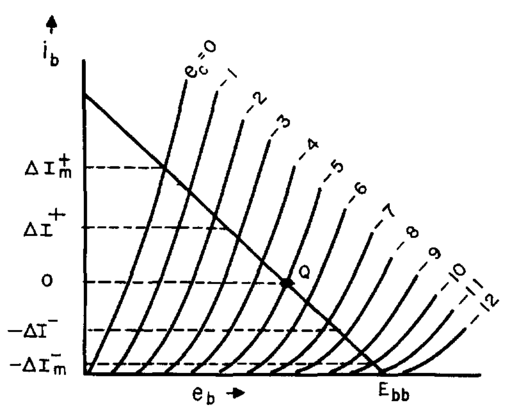


Fig. 11.14. Graphical constructions and terminology used in the evaluation of the harmonics; 5v grid swing assumed.

terms. Therefore the evaluation of the coefficients will be performed for the special case of a triode, but it must be understood that the procedure itself is perfectly general and can be applied to any nonlinear circuit element. The method can be briefly summarized as follows:

(1) Draw the load line on the static plate characteristics of the tube and locate the Q point.

(2) Specify the amplitude of the grid signal.

(3) For this value of E_g, evaluate $e_g = E_g \cos \omega t$ and the corresponding value of i_p for each of the five values of ωt assumed.

(4) Substitute these values for i_p and ωt into equation (11.5).

The graphical constructions and operations are shown in figure (11.14). The resulting five equations are

when $\omega t = 0°$;

$$\Delta I_m^+ = A_0 + A_1 + A_2 + A_3 + A_4$$

when $\omega t = 60°$;

$$\Delta I^+ = A_0 + \tfrac{1}{2}(A_1 - A_2) - A_3 - \tfrac{1}{2}A_4$$

when $\omega t = 90°$;

$$0 = A_0 + 0 - A_2 + 0 + A_4$$

when $\omega t = 120°$;

$$-\Delta I^- = A_0 - \tfrac{1}{2}(A_1 + A_2) + A_3 - \tfrac{1}{2}A_4$$

when $\omega t = 180°$;

$$-\Delta I_m^- = A_0 - A_1 + A_2 - A_3 + A_4$$

Solve these five equations simultaneously for the five unknown coefficients to obtain the following results:

$$A_0 = \frac{1}{6}(\Delta I_m^+ - \Delta I_m^-) + \frac{1}{3}(\Delta I^+ - \Delta I^-) \tag{11.6}$$

$$A_1 = \frac{1}{3}(\Delta I_m^+ + \Delta I_m^-) + \frac{1}{3}(\Delta I^+ + \Delta I^-) \tag{11.7}$$

$$A_2 = \frac{1}{4}(\Delta I_m^+ - \Delta I_m^-) \tag{11.8}$$

$$A_3 = \frac{1}{6}(\Delta I_m^+ + \Delta I_m^-) - \frac{1}{3}(\Delta I^+ + \Delta I^-) \tag{11.9}$$

$$A_4 = \frac{1}{12}(\Delta I_m^+ - \Delta I_m^-) - \frac{1}{3}(\Delta I^+ - \Delta I^-) \tag{11.10}$$

Thus, by using these equations and a graphical evaluation of the type shown in figure (11.14), it is a relatively simple problem to compute the amplitudes of the various harmonic components of the plate current.

It is common practice to express the harmonic content of a wave in terms of the amplitude of the harmonic relative to the amplitude of the fundamental. Hence, in terms of per cent,

$$\%\text{ 2nd harmonic} = \frac{A_2}{A_1} \times 100\% = D_2$$

$$\%\text{ 3rd harmonic} = \frac{A_3}{A_1} \times 100\% = D_3$$

$$\%\text{ 4th harmonic} = \frac{A_4}{A_1} \times 100\% = D_4$$

and so on. The formulas for the harmonics, expressed in percent, can then be written directly from equations (11.6) through (11.10) as

$$D_2 = \frac{3}{4} \frac{(\Delta I_m^+ - \Delta I_m^-)}{(\Delta I_m^+ + \Delta I_m^-) + (\Delta I^+ + \Delta I^-)} \times 100\% \tag{11.11}$$

$$D_3 = \frac{1}{2} \frac{(\Delta I_m^+ + \Delta I_m^-) - 2(\Delta I^+ + \Delta I^-)}{(\Delta I_m^+ + \Delta I_m^-) + (\Delta I^+ + \Delta I^-)} \times 100\% \tag{11.12}$$

$$D_4 = \frac{1}{4} \frac{(\Delta I_m^+ - \Delta I_m^-) - 4(\Delta I^+ - \Delta I^-)}{(\Delta I_m^+ + \Delta I_m^-) + (\Delta I^+ + \Delta I^-)} \times 100\% \tag{11.13}$$

Occasionally the total harmonic content of a waveform is specified in terms of a single factor as follows:

$$D = \sqrt{D_2^2 + D_3^2 + D_4^2 + \dots}$$

However, the relative undesirability of odd and even harmonics can often be a major factor. The expression for the total harmonic content does not reveal the relative magnitudes of the harmonics involved, and its use then acts to confuse the problem. It is generally best to specify the individual harmonic terms separately.

The formulas developed in this section included only the first five terms in the original series. If this analysis does not provide enough accuracy, the same method can be followed for a larger number of points to obtain any specified precision. However, it must be remembered that when high precision is required, the published characteristics of vacuum tubes are not suitable for computational work because of the permissible latitude in manufacturing tolerances. When a higher degree of accuracy is required and more points are used in the analysis, the characteristics of the tubes actually being used should be evaluated experimentally. Of course, if a harmonic wave analyzer of the right characteristics is available, the entire computation may be obtained experimentally.

11.5. Some Practical Considerations of Nonlinearity

The effects and results of nonlinearity in the transfer characteristics of many electronic circuit elements are of great interest. The effect is responsible for many major problems, and in other cases it provides the essential operating principle.

An obvious *application* of the effects of nonlinearity is implied by the nature of the result. The fact that a single frequency sinusoidal input generates a multiple frequency output suggests the use of nonlinear circuit elements as *frequency multipliers*. In other words, an input signal of 1000 c will generate an output having components at 1000 c, 2000 c, 3000 c, and so on. If the output circuit is designed to pass only one of the harmonic terms it is clear that frequency multiplication is produced.

Of course, the existence of harmonics in the output indicates the presence of nonlinear distortion. If fidelity of transmission of the input signal is important, the accuracy with which the input is reproduced in the output can be specified in terms of the amount of distortion present. It will be shown later that harmonic distortion is a primary determinant in the selection of the operating point and load resistance of power amplifiers.

The input signal is sinusoidal and does not have a d-c component. It was shown that the output current did have a constant component A_0. This constant term is important for two reasons:

(1) It shows that the application of a sinusoidal signal input to nonlinear circuit elements causes the Q point to shift from the quiescent value I_b to a higher current, $I_b + A_0$. This is called the *dynamic shift of the Q point*.

(2) Because a direct current is generated as a result of a sinusoidal signal input, the nonlinear element might be used to convert a high frequency signal into a direct current. If the magnitude of the input voltage varies at a slow rate compared with the frequency of the signal, the d-c component fluctuates about a mean value in approximately the same manner as the input signal varies in magnitude about a mean value. Devices operating in this manner are used as *detectors* to convert *amplitude modulated* waves into an audio frequency signal. It is often

called a *square law detector* because the d-c output is mainly developed by the square term in the original power series.

Thus far it has been assumed that the signal input to the nonlinear element consisted of a single frequency sine wave. Suppose the input contains two terms of different frequency, such as $e_s = E_1 \cos \omega_1 t + E_2 \cos \omega_2 t$. Now substitute this function into the general power series previously given for the plate current. Multiply terms out and simplify. The trigonometric identities given in section (11.3) will be helpful, together with

$$\cos \omega_1 t \cos \omega_2 t = \tfrac{1}{2} \cos (\omega_1 - \omega_2)t + \tfrac{1}{2} (\cos (\omega_1 + \omega_2)t$$

After simplification, and if only the first two terms in the power series are considered, the results will have the following form:

$$i_p = B_0 + B_1 \cos \omega_1 t + B_2 \cos \omega_2 t + B_3 \cos 2\omega_1 t + B_4 \cos 2\omega_2 t$$
$$+ B_d \cos (\omega_1 - \omega_2)t + B_s \cos (\omega_1 + \omega_2)t$$

In addition to the expected d-c term, fundamentals, and second harmonics, there are also two new frequencies equal to the sum and difference of the input signal frequencies. If more terms had been considered in the power series, more complicated terms would appear in the output. The sum and difference frequency components are called the *cross-modulation products*. This name is partially governed by the fact that the terms arise from the cross products of the two frequencies in the square factor in the power series.

The cross-modulation products do not bear any particular harmonic relationship to either signal frequency, so that they are generally inharmonious.

Cross modulation has a potential application. Because the sum and difference frequencies represent a frequency displacement of one signal up or down the frequency scale by an amount equal to the other frequency, *frequency translation* is effected. This effect is required in all superheterodyne radio receivers.

The technique of frequency translation is widely used in practical systems to produce *square law modulation*, *mixers*, and *beat frequency oscillators*. In point of fact, however, most frequency translation systems in practical use employ the switching characteristics of electronic devices and not the square law characteristics.

The discussion has been restricted to those circuits in which the amplitude of the input signal is large and nonlinear distortion results. There are other circuits in which the signal amplitudes are small and the harmonic distortion is negligible; operation is essentially linear. Still, a practical application is made of the nonlinear transfer characteristic. By using a special type of construction, it is possible to make a tube have a very curved static transfer characteristic. The slope of this curve at any point is the mutual transconductance g_m of the tube. Thus, because of the curvature in the transfer characteristic, g_m varies over a wide range of values. Tubes of this type are called *variable mu* or *remote cutoff* tubes.

The variation in g_m possible with remote cutoff tubes finds extensive application in gain control of linear amplifiers, particularly in automatic control systems. It was shown in chapter 3 that the gain of any voltage amplifier can be expressed by $A = -g_m' Z_m$. Variations in g_m, and hence in g_m', cause corresponding changes in the voltage gain. Thus, by using variable mu tubes, a convenient method of gain control is available simply by adjusting the Q point of the tube. This technique is widely used in *automatic volume control* (AVC) systems in commercial radio receivers.

11.6. An Idealized Class A Vacuum Tube Power Amplifier

Nonlinear distortion is a major factor in the design of class A power amplifiers; the distortion is caused by the large signal amplitudes required in power amplifiers. It will be proved here that large signals are required.

The class A triode circuit in figure (11.1) is the basic circuit diagram of a class A vacuum tube power amplifier. The corresponding class A equivalent plate circuit is shown in figure (11.12). The principle of superposition can be applied, so that the fundamental component of load current can be computed independently of the distortion components. That is, the fundamental plate current component is

$$I_p = g_m E_g\left(\frac{r_p}{r_p + R_L}\right) = g_m E_g\left(\frac{1}{1 + R_L/r_p}\right) \qquad (11.14)$$

The power dissipated in the load resistance is

$$P_L = \frac{1}{2} I_p^2 R_L = g_m^2 E_g^2 \left(\frac{1}{1 + R_L/r_p}\right)^2 \frac{R_L}{2} \qquad (11.15)$$

Replace one of the g_m terms by μ/r_p and the equation for the load power becomes

$$P_L = \mu g_m \frac{R_L/r_p}{(1 + R_L/r_p)^2} \cdot \frac{E_g^2}{2} \qquad (11.16)$$

This is an important equation even though its practical use is somewhat limited. It shows that the power output can be increased by increasing μ and g_m. The μg_m term can be used as one convenient criterion or figure of merit by which different power amplifier tubes can be compared.

Equation (11.16) also shows that the power output is proportional to the square of the grid signal voltage E_g. This dependence is illustrated graphically in figure (11.15). It is immediately clear that it is generally desirable to use large grid signal voltages. Thus power amplifiers are often called *large signal* amplifiers. The method of computing the distortion that results was outlined in the preceding section.

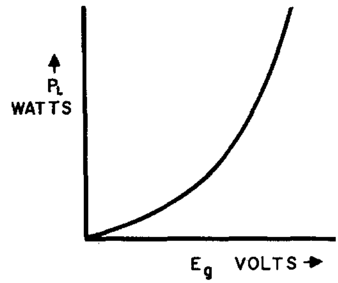


Fig. 11.15. Effect of variations in the amplitude of the grid signal voltage on the power output.

You can see from equation (11.16) that the power output is governed by the size of the plate load resistance R_L relative to the plate resistance

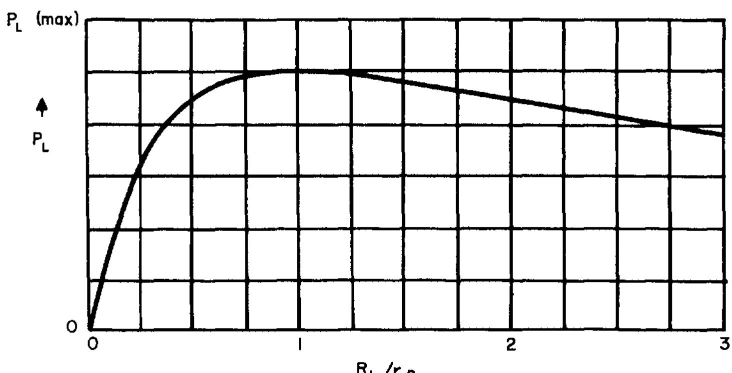


Fig. 11.16. Effect of variations in load resistance on power output.

of the tube. If this dependence is computed and plotted, the result appears as shown in figure (11.16). The maximum power output is obtained when $R_L = r_p$. However, the establishment of this operating condition might result in an undue amount of distortion, so that this does not necessarily represent the *optimum* condition. Actually, it is common practice to use values of R_L two or three times larger than r_p, though this should not be taken as a specific rule of thumb. The reason for this will be apparent later.

Although the preceding analysis assumed a triode, exactly the same results are obtained for a pentode. However, for a pentode, r_p is much larger than R_L, and

$$P_L \doteq \mu g_m \frac{R_L}{r_p} \cdot \frac{E_g^2}{2} \tag{11.17}$$

Hence the power output increases almost linearly with increasing load resistance.

11.7. Harmonic Distortion in Triodes and Pentodes

The preceding section proved that class A power amplifiers are customarily operated with large signal inputs that move the operating point over a nonlinear range of the transfer characteristics. Because the power output is governed by the size of the load resistance, it is important to determine the effect of load resistance variations on the harmonic distortion.

A method of computing the harmonics caused by nonlinear distortion was given in section (11.4) and some useful formulae were derived. This same technique is followed to compute the harmonics generated for various values of load resistances, so that the graphical steps required appear as shown on figure (11.17). These figures show three possible load lines all going through the same Q point and intercepting the same limiting characteristic curves. In other words, the Q point and grid signal voltage are the same for all values of load resistance. By following the method given in section (11.4), the harmonics generated in the plate current can be computed and the results plotted as shown in figure (11.18). The same general kind of results will be obtained for any power triode, beam tetrode, or pentode.

These curves show that the harmonic distortion is determined to a large extent by the value of the load resistance. Therefore the final choice of the value for R_L may be made on the basis of the allowable distortion rather than upon the condition for maximum power output. As a result, it is rare to find the load resistance set equal to the plate resistance of the tube in practical cases. For triodes it is often two or three times the size of r_p, because this reduces the harmonic distortion by a large factor without causing too great a reduction in the power output. Pentodes are often operated with the value of R_L required for minimum distortion and this is much less than the r_p of the tube.

Superficially it would appear desirable to operate a pentode with the load resistance set to produce minimum over-all distortion. This may or may not be true, depending upon the application. For audio work,

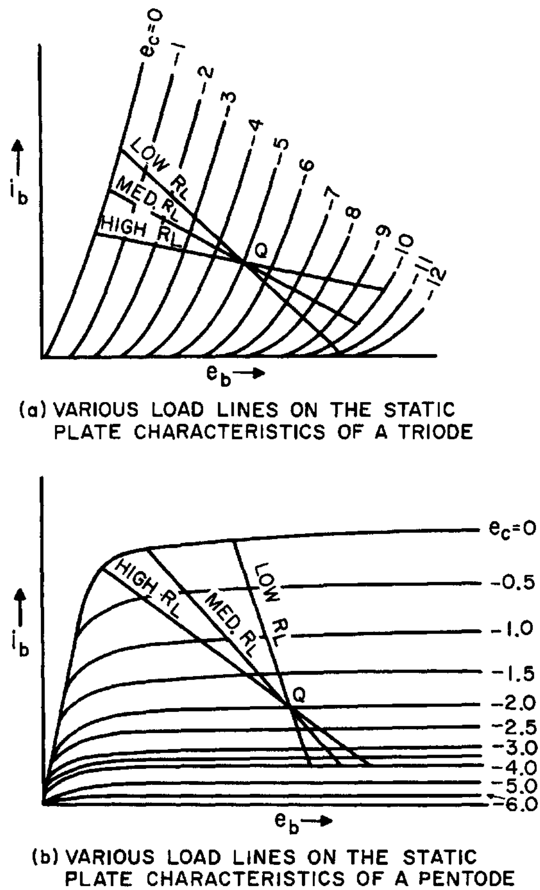
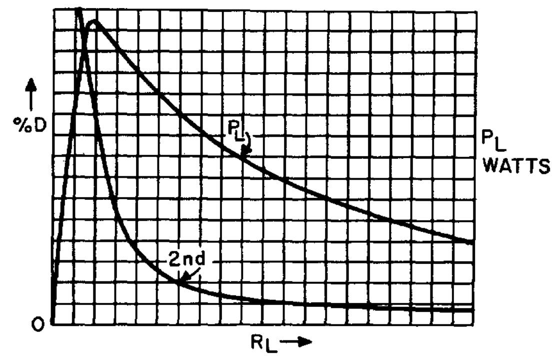


Fig. 11.17. Preliminary graphical steps required in the computation of the harmonic distortion as a function of load resistance.

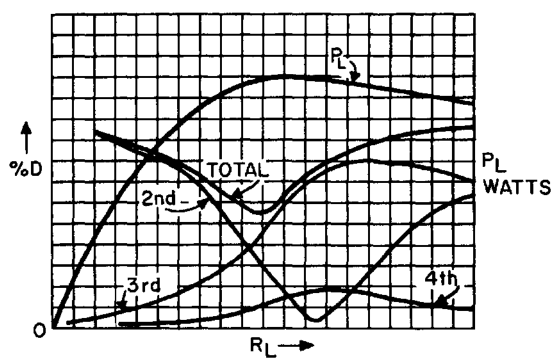
the third harmonic is generally more objectionable than the second harmonic; higher percentages of second harmonic distortion can be tolerated.

The correct choice of the load resistance is obviously a difficult task. Because the distortion will also be affected by the location of the Q

point, common sense suggests that the preceding procedure and results would have to be developed for a series of operating points. Finally, a selection of R_L and the Q point could be made.



(a) POWER OUTPUT AND SECOND HARMONIC DISTORTION IN A TRIODE AS A FUNCTION OF LOAD RESISTANCE



(b) POWER OUTPUT AND HARMONIC DISTORTION IN A PENTODE AS A FUNCTION OF LOAD RESISTANCE

Fig. 11.18. Principal harmonics generated in triodes and pentodes as a function of load resistance.

The results of all this graphical construction and evaluation can be represented by a series of curves such as those shown in figure (11.19). This illustrates the dependence of the distortion terms and signal input on the power output.

Fortunately, in most cases this evaluation has been performed by the tube manufacturer and published in tube manuals. It is generally advisable to follow his recommendations unless some special requirements dictate otherwise.

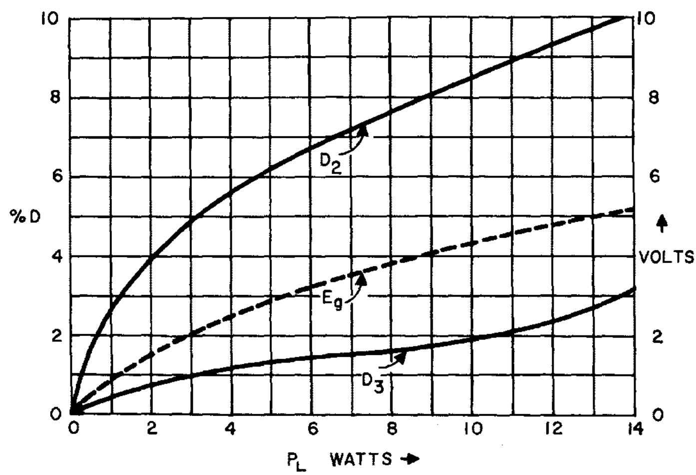


Fig. 11.19. Distortion and grid signal voltage vs. power output.

11.8. Plate Circuit Efficiency

In the consideration of a power amplifier circuit it is inevitable that the efficiency of power conversion from direct to alternating current will be important. This quantity is defined as follows:

$$\eta_p = \text{plate circuit efficiency} = \frac{P_{ac}}{P_{dc}} \tag{11.18}$$

where P_{ac} = signal power output; P_{dc} = d-c power input from the plate power supply. This factor is not important in the design of voltage amplifiers because the power levels are quite low. However, in power amplifiers, the conversion efficiency is important because both the signal power output and d-c power input are appreciable.

It will be shown later that the efficiency of power conversion can be materially increased by operating the power tube in the switching mode. This is one of the most important advantages of operation in the switching mode.

Consider the elementary class A power amplifier shown in figure (11.1). Define terms as follows:

i_b = instantaneous plate current $= I_b + I_p \sin \omega t$
I_b = quiescent plate current
I_p = amplitude of signal current component
e_b = instantaneous plate voltage $= E_b + E_p \sin \omega t$
E_b = quiescent plate voltage
E_p = amplitude of signal component of plate voltage
E_{bb} = plate supply voltage
R_L = load resistance

The signal power output is the signal power dissipated in the load resistance R_L. Hence

$$P_{ac} = \frac{1}{2} I_p^2 R_L = \frac{1}{2} \frac{E_p^2}{R_L} = \frac{1}{2} E_p I_p \tag{11.19}$$

The d-c power input from the plate power supply is

$$P_{dc} = E_{bb} I_b. \tag{11.20}$$

The d-c power input must supply all the power in the circuit. The various components of the total power supplied from this one source are: P_{ac} = signal power dissipated in R_L; P_L = d-c power loss in R_L; P_p = power lost in the tube = plate dissipation. Hence $P_{dc} = P_{ac} + P_L + P_p$, and the plate circuit efficiency can be written

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_L + P_p} \tag{11.21}$$

It is immediately clear that the efficiency of power conversion is prevented from attaining 100% by two terms: P_L, the power lost through d-c heating of R_L, and P_p, the power dissipated on the plate of the tube. A reduction in either of these two losses would increase the plate circuit efficiency.

The plate circuit efficiency is easily computed by substituting equations (11.19) and (11.20) into (11.18). The result is

$$\eta_p = \frac{1}{2} \frac{I_p^2 R_L}{E_{bb} I_b} = 50 \frac{I_p^2 R_L}{E_{bb} I_b} \quad \text{per cent} \tag{11.22}$$

The transfer characteristic of a triode is shown in figure (11.20). Operation in class A over the maximum possible range results if the plate current swings over a range of values from 0 to $2I$ as shown in figure (11.20). The same operation is shown on the idealized static plate characteristics of figure (11.21). If the Q point is located midway between 0 and $2I$, the amplitude of the signal component of plate current must be exactly equal to I_b. Therefore $I_p = I_b$ and the plate circuit efficiency becomes

$$\eta_p = 50\frac{I_bR_L}{E_{bb}} \quad \text{per cent} \tag{11.23}$$

In the absence of any signal input voltage,

$$E_{bb} = E_b + I_bR_L$$

or

$$I_bR_L = E_{bb} - E_b \tag{11.24}$$

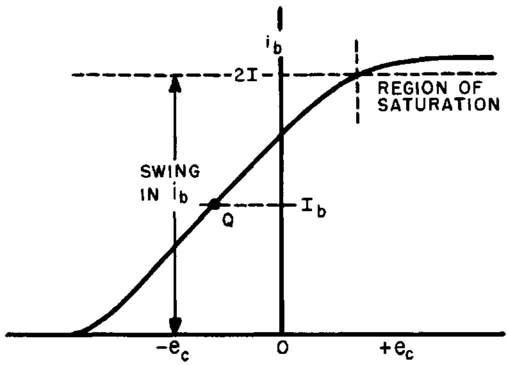


Fig. 11.20. Transfer characteristic showing range of operation.

Substitute this equation into (11.23) for the plate circuit efficiency to obtain

$$\eta_p = 50\left(1 - \frac{E_b}{E_{bb}}\right) \quad \text{per cent} \tag{11.25}$$

To obtain the maximum possible theoretical efficiency it is necessary to determine the condition that will make the signal power a maximum. The signal power is $P_{ac} = \frac{1}{2}I_p^2R_L$. Maximize this with respect to R_L. This requires that $I_p^2 = 0$. Thus the power output will be a maximum when the signal component of the plate current is zero. This can occur only by making the load resistance infinite so that the load line becomes horizontal. This makes the load line correspond with the horizontal axis of figure (11.21) and makes $E_p = E_b = \frac{1}{2}E_{bb}$. This makes the plate circuit efficiency given in equation (11.25) have a maximum theoretical value of 25%.

Twenty-five per cent is a rather low figure. Moreover, it results only under the highly theoretical condition of an infinite load resistance. Actual values of efficiency are about 10% to 12%.

The two factors that hold the maximum theoretical efficiency down to 25% are the d-c power loss in the load resistance and the plate dissipation in the tube. It will be shown in the next section that the theoretical efficiency can be increased to 50% if the d-c loss in the load can be avoided.

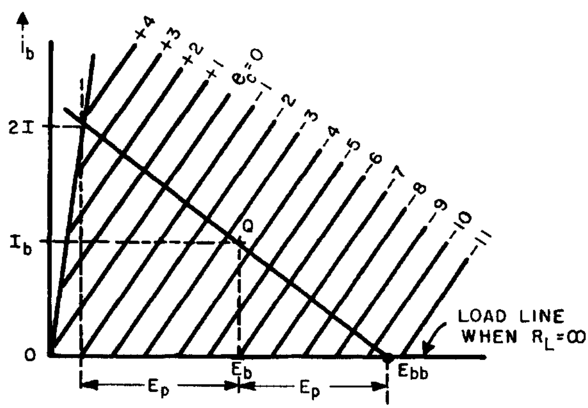


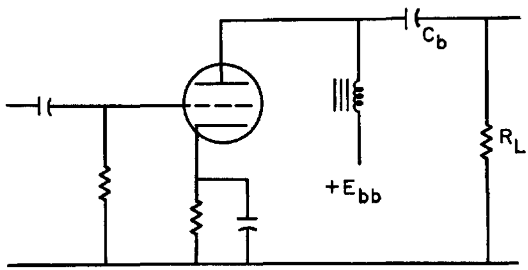
Fig. 11.21. Relations in a power amplifier with an idealized triode.

11.9. Methods of Feed, a-c and d-c Load Lines

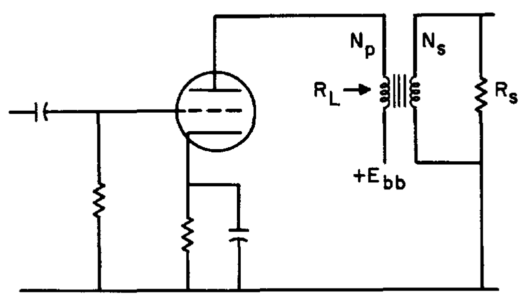
The d-c power loss in the load resistor acts to limit the maximum theoretical efficiency of a class A vacuum tube power amplifier to only 25%. If this loss can be avoided, an increase in efficiency will be observed.

Two methods that virtually eliminate the d-c loss in R_L are shown in figure (11.22). The first scheme is called *shunt feed* because the d-c power input to the circuit is supplied in parallel with the load. Thus the d-c current flows through the low resistance choke coil L to the tube, but is prevented from flowing through the load resistance by the blocking capacitor C_b. The only power loss in the circuit is the d-c loss in the choke coil, and this is small. The choke coil prevents the signal power from being partially dissipated in the internal impedance of the power supply or being short circuited to ground. The impedance of the choke must be high at all frequencies of interest so that no appreciable amounts of signal current flow through it. Thus, for all practical purposes, all the signal current passes through R_L. The term *shunt feed* is used to distinguish the circuit from the *series feed* connection shown in the rudimentary power amplifier of figure (11.1).

Transformer coupling, shown in figure (11.22), accomplishes the same effect as shunt feed. In the circuit, in the absence of any grid signal, the impedance between the tube and the power supply is the d-c resistance of the transformer primary. This will be quite small in



(a) SHUNT FEED



(b) TRANSFORMER COUPLING

Fig. 11.22. Circuit connections that will reduce the d-c power loss in the load circuit.

a well-designed transformer; the d-c power loss will be correspondingly small. When a signal input is provided, the load is coupled to the tube through the mutual inductance of the transformer.

The distinction between shunt feed and transformer coupling is essentially a practical matter. The shunt fed amplifier is used when the load resistance is comparatively large; the transformer coupled amplifier is used when the load resistance is small compared with the plate resistance of the tube.

The plate circuit efficiency, when there is no d-c power loss in the load resistance, is

$$\eta_p = \frac{P_{ac}}{P_{dc}} = \frac{P_{ac}}{P_{ac} + P_p} \tag{11.26}$$

In the absence of any grid signal voltage the plate circuit impedance is zero. Hence the d-c or zero-signal load line is nearly vertical as shown in figure (11.23). The Q point is located using this load line. When a grid signal voltage is supplied, the impedance of the choke at the signal frequency is so high that it is effectively an open circuit, while the blocking capacitor is virtually a short circuit. Thus the total plate circuit impedance is equal to the load resistance R_L. Therefore a new

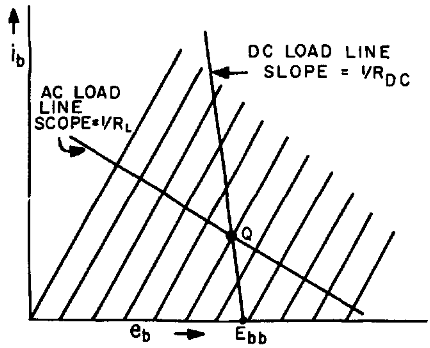


Fig. 11.23. A-c and d-c load lines for the shunt fed or transformer coupled amplifier.

load line, called the a-c load line, must be drawn through the Q point with a slope equal to $1/R_L$. This is shown in figure (11.23).

The plate circuit efficiency is defined in equation (11.26). If the cathode bias voltage of either of the amplifiers of figure (11.22) is neglected, the plate voltage is equal to the power supply voltage. Thus $E_b = E_{bb}$. Hence, the d-c power input from the power supply is

$$P_{dc} = E_{bb}I_b = E_bI_b \tag{11.27}$$

The signal power in the load resistance is

$$P_{ac} = \tfrac{1}{2}E_pI_p \tag{11.28}$$

The plate circuit efficiency is therefore

$$\eta_p = 50\,\frac{E_pI_p}{E_bI_b} \quad \text{per cent} \tag{11.29}$$

If the tube characteristics are assumed to be straight, parallel, and equally spaced, the plate current can be driven from zero to a maximum value of $2I$. With the Q point located midway between these two extremes, $I_b = I_p = I$. The plate circuit efficiency is then

$$\eta_p = 50\frac{E_p}{E_b} \quad \text{per cent} \tag{11.30}$$

Now suppose that the load resistance is infinite. In this purely theoretical case the a-c load line is horizontal and passes through the Q point. The maximum value of the signal component of plate voltage is equal to E_{bb}, because any larger value would drive the plate negative. Hence, if $E_p = E_b = E_{bb}$, the theoretical maximum efficiency of the shunt feed or transformer coupled amplifier is 50%. This is a substantial improvement over the 25% computed for the series fed amplifier.

Although the theoretical maximum efficiency is 50%, practical values will range in the neighborhood of 25%.

11.10. Plate Dissipation

The equation for the plate circuit efficiency of a shunt fed or transformer coupled amplifier can be written

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_p} \tag{11.31}$$

Solve equation (11.31) for the signal power output; the result is

$$P_{ac} = P_p\frac{\eta_p}{1 - \eta_p} \tag{11.32}$$

This shows that the signal power output is linearly dependent upon the power dissipated on the plate of the tube. For a fixed plate circuit efficiency η_p, the largest possible plate dissipation is used to achieve the maximum possible power output. In fact, the maximum possible power output from the tube is controlled by the amount of power that can be dissipated on the plate.

The plate dissipation is the power consumed by the tube. It is an unavoidable loss caused by the release of the kinetic energy of the electrons when they strike the plate. Proper operation of the tube requires that the power supply energize the electrons and cause them to traverse the interelectrode space. It then becomes the duty of the plate to absorb this kinetic energy from the electrons. The release of electronic kinetic energy causes the plate temperature to rise to some particular value governed by the rate that energy is supplied by the electrons and removed by various agencies from the plate. In small tubes radiation may be the only heat removing agency. In larger tubes, the plate may be cooled by radiation and circulating water or air.

As the power input to the plate is increased the plate temperature rises. Eventually the temperature is so high that the occluded gases, mainly hydrogen, carbon monoxide, nitrogen, and carbon dioxide, are driven out of the plate and into the interelectrode space. These released gas atoms are ionized by the high energy electrons, and the tube arcs over into heavy conduction through the formation of a sustained gas discharge. As a result, it is customary to rate a tube in terms of an allowable plate dissipation that will not permit the plate temperature to exceed the critical point. According to equation (11.32), this automatically imposes an upper limit on the maximum signal power output.

Solve equation (11.31) for the plate dissipation. This yields

$$P_p = P_{ac}\left(\frac{1}{\eta_p} - 1\right) \tag{11.33}$$

A given design problem will generally specify the required signal output P_{ac}. The plate circuit efficiency η_p may be specified by other considerations. Then equation (11.33) can be used to compute the required amount of plate dissipation. This value is then used to select an appropriate tube for use in the projected amplifier.

11.11. The Parallel Connection of Power Amplifiers

The preceding sections concentrated upon the characteristics of single tube power amplifiers. It is useful now to inquire into the possible effects produced by various connections of multitube power amplifiers. The simplest such combination of tubes is the *parallel connection.*

A typical parallel connection is shown in figure (11.24). In terms of the series expansion of the plate current, the current through either tube can be written

$$i_b = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \ldots \tag{11.34}$$

If the tubes are assumed to be identical, the two plate currents are equal and the total current is

$$i_T = 2(A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \ldots) \tag{11.35}$$

Hence the total current is simply twice the current through a single tube. Consequently, the fundamental, harmonics, and d-c term are doubled. Of course, when two tubes are operated in parallel, the *total load* will be *half* the *load resistance* of a single tube. Therefore the load power will be doubled when two identical amplifier stages are paralleled. The percentages of harmonic distortion remain unchanged. Thus the main advantage of the parallel connection is the increased signal power output.

It might appear that the increase in power output is not advantageous because the same effect can be achieved by using a single tube with a larger rated plate dissipation. This is quite true, but it might be more economical to buy two small tubes than one large one. Moreover, the parallel connected amplifier is generally more reliable because it would continue to operate even if one tube fails. The decision as to the

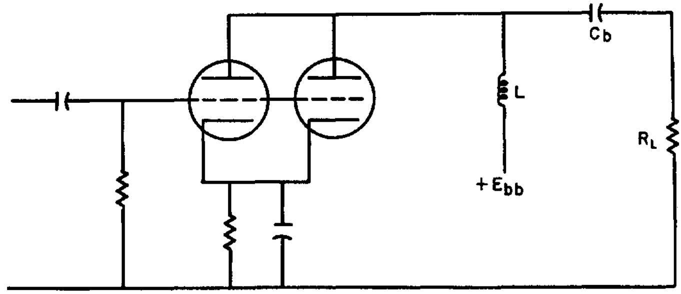


Fig. 11.24. Parallel connection of shunt fed amplifiers.

advisability of paralleling two tubes, or more, or using a single tube with a large rated plate dissipation can be made only in the light of a specific application.

11.12. Push-Pull Amplifiers

In parallel connected vacuum tube power amplifiers, the grid voltages of the two tubes are identically the same. This *in-phase* relationship between grid voltages produces plate currents that are also in phase. When the grids are driven by voltages that are mirror images of one another, 180 degrees out of phase in the case of sine waves, the amplifier is said to be connected in *push-pull.* A typical transformer coupled push-pull amplifier is shown in figure (11.25). The only difference between this circuit and a parallel connected transformer coupled amplifier is the relative phase between the grid voltages and the center tap connection of the transformer.

In the general case the plate current of tube 1 can be expressed as a series of the form

$$i_{b_1} = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \dots \quad (11.36)$$

If the tubes are identical and if the grid voltages are 180° out of phase, the plate current in tube 2 is the same as that in tube 1 except that ωt is replaced $(\omega t + \pi)$. Hence

$$i_{b_2} = A_0 + A_1 \cos(\omega t + \pi) + A_2 \cos 2(\omega t + \pi) + A_3 \cos 3(\omega t + \pi) + \dots$$

This expression reduces to

$$i_{b_2} = A_0 - A_1 \cos \omega t + A_2 \cos 2\omega t - A_3 \cos 3\omega t + \dots \tag{11.37}$$

A term-by-term comparison of the equations for the two plate currents reveals an interesting fact: the fundamental and odd harmonics

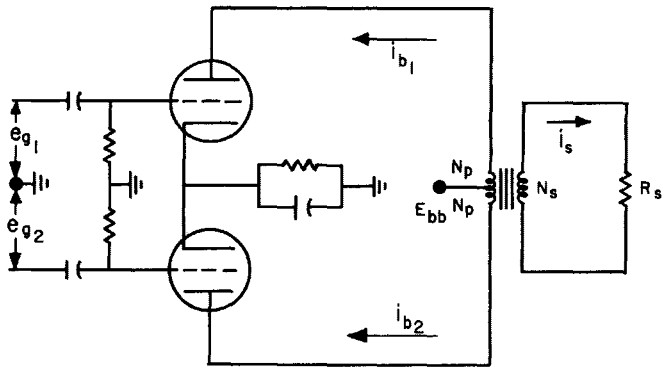


Fig. 11.25. Push-pull connection of transformer coupled amplifiers.

of tube 2 are 180 degrees out of phase with corresponding terms in tube 1, while the even harmonics in the two plate currents are in phase.

Consider the magnetic circuit of the transformer. If leakage fluxes are neglected the primary ampere-turns will equal the secondary ampere-turns. The primary magnetizing force is made up of the two components caused by the currents i_{b_1} and i_{b_2}. These currents flow in opposite directions through the transformer primary so they produce mmf's in opposite directions. Hence

$$N_s i_s = N_p i_{b_1} - N_p i_{b_2} \tag{11.38}$$

or

$$i_s = \frac{N_p}{N_s}(i_{b_1} - i_{b_2}) = \text{secondary current} \tag{11.39}$$

Therefore the load current depends upon the transformer turns ratio and the difference between the two plate currents of the tubes. Substitute for i_{b_1} and i_{b_2} in equation (11.39) to obtain

$$i_s = \frac{N_p}{N_s}(A_1 \cos \omega t + A_3 \cos 3\omega t + \dots) \tag{11.40}$$

Note that only the fundamental and the *odd* harmonics appear in the output; the even harmonics are cancelled by the push-pull connection. This is the main advantage of the push-pull connection, because it occurs with an increase in fundamental power output exactly as if the tubes were in parallel.

The harmonic distortion in triodes, operated Class A, is predominantly second harmonic in nature; thus the push-pull connection will yield an output that is remarkably free from distortion. The distortion in class A tetrodes and pentodes was shown to contain a strong third harmonic so that it appears that the push-pull connection has little to offer when used with such tubes. However, by using a smaller load resistance, the magnitude of the third harmonic can be reduced considerably, while the second harmonic is increased. The increase in D_2 is not too serious because the second harmonic is cancelled from the output in the push-pull connected amplifier. Thus the amplifier is made to have a smaller third harmonic by this method, but the reduction is *not* an inherent property of the push-pull connection.

The *complete* cancellation of even harmonics should not be interpreted too literally for it seldom occurs in practice. The effect depends upon *identical* tubes, and this is an academic fiction. Tube characteristics vary over a wide range, and it would be pure coincidence to locate two that were identical. Thus, while the push-pull connection will reduce even harmonic distortion, in practical cases it will not cancel it to zero.

The total current flowing through the cathode resistor is the sum of the two plate currents, or $i_k = i_{b_1} + i_{b_2}$. Therefore the cathode voltage is

$$\begin{aligned} e_k &= i_k R_k = (i_{b_1} + i_{b_2}) R_k \\ &= 2(A_0 + A_2 \cos 2\omega t + A_4 \cos 4\omega t + \dots) R_k \qquad (11.41) \end{aligned}$$

Thus the odd harmonics and fundamental cancel in the cathode resistor, leaving only the d-c term and even harmonics. The phase of the even harmonic cathode voltage components is such as to introduce negative feedback, thereby reducing the harmonic distortion in the event of an unbalance between tubes. This appears to be desirable and that it would be advantageous to leave the cathode resistor unbypassed. However, if the tubes are not identical, the *odd* harmonics will not cancel completely in R_k. Unfortunately, the phase of the odd harmonic components is such as to produce positive feedback, thereby accentuating the unbalance and increasing the odd harmonic distortion.

Hence it is almost always desirable to bypass the cathode resistor adequately. If this is done, the cathode voltage is $e_k = 2A_0R_k$.

The principal advantage of the push-pull connection is the reduction of even harmonic distortion. Another real advantage is obtained because the mmf's caused by the d.c components of the plate currents cancel in the transformer primary. This reduces the constant core magnetization virtually to zero and allows the use of smaller transformers than those permissible with single tube amplifiers.

11.13. Vacuum Tube Power Amplifiers, Concluding Remarks

Once it has been determined that a class A vacuum tube power amplifier is to be designed, and once the specifications of power output and allowable harmonic distortion have been set, the most immediate problem is the choice of the power amplifier tube. The list of available tubes is a bit bewildering at first because of the many possibilities. The field of choice is narrowed considerably by the specified requirements of power and allowable distortion. The selection is also partly controlled by the power supply requirements. For example, types 25L6 and 50L6 are designed for use in small a-c/d-c equipment. The 3S4 and 3V4 are intended to be used with battery-operated portable equipment where 90 and 67.5 volts are available. The filament voltage required also helps limit the selection. For example, if batteries are used for filament power, low drain 3 volt filaments could be used. On the other hand, it might be desirable to use a series connection of filaments of a number of tubes directly across the 110v a-c line; this makes a filament transformer unnecessary. Hence a 35L6 or 50L6 might be selected. Finally, the amount of grid driving voltage might be a determinant as well as the cost of the tube. Thus, even though a large choice exists, the selection rapidly narrows down to a few tubes.

As a final consideration, it is generally advisable to use the tubes specified by the manufacturer on his *preferred list*. This usually insures better quality, lower initial cost, and ready availability.

For most practical cases the equivalent plate circuit of class A vacuum tube power amplifiers can be drawn as shown in figure (11.12) earlier. The magnitudes of the various current generators are specified as follows:

$A_0 =$ d-c term accounting for the dynamic shift of the Q point.
$A_1 =$ amplitude of the desired fundamental.
$A_2 =$ amplitude of the second harmonic.
$A_3 =$ amplitude of the third harmonic.

These amplitudes are computed from the formulas given in section (11.4).

The power output at the fundamental frequency is

$$P_{L_1} = A_1^2\left(\frac{r_p}{r_p + R_L}\right)^2 \frac{R_L}{2} \tag{11.42}$$

Similarly, the power developed in the load at the second harmonic frequency is

$$P_{L_2} = A_2^2\left(\frac{r_p}{r_p + R_L}\right)^2 \frac{R_L}{2} \tag{11.43}$$

and so on for each generator in the equivalent circuit. Now substitute the formulas for the harmonic amplitudes derived previously in section (11.4) so that

$$P_{L_1} = \frac{R_L}{18}\left[(\Delta I_m^+ + \Delta I_m^-) + (\Delta I^+ + \Delta I^-)\right]^2\left(\frac{r_p}{r_p + R_L}\right)^2$$

$$P_{L_2} = \frac{R_L}{32}(\Delta I_m^+ - \Delta I_m^-)^2\left(\frac{r_p}{r_p + R_L}\right)^2$$

and so on for each term.

11.14. Power Gain of Transistor Amplifiers

The equivalent circuit of any type of transistor amplifier with resistance load was derived in chapter 8 and is reproduced in figure (11.26).

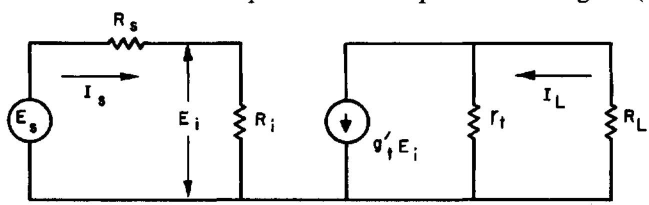


Fig. 11.26. Class A linear equivalent circuit of a simple transistor power amplifier.

Only the mid-frequency case is considered here, so that no capacitive elements are shown. The circuit can be used for any of the three basic amplifier configurations by using the appropriate value for R_i and g_f' derived in chapter 8.

Although the problem of nonlinearity in transistors is more complex than in vacuum tubes, in general the same methods and considerations apply. The actual performance of transistors as power amplifiers and their resulting properties of harmonic distortion were not thoroughly reported at the time this book was prepared. Indeed, the whole problem is somewhat obscured by the lack of a generally acceptable and suitable equivalent circuit. Once this has been established, the methods outlined in this chapter can be used to evaluate the harmonic distortion generated in large signal transistor amplifiers.

Calculations of power gain for various transistor amplifiers can be made with the aid of the formulas derived in chapter 8.

11.15. Push-Pull Transistor Power Amplifiers

Some aspects of push-pull amplification with transistors are interesting and will be discussed briefly.

Two identical transistor amplifiers can be operated in push-pull in essentially the same manner as vacuum tubes. A phase inverting circuit

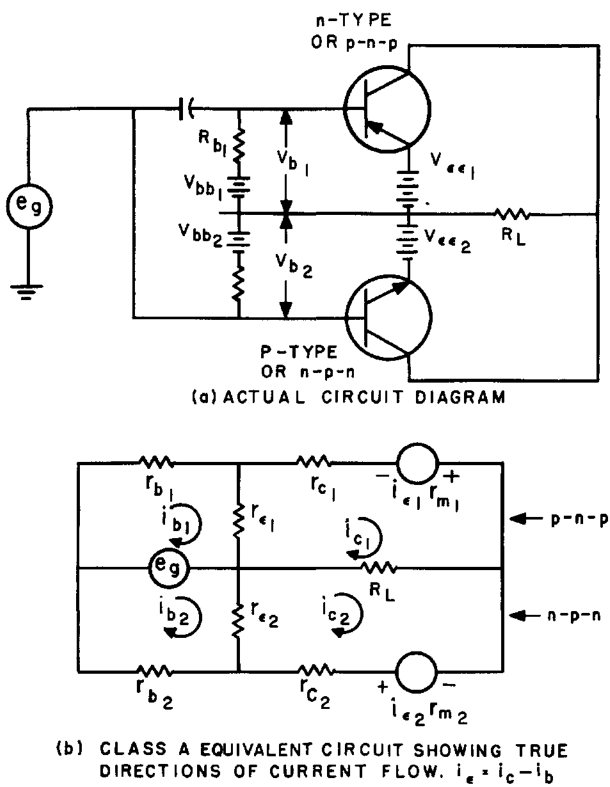


Fig. 11.27. Grounded emitter push-pull transistor amplifier.

on the input is required, together with transformer coupling to the load. Circuit operation then closely follows that of vacuum tube amplifiers.

A much simpler and more ingenious arrangement uses a combination of *n-p-n* and *p-n-p* types. This removes the necessity of a special phase inverting circuit, and the load can be connected directly to the transistors without using an output transformer. A representative circuit is shown in figure (11.27) together with the class A equivalent circuit. The *true* directions of current flow in the transistors are shown (see chapter

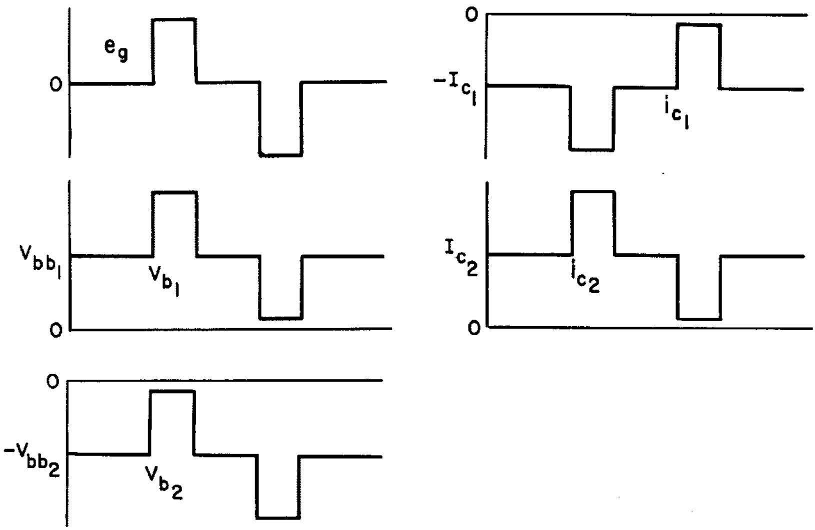


Fig. 11.28. Waveforms in the push-pull transistor amplifier.

8). Note that the currents flow in opposite directions through the load resistance.

Although the same signal voltage is applied to both amplifier inputs, the collector currents so produced are necessarily inverted with respect to one another because one transistor is *p*-type while the other is *n*-type. Thus the two collector currents are

$$i_{c_1} = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + A_3 \cos 3\omega t + \dots$$
$$i_{c_2} = A_0 - A_1 \cos \omega t + A_2 \cos 2\omega t - A_3 \cos 3\omega t + \dots$$

Therefore the load current is

$$i_L = i_{c_1} - i_{c_2} = 2(A_1 \cos \omega t + A_3 \cos 3\omega t + \dots)$$

and push-pull operation clearly results.

The key factor in the operation of this circuit is the inherent phase inverting characteristics of a p-type transistor relative to an n-type. This is illustrated by the waveforms of figure (11.28).

11.16. Square Law Amplitude Modulation

The general problem of modulation is discussed at considerable length in chapter 13. However, it is a process that may be accomplished as a result of tube nonlinearities; thus it seems appropriate to mention the subject at this point.

It is shown in chapter 13 that an amplitude modulated carrier wave has a mathematical form

$$i = I_c \cos \omega_c t + \frac{m_a I_c}{2} [\cos (\omega_c + \omega_m)t + \cos (\omega_c - \omega_m)t]$$

where m_a = modulation index = I_m/I_c.

Now the question is, can a current of this type be obtained as a result of the nonlinearities in vacuum tube characteristics? When two different frequencies are impressed upon a nonlinear impedance having an approximately square law characteristic, it was shown in section (11.5) that the resulting current has the form

$$\begin{aligned} i = B_0 + B_1 \cos \omega_1 t + B_2 \cos \omega_2 t + B_3 \cos 2\omega_1 t \\ + B_4 \cos 2\omega_2 t + B_d \cos (\omega_1 - \omega_2)t \\ + B_s \cos (\omega_1 + \omega_2)t \end{aligned}$$

Now, by defining terms as follows:

$$\omega_1 = \omega_c; \qquad B_1 = I_c; \qquad B_d = B_s = \frac{m_a I_c}{2}$$

$$\omega_2 = \omega_m; \qquad B_2 = I_m$$

the preceding equation assumes the form

$$i = I_c \cos \omega_c t + \frac{m_a I_c}{2} [\cos (\omega_c - \omega_m)t + \cos (\omega_c + \omega_m)t]$$

$$+ B_0 + I_m \cos \omega_m t + B_3 \cos 2\omega_c t + B_4 \cos 2\omega_m t$$

It is clear that the bracketed quantity has precisely the same form as that of the amplitude modulated wave. Hence nonlinear impedance elements such as diodes, triodes, or pentodes, which have approximately square law characteristics, can be used to produce amplitude modulation. Of course, some additional terms are also produced, but it is assumed that it is possible to remove them or to nullify their effects.

Vacuum tube square law modulators are often called *Van der Bijl* modulators. A typical circuit is shown in figure (11.29). The load circuit is usually adjusted so that an appreciable output is obtained only for the three terms required for the modulated wave; these are the carrier and the sum and difference frequency components.

Except in special cases, modulation processes based upon the curvature in the transfer characteristics of vacuum tubes are not too successful in practice. This is so mainly because the amount of nonlinearity is inclined to change with tube age, replacement, and environment. This directly affects the operation of the modulator, tending to make it

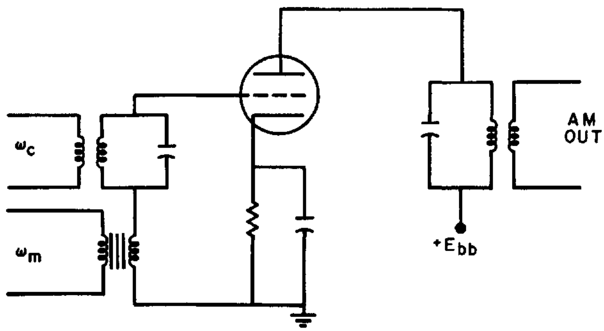


Fig. 11.29. Van der Bijl modulator.

unreliable and unpredictable. More practical methods use tubes operating in the switching mode.

11.17. Square Law Detectors

Detection or *demodulation* is the reverse of the modulation process; the original intelligence signal is recovered from the modulated carrier in a detector.

It was shown in section (11.4) that the load current resulting from the application of a sinusoidal signal input to a nonlinear circuit element is of the form

$$i = A_0 + A_1 \cos \omega_c t + A_2 \cos 2\omega_c t + A_3 \cos 3\omega_c t + \ldots$$

The d-c component A_0 is approximately linearly proportional to the amplitude of the signal input. Hence if the signal amplitude varies at a rate that is slow compared with ω_c, then A_0 will vary in a similar manner. In the case of an amplitude modulated wave, the amplitude of the signal input to the nonlinear circuit element varies as $I_c(1 + m_a$

$\cos \omega_m t$); the d-c component of the resulting current undergoes similar variations as a function of time. Hence, a nonlinear impedance can be used to demodulate an amplitude modulated wave because A_0 varies at the modulation frequency ω_m.

An amplitude modulated wave contains three components: the carrier frequency ω_c, and the sum and difference frequencies $(\omega_c + \omega_m)$ and $(\omega_c - \omega_m)$. This is the input to the detector; the current through the detector will then contain the following components:

(1) The original frequencies

$$\omega_c; \qquad (\omega_c + \omega_m); \qquad (\omega_c - \omega_m)$$

(2) The second harmonics

$$2\omega_c; \qquad 2(\omega_c + \omega_m); \qquad 2(\omega_c - \omega_m)$$

(3) The sum and difference frequencies

$$2\omega_c; \qquad 2\omega_m; \qquad \omega_m; \qquad (2\omega_c + \omega_m); \qquad (2\omega_c - \omega_m)$$

Of course, out of this whole array only the ω_m term is desired, because it is the original frequency component representing the intelligence signal. Presumably the other terms can be filtered out, and this is generally possible with all terms except the second harmonic $2\omega_m$ of the desired signal. This results in second harmonic distortion, and it can be shown[2] that the second harmonic has a relative amplitude of $m_a/4$ when the detector has a resistance load. If $m_a = 1$, then 25% second harmonic distortion results. This is excessive for most applications and is one reason for the fact that square law detection of the type described here is not widely used in practical systems.

11.18. Applications in Gain Control

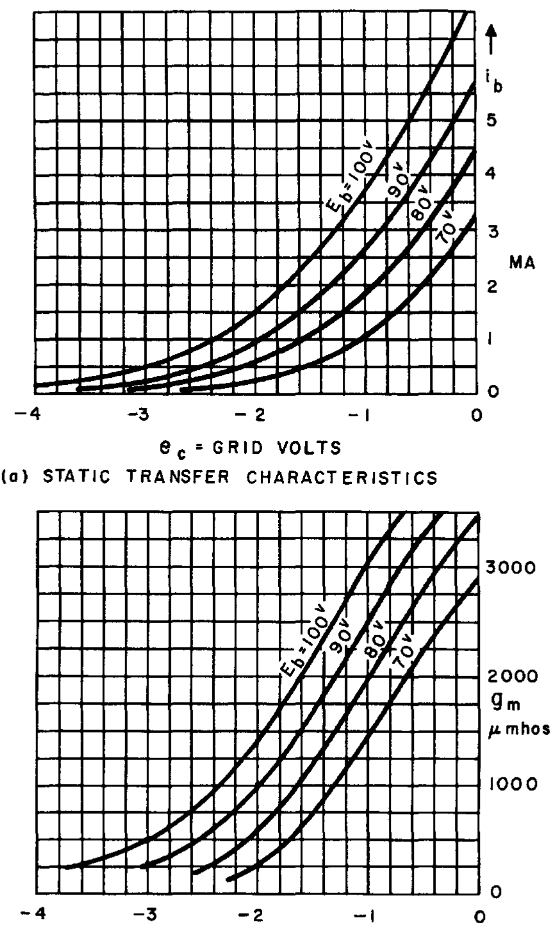
Tubes that are designed to exhibit a wide range of variability in transconductance g_m are called *variable mu, remote cutoff,* or *super control* tubes. The essential characteristics of such tubes are illustrated in figure (11.30).

Super control pentodes are extensively used in superheterodyne radio receivers as bandpass amplifier tubes. In such service they are either single or double tuned voltage amplifiers, usually of the grounded cathode type. The super control feature is useful because it permits the stage gain, $A = -g_m Z_m$, to be varied by the simple process of varying the bias on the tube. This is clear from the relationships in figure (11.30).

[2] Cruft Lab Staff, *Electronic Circuits and Tubes*, McGraw-Hill Book Co., Inc., New York, 1947, pp. 699–702.

This would have little advantage if the adjustment were made manually; the great advantage derives from the possibility of automatic control.

Automatic volume control (AVC) is obtained by using a detector to



(a) STATIC TRANSFER CHARACTERISTICS

(b) VARIATIONS IN g_m WITH GRID BIAS

Fig. 11.30. Characteristics of a super control tube.

develop a voltage proportional to the amplitude of the signal amplified by the super control tube. The detector voltage is then used to bias the super control tube so that the gain of the stage varies inversely with the strength of the signal. This acts to keep the amplifier output fairly constant and independent of variations in signal input or tube deterioration with time.

The same technique is used in audio recording systems. In this application it is called *automatic volume expansion* and *compression* and is used to alter and restore the true dynamic range of the recorded sound.

Another application for super control tubes occurs in the design of *reactance tube modulators*, which are used at times in frequency modulated transmitters (see chapters 13 and 15). This particular type of circuit has an input reactance proportional to the g_m of the tube. By varying the g_m, the effective reactance varies, and this is a part of an oscillator tank circuit. As a result, the oscillator frequency can be made to vary in proportion to the g_m of the tube. If the intelligence signal is used to vary the grid bias of the tube, the g_m varies, and the oscillator is frequency modulated.

PROBLEMS

11.1. Using the published plate characteristics for a triode connected 6F6 tube, compute the fundamental power output and per cent second and third harmonic distortion with the following load resistances: 2000, 3000, 4000, 5000, 6000, 7000, 8000, and 9000 ohms. Locate the Q point, using $E_b = 275$ v and $E_{cc} = -20$ v. The grid signal amplitude is 20 v. Neglect the dynamic shift of the Q point.

11.2. The power amplifier stage for a small a-c/d-c radio receiver is to be designed. The power output must not be less than 1.2 w with not more than 10% harmonic distortion. All filaments of the tubes in the receiver are connected in series across the 115-v line. The other tubes in the receiver are 12SA7 = mixer and local oscillator; 12SK7 = intermediate frequency amplifier (1 stage); 12SQ7 = detector and audio voltage amplifier; 35Z4–GT = rectifier. Select a suitable power amplifier tube and explain your selection in detail.

11.3. In problem (11.2) it is obvious that the current requirements for all the tube filaments must be the same. Bearing in mind that the series 12 tubes require 12.6 v for the filament supply, compute the resistance and wattage rating of any series dropping resistor required in the filament circuit.

11.4. If the output voltage from the 12SQ7 detector unit is 0.25 v, compute the voltage gain required in the audio voltage amplifier to drive the power amplifier at its rated power output. Compute the correct load resistance for the 12SQ7 voltage amplifier section if $g_m = 1100\ \mu$mhos, $r_p = 91{,}000$ ohms, and if the gridleak resistor of the power amplifier is 500,000 ohms.

11.5. Draw the circuit diagram of the power amplifier stage showing transformer coupling between the tube and loudspeaker. The impedance of the loudspeaker voice coil is 3.2 ohms. Compute the proper turns ratio for the transformer. Compute the minimum possible efficiency of power transfer for the transformer, assuming that 1.0 w of audio power must be available at the speaker input terminals.

11.6. From the data computed in problem (11.1), plot the second and third harmonics and power output *vs.* the load resistance. Select the value of load resistance that will keep the total harmonic distortion to 5% or less. What power output is obtained?

11.7. Construct the dynamic transfer characteristic of the 6F6, using the load resistance computed in (11.6).

11.8. Discuss the problem of push-pull amplification with shunt feed, rather than transformer coupling. In what particular ways does the end result differ from transformer coupling?

11.9. Draw the circuit diagram and explain the operation of a push-pull transistor amplifier, using grounded base and grounded emitter stages together in the same circuit. What are some precautions that must be observed here?

Part III

OPERATION IN THE SWITCHING MODE

Chapter 12

VACUUM TUBE POWER AMPLIFIERS IN THE SWITCHING MODE

This is the first of a sequence of chapters devoted to the operation of electronic devices in the *switching mode*. This chapter is a general exploration of the factors involved in the operation of vacuum tubes in the switching mode as power amplifiers. The discussion is purely hypothetical at the outset, but concludes with a practical design procedure for power amplifiers.

The approach is analytical, rather than graphical, for consistency with the rest of the text. The actual technique depends upon a Fourier analysis of the plate current waveform produced by a sinusoidal grid signal of sufficient amplitude to cause the switch in the equivalent circuit to operate.

This chapter is important in the design of audio power amplifiers and high-efficiency tuned amplifiers for unmodulated signals. It is also a prerequisite for chapter 13, which covers the design of similar amplifiers used to amplify modulated signals.

The theory of transistor amplifiers operating in the switching mode was not fully developed at the time of manuscript preparation and it is not included for that reason.

1.21. Factors of Interest

The first step in the design of an amplifier for use in a given application is the selection of the tube. This is a complicated matter because it depends upon a multitude of factors too numerous to mention. A wide latitude of *designer's choice* is nearly always present and the application of sound engineering judgment tempered by experience is required.

Once the tube is selected, the design of the circuit can be approached. In the derivation of design equations you should always keep the obviously important factors in mind. For example, it is reasonably clear that the actual signal power output will be of considerable interest. Also, as in any power application, the amount of wasted power is important. Hence the efficiency of conversion of d-c power input to signal power output is a significant factor. This is the *plate circuit efficiency* of the amplifier and is defined as

$$\eta_p = \frac{\text{signal power output}}{\text{d-c power input}} = \frac{P_{ac}}{P_{dc}}$$

The power amplifiers covered in this chapter operate in the switching mode. Consequently, the *conduction time*, or *conduction angle* when sinusoidal signals are involved, is a major factor of interest. This requires an exact knowledge of the grid bias and grid signal voltage necessary to produce this conduction angle. It may also develop that grid current flows and that grid driving power is required.

Because the tube operates in the switching mode, a sinusoidal grid signal will produce a highly distorted plate current signal. Thus a large number of harmonics of varying amplitudes will appear in the output. Clearly, the relative magnitudes and frequencies of these terms will be important.

The frequency of the grid signal may vary over some specified range. Thus the frequency response of the amplifier must be considered. This is an important factor in any case because of the presence of harmonics.

The foregoing represent only a few of the factors that must be considered in designing an amplifier. By way of summary, these factors can be listed as follows:

(1) Signal power output.
(2) Harmonics in the output.
(3) Conduction angle of the tube.
(4) Plate circuit efficiency.
(5) Grid bias.
(6) Grid signal voltage and driving power.
(7) Frequency response.

12.2. Operation in the Switching Mode

If the static plate characteristics of vacuum tubes are equally spaced, parallel, straight lines as shown in figure (12.1a), the behavior of the tube can be accurately represented over this region by the equivalent circuit of figure (12.1b). In most cases this equivalent circuit provides a close approximation to the actual behavior of the tube. Moreover, tubes that are designed for use as power amplifiers are made to approach this hypothetical linear status as nearly as is consistent with other requirements. Nevertheless, some nonlinearity, lack of parallelism, and nonuniformity in spacing will always exist. Theoretical calculations based upon the linear idealization will be somewhat at variance with experiment. The most noticeable deviation occurs in the positive grid region as shown in figure (12.1a). For very positive values of grid voltage approaching the plate voltage, the plate current

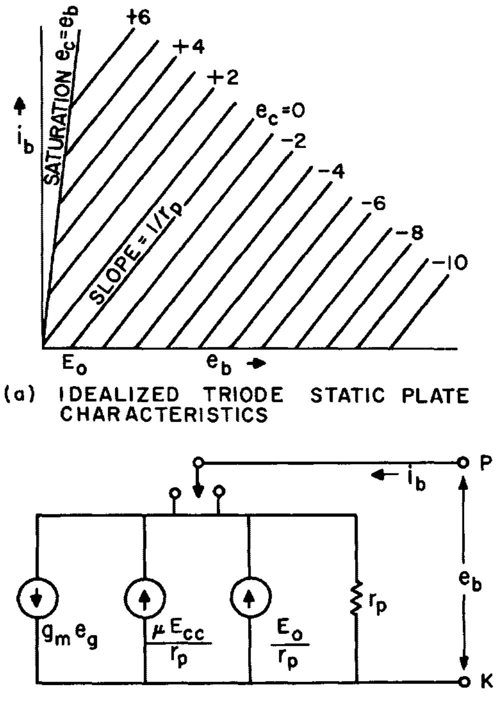


Fig. 12.1. Representation of an idealized triode.

becomes virtually independent of the grid voltage. This is the *saturation* region of the tube characteristics, and establishes the upper limit on the grid voltage for which the linear equivalent circuit is valid.

The static *transfer* characteristics of the idealized triode are shown in figure (12.2a). Suppose that the plate voltage and the Q point are determined so that operation is confined to one of the transfer characteristics as shown in figure (12.2b). By proper adjustment of the grid bias voltage E_{cc} and the grid signal e_g, the plate current will have a signal component that is an undistorted replica of the grid signal voltage. The tube operates as shown in figure (12.2b). This is clearly class A operation and this mode was treated in detail in part II.

Now consider figure (12.2c). Here the grid bias and grid signal voltage are both increased over the values they had in class A operation.

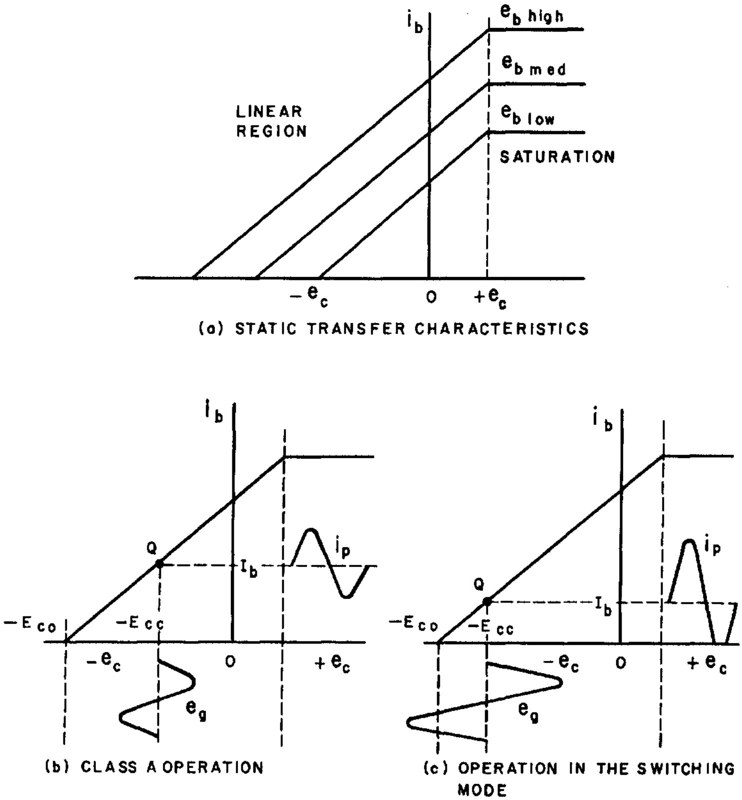


Fig. 12.2. Vacuum tube operation expressed in terms of the transfer characteristics.

It is apparent that the signal component of plate current has been distorted to a remarkable degree because of nonconduction of the tube over some fraction of the period of the grid signal voltage. During this nonconducting interval the plate current is zero, and this situation prevails as long as the total grid voltage is more negative than the cutoff voltage E_{co}. Hence the tube is being operated in the switching mode.

It has become common practice to associate certain names with various conduction intervals of the tube. These are summarized as follows:

If the tube conduction angle is	*Operation is said to be*
$360°$	class A
between $180°$ and $360°$	class AB
$180°$	class B
less than $180°$	class C

In addition, subscripts are used to indicate the presence or absence of grid current. That is, subscript 1 = no grid current, subscript 2 = grid current flows.

Thus a typical amplifier designation might be given as B_2. This signifies that the tube conducts plate current over a $180°$ interval and that grid current flows at some time during the conducting period. Subscripts are often omitted on the designations of class A and class C amplifiers. Usage has established that this means class A_1 and class C_2.

The operation of the switch in the equivalent circuit makes it clear that the circuits in Part III must be analyzed somewhat differently than class A circuits. However, it will be apparent later that the difference is mainly one of viewpoint and emphasis and that a great deal of the theoretical technique developed for class A circuits will be applicable here.

12.3. Preliminary Steps in the Analysis

It was shown in the preceding section that operation of a vacuum tube in the switching mode causes a considerable distortion in the signal component of plate current. The sharp discontinuity in the transfer characteristic of the tube is clearly a form of nonlinearity, so it should be anticipated that the effects caused by this discontinuity will be similar to those computed for the nonlinear class A circuits of chapter 11. Thus operation of a vacuum tube in the switching mode is expected to produce a d-c term, a fundamental, and an infinite series of harmonics in the output. Also, because the degree of distortion of the waveform depends upon the interval of nonconduction of the tube, it is reasonable to conclude that the amplitudes of these various components in the output will have some functional relationship to the nonconducting interval of time. The purpose of this section is to determine this relationship.

A representative case involving operation in the switching mode was shown in figure (12.2c). Assuming a sinusoidal grid signal, the plate current waveform appears as shown in figure (12.3a). The complete equivalent circuit of the tube during the conducting interval is shown in figure (12.3b), where the current source equivalent for the tube is used.

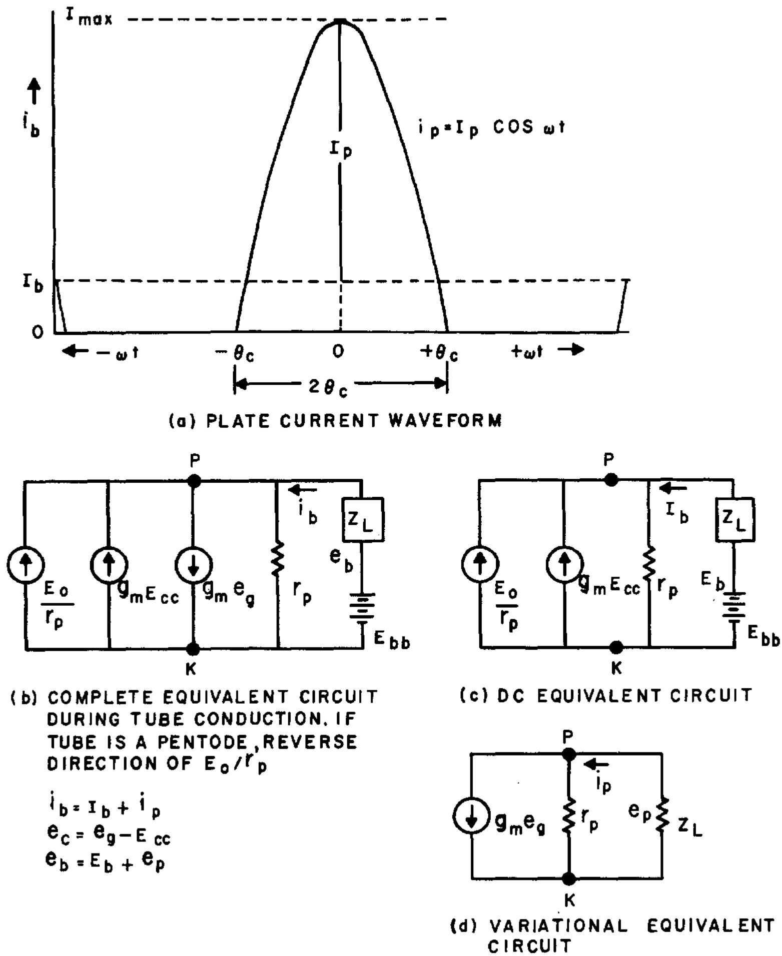


Fig. 12.3. Waveform and equivalent circuits during tube conduction.

The Principle of Superposition permits simplification of this general circuit into two specialized cases as follows:

(1) Figure (12.3c) shows only the d-c components in the circuit.

(2) Figure (12.3d) shows only the variational components.

Reference to the d-c, or static, equivalent circuit shows that the direct plate current in the circuit is

$$I_b = -g_m E_{cc} + \frac{E_b}{r_p} - \frac{E_0}{r_p}$$

or

$$I_b = -g_m\left(E_{cc} - \frac{E_b - E_0}{\mu}\right) \tag{12.1}$$

where $E_{cc} =$ grid bias voltage; $E_b =$ plate-to-cathode voltage; $E_0 =$ intercept voltage. Let

$$E_b' = E_b - E_0 \tag{12.2}$$

so that the equation for the plate current is

$$I_b = -g_m\left(E_{cc} - \frac{E_b'}{\mu}\right) \tag{12.3}$$

In a similar manner, from the variational equivalent circuit,

$$i_p = g_m e_g + \frac{e_p}{r_p}$$

or

$$i_p = g_m\left(e_g + \frac{e_p}{\mu}\right) \tag{12.4}$$

From the relationships given in (12.3) and (12.4) it is clear that each plate current component is produced by a *composite voltage.* Thus the signal component of plate current i_p is produced by the composite voltage $(e_g + e_p/\mu)$. Now, because the signal current is

$$i_p = I_p \cos\theta = g_m\left(e_g + \frac{e_p}{\mu}\right); \qquad \pm\theta_c < \pm\theta < 0 \tag{12.5}$$

$$i_p = 0; \qquad \pm\pi < \pm\theta < \pm\theta_c$$

the angular variation of the composite voltage must be of the same form. That is,

$$i_p = I_p \cos\theta = g_m\left(E_g + \frac{E_p}{\mu}\right)\cos\theta$$

during the conducting period. If the magnitude of the composite voltage is defined as

$$E_1 = E_g + \frac{E_p}{\mu} \tag{12.6}$$

the signal component of plate current is

$$i_p = g_m E_1 \cos\theta \tag{12.7}$$

where $g_m E_1 = I_p$ during the conducting interval $2\theta_c$.

The d-c component can be expressed in terms of a composite voltage as well as in terms of I_p and θ_c. That is,

$$I_b = -g_m\left(E_{cc} - \frac{E_b'}{\mu}\right) = -g_m E_x \tag{12.8}$$

where

$$E_{cc} - \frac{{}^{\prime}_{b}}{\mu} = E_x \tag{12.9}$$

Also, from figure (12.3a) it is clear that

$$I_b = -I_p \cos\theta_c \tag{12.10}$$

Therefore

$$I_b = -g_m E_1 \cos\theta_c \tag{12.11}$$

and so

$$E_x = E_1 \cos\theta_c \tag{12.12}$$

Combining the preceding results, the total plate current can be expressed as:

$$\begin{aligned} i_b(\theta) &= I_p(\cos\theta - \cos\theta_c) \\ &= g_m E_1(\cos\theta - \cos\theta_c); \qquad \pm\theta_c < \pm\theta < 0 \qquad (12.13) \\ i_b(\theta) &= 0; \qquad \pm\pi < \pm\theta < \pm\theta_c \qquad (12.14) \end{aligned}$$

The results of operating a vacuum tube in the switching mode are computed in the next section using equations (12.13) and (12.14).

12.4. Harmonic Generation vs. Conduction Angle

In the waveform of figure (12.3a) the tube conducts over a total angle of $2\theta_c$. Hence $2\theta_c$ is called the *total conduction angle* of the tube. Also, by choosing $\theta = 0$ as shown in the figure, the waveform becomes symmetrical about the axis. That is, $i_b(\theta) = i_b(-\theta)$. This is the definition of an *even function.*

An even function can be expanded into an infinite series of cosine terms through application of *Fourier Analysis.* That is, the plate current waveform $i_b(\theta)$ can be written

$$i_b(\theta) = I_b' + I_{p_1}\cos\theta + I_{p_2}\cos 2\theta + I_{p_3}\cos 3\theta + \ldots$$

or

$$i_b(\theta) = I_b' + \sum_{n=1}^{\infty} I_{p_n}\cos n\theta$$

where the coefficients in the series are given by the following formulas:[1]

$$I_b' = \text{d-c component} = \frac{1}{\pi}\int_0^{\pi} i_b(\theta)\, d\theta \tag{12.15}$$

$$I_{p_1} = \text{fundamental} \quad = \frac{2}{\pi}\int_0^{\pi} i_b(\theta)\cos\theta\, d\theta \tag{12.16}$$

$$I_{p_n} = \text{harmonics} \quad = \frac{2}{\pi}\int_0^{\pi} i_b(\theta)\cos n\theta\, d\theta \tag{12.17}$$

The amplitudes of these various components are easily computed from these formulas as soon as $i_b(\theta)$ is specified. This was calculated in the preceding section and the results given in equations (12.13) and (12.14). Substitution of these relationships into the preceding formulas leads to the following relationships:

$$I_b' = \frac{g_m E_1}{\pi}\int_0^{\theta_c}(\cos\theta - \cos\theta_c)\, d\theta \tag{12.18}$$

$$I_{p_1} = \frac{2g_m E_1}{\pi}\int_0^{\theta_c}(\cos\theta - \cos\theta_c)\cos\theta\, d\theta \tag{12.19}$$

$$I_{p_n} = \frac{2g_m E_1}{\pi}\int_0^{\theta_c}(\cos\theta - \cos\theta_c)\cos n\theta\, d\theta \tag{12.20}$$

These integrals are easily evaluated from integral tables.[2] The results are

$$I_b' = \frac{g_m E_1}{\pi}(\sin\theta_c - \theta_c\cos\theta_c) \tag{12.21}$$

$$I_{p_1} = \frac{2g_m E_1}{\pi}\left(\frac{\theta_c}{2} + \frac{\sin 2\theta_c}{4} - \cos\theta_c\sin\theta_c\right) \tag{12.22}$$

$$I_{p_n} = \frac{2g_m E_1}{\pi}\left[\frac{\sin(n-1)\theta_c}{2(n-1)} + \frac{\sin(n+1)\theta_c}{2(n+1)} - \frac{\cos\theta_c\sin n\theta_c}{n}\right] \tag{12.23}$$

[1] See for example, Philip Franklin, *Methods of Advanced Calculus*, McGraw-Hill Book Co., Inc., New York, 1944, pp. 405–406.

[2] See for example, Burington, *Handbook of Mathematical Tables and Formulas*, Handbook Publishers Inc., Sandusky, Ohio, 1940, pp. 73–74, integrals 200–202, 204.

Through the use of standard trigonometric identities, a few of these coefficients can be simplified and written

$$I_b' = \frac{g_m E_1}{\pi}(\sin\theta_c - \theta_c\cos\theta_c) \tag{12.24}$$

$$I_{p_1} = \frac{g_m E_1}{\pi}(\theta_c - \sin\theta_c\cos\theta_c) \tag{12.25}$$

$$I_{p_2} = \frac{g_m E_1}{\pi}\frac{2}{3}\sin^3\theta_c \tag{12.26}$$

$$I_{p_3} = \frac{g_m E_1}{\pi}\frac{2}{3}\sin^3\theta_c\cos\theta_c \tag{12.27}$$

and so on.

It is also desirable to express these coefficients in terms of the maximum plate current. Because

$$I_{\max} = I_p(1 - \cos\theta_c) \tag{12.28}$$

and

$$I_p = g_m E_1$$

then

$$g_m E_1 = \frac{I_{max}}{1 - \cos\theta_c} \tag{12.29}$$

Thus if the $g_m E_1$ in the preceding equations is replaced by the value given in equation (12.29), another useful set of formulas results.

Now suppose that the maximum plate current is held constant while the conduction angle of the tube is made variable. Under these conditions it is a simple matter to compute the various Fourier components, using formulas (12.24) through (12.27) and with $g_m E_1$ replaced by equation (12.29). The results of this calculation are shown in figure (12.4), which shows the relative magnitudes of the d-c term, fundamental, and second and third harmonics as functions of the total conduction angle. All magnitudes are referred to the magnitude of the fundamental in class A operation. Bear in mind that harmonic distortion caused by curvature in the characteristics has been neglected.

Figure (12.4) shows that for a fixed plate current swing, the fundamental amplitude reaches a maximum value that is about 8% larger than that obtained in class A. This value results in class AB operation. For class B operation the fundamental is again equal to the value obtained in class A. Thereafter, as operation moves into the class C region, the fundamental amplitude decreases continuously as the angle of tube conduction decreases. It is significant that the third harmonic is zero for class B operation.

The same data can be plotted in a slightly different manner by computing the harmonic percentages. That is,

$$D_2 = \frac{I_{p_2}}{I_{p_1}} \times 100\% = \text{per cent 2nd harmonic}$$

$$D_3 = \frac{I_{p_3}}{I_{p_1}} \times 100\% = \text{per cent 3rd harmonic}$$

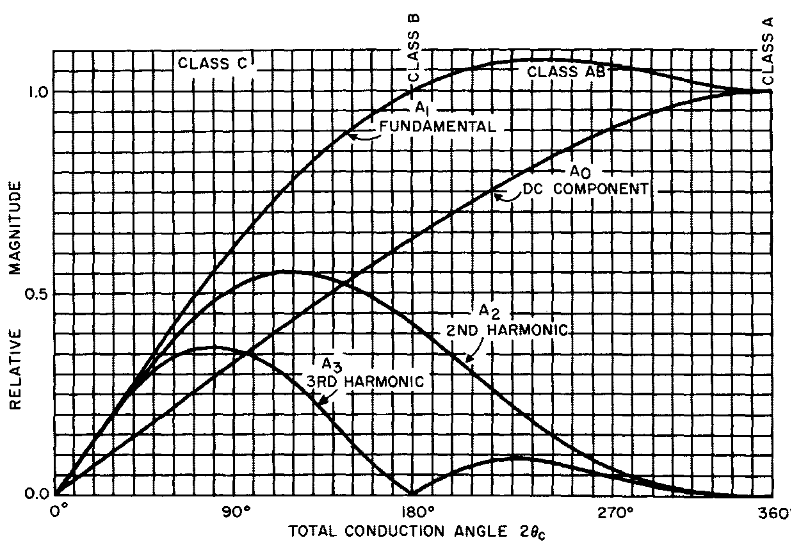


Fig. 12.4. Effect of conduction angle on the relative magnitudes of the Fourier coefficients.

These two factors are plotted against the tube conduction angle in figure (12.5). The curves show that as long as the tube conducts for an angle exceeding about 290°, the harmonic distortion introduced by plate current cutoff is not excessively large. For shorter conduction angles, the 2nd harmonic increases rapidly and continuously. The 3rd harmonic increases, reaches a maximum of about 9%, and then decreases to zero in class B operation. In class C operation, the 3rd harmonic increases very rapidly as the tube conducts for shorter periods.

The high harmonic content obtained in class C operation suggests the possibility of using such circuits for frequency multiplication.

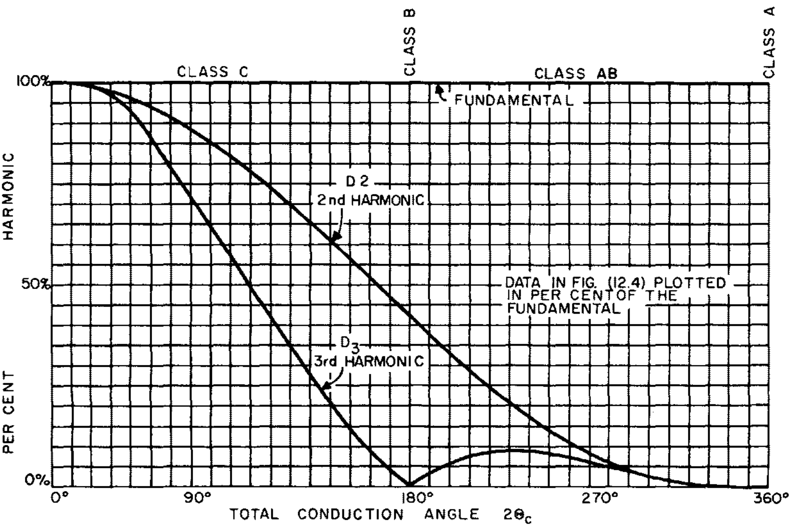


Fig. 12.5. Per cent harmonic vs. conduction angle. Data plotted in per cent of the fundamental.

12.5. Plate Circuit Efficiency

In nearly all cases in which vacuum tubes are used as amplifiers, the tube serves mainly as a *power converter*, changing the d-c power available from the plate power supply into signal power that conveys intelligence in one form or another. In the final analysis, every amplifier is a power amplifier because the signal power in the output is always considerably larger than that in the input.

The plate circuit efficiency measures the effectiveness of the tube in converting DC power into signal power. That is,

$$\eta_p = \text{plate circuit efficiency} = \frac{P_{ac}}{P_{dc}}$$

where P_{ac} = signal power output; P_{dc} = d-c power input from the plate supply.

The problems relating to the plate circuit efficiency are relatively complex. Some of the factors involved were discussed in connection with class A power amplifiers in chapter 11. It was shown there that

$$P_{dc} = P_{ac} + P_L + P_p$$

where P_L = d-c power loss in the load resistance; P_p = plate dissipation. Hence the plate circuit efficiency can be written

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_L + P_p}$$

It was then shown that the maximum theoretical efficiency of the class A series fed amplifier was 25%. By using shunt feed or transformer coupling, the d-c power loss in the load resistance could be reduced practically to zero and the plate circuit efficiency written

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_p}$$

In this case the theoretical maximum efficiency was shown to be 50%. Further increases in efficiency are possible only by reducing the plate dissipation P_p.

It will be shown in the next section that the plate dissipation can be reduced by operating vacuum tubes in the switching mode. Thus such operation leads to higher plate circuit efficiencies than those obtainable in class A. Indeed, this is the main reason for amplifier operation in class AB, B, and C, because the large amount of harmonic distortion produced is generally a severe disadvantage.

12.6. Effect of Conduction Angle on Efficiency

In the fairly general case of a shunt fed or transformer coupled amplifier of arbitrary conduction angle $2\theta_c$ the plate circuit efficiency is

$$\eta_p = \frac{P_{ac}}{P_{dc}}$$

The term *signal power* will generally refer to the power developed by the fundamental component of plate current. Hence

$$P_{ac} = \tfrac{1}{2} I_{p_1}^2 R_L \quad \text{and} \quad P_{dc} = E_{bb} I_b'$$

Therefore the plate circuit efficiency is

$$\eta_p = \frac{1}{2} \frac{I_{p_1}}{I_b'} \cdot \frac{I_{p_1} R_L}{E_{bb}} \tag{12.30}$$

It was previously shown in section (12.4) that the d-c component and fundamental could be expressed in terms of the total plate current swing and conduction angle as follows:

$$I_{p_1} = \frac{I_{max}}{\pi} \cdot \frac{\theta_c - \sin\theta_c \cos\theta_c}{1 - \cos\theta_c}$$

$$I_b' = \frac{I_{max}}{\pi} \cdot \frac{\sin\theta_c - \theta_c \cos\theta_c}{1 - \cos\theta_c}$$

Substitute these two expressions into the equation for the plate circuit efficiency. The result, expressed in terms of per cent, is

$$\%\eta_p = 50\,\frac{I_{max}}{\pi}\cdot\frac{R_L}{E_{bb}}\cdot\frac{(\theta_c - \sin\theta_c\cos\theta_c)^2}{(1-\cos\theta_c)(\sin\theta_c - \theta_c\cos\theta_c)} \tag{12.31}$$

$$= 15.9\,\frac{I_{max}R_L}{E_{bb}}\cdot\frac{(\theta_c - \sin\theta_c\cos\theta_c)^2}{(1-\cos\theta_c)(\sin\theta_c - \theta_c\cos\theta_c)} \tag{12.32}$$

This is the plate circuit efficiency for any conduction angle, load resistance, plate supply voltage, and maximum plate current swing.

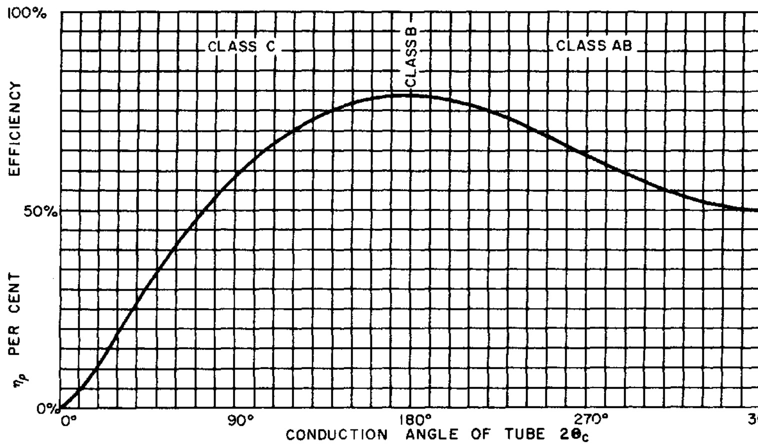


Fig. 12.6. Plate circuit efficiency vs. conduction angle, resistance load.

Now suppose that the plate current swing is adjusted to the maximum possible theoretical value. In this case the plate voltage swings from 0 to $2E_{bb}$, centered about E_{bb}. Hence

$$I_{\max}R_L = 2E_{bb}$$

and the maximum possible theoretical efficiency is

$$\%\eta_p = 31.8\,\frac{(\theta_c - \sin\theta_c\cos\theta_c)^2}{(1-\cos\theta_c)(\sin\theta_c - \theta_c\cos\theta_c)} \quad \text{per cent} \tag{12.33}$$

This is often called the *asymptotic efficiency* of a power amplifier with a pure resistance load circuit where the load resistance is independent of frequency.

Equation (12.33) has been computed and plotted and the results are shown in figure (12.6) as a function of the tube conduction angle. The results are striking. The graph indicates that the plate circuit efficiency can be increased from 50% up to a theoretical maximum value of 78.5%

by operating the tube in class B. From the standpoint of efficiency, no advantage results from class C operation. Moreover, the harmonics generated in class C are more pronounced than in class B so that class C operation of amplifiers with resistance loads, which are independent of frequency, should usually be avoided.

Consideration of figure (12.6) indicates that power amplifiers having load impedances that are constant over the pass band that includes all the significant harmonics attain their maximum efficiency in class B. Audio power amplifiers fit this category, and for this reason, are never operated in class C except in trick circuit arrangements.

12.7. Low Pass Power Amplifiers

Class A power amplifiers were discussed in some detail in chapter 11. It was shown there that the push-pull connection can be used to advantage to reduce the even harmonic distortion to small values. The odd harmonics were not appreciably affected by the connection.

Low pass amplifiers have load impedances that are essentially constant resistances over a specified range of frequencies up to some given high frequency limit. Audio amplifiers typify such circuits. Thus such amplifiers have the operating characteristics shown in figures (12.4) and (12.6). These figures show that these amplifiers can be used at higher efficiencies than class A circuits, but that the harmonic distortion generally increases. Consequently, class AB and B operation of a single tube audio power amplifier is seldom used, because the harmonic distortion introduced is excessive. Of course, some tube nonconduction over short angles might be tolerated because the distortion is not high in such cases. Neither is the efficiency much improved over class A operation. Thus, as a general rule, low pass, single stage power amplifiers are most advantageously operated in class A.

Push-pull amplifiers can be used to considerable advantage when the tubes operate in the switching mode, because the circuit connection can be used to reduce the even harmonic distortion, and the switching mode can be used to increase the efficiency. Thus two tubes operated in push-pull class AB or B will provide a greater power output than the same tubes operated in class A push-pull or parallel. The advantage of the push-pull connection for class B operation is clear from figure (12.4) or (12.5). Figure (12.5) shows that the third harmonic is zero in class B. The second harmonic has a value of about 42.5%, but this can be reduced to a small figure by the push-pull connection. Moreover, the circuit now has a maximum theoretical efficiency of 78.5%, a considerable improvement over the 50% possible in class A operation. Furthermore, practical push-pull class B amplifiers come closer to 78.5% than class A amplifiers come to 50% for the same amount of harmonic distortion.

The importance of this efficiency improvement will be apparent from the following comparison of hypothetical cases. It was previously shown that

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_p}$$

Solve this equation for P_{ac}. The result is

$$P_{ac} = \frac{\eta_p}{1 - \eta_p} P_p$$

For a class A amplifier operating at a hypothetical maximum efficiency of 50%, $P_{ac} = P_p$. Thus the maximum theoretical power output is equal to the plate dissipation of the tube. If the same tube is operated in class B with a theoretical maximum efficiency of 78.5%, the signal power output is $P_{ac} = 3.65P_p$. The signal power output in this case is 3.65 times as large as that obtainable in class A operation.

These figures also make it clear that class C operation is inadvisable even with the push-pull connection. The second harmonic could be largely eliminated, of course, but the third harmonic is large and is not removed by the push-pull circuit. Also, the increase in distortion is accompanied by a decrease in efficiency.

12.8. Efficiency with a Tuned Load Circuit[3]

For the case of a pure resistance load circuit, section (12.6) indicated that no advantage resulted from class C operation. However, the situation is radically altered when the load on the tube is changed to an impedance having a band pass characteristic. The simple high Q tuned circuit is the most common example of such a circuit.

Assume the amplifier to be shunt fed and to have a load circuit tuned to the fundamental frequency. Further assume that the Q of the circuit is so high that its impedance is zero at zero frequency and at all the

[3] This discussion of tuned amplifiers follows that given by W. L. Everitt in: *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., New York, 1937, chap. 17; "Optimum Operating Conditions for Class C Amplifiers," *Proc. IRE*, vol. 22, p. 152, February, 1934; and "Optimum Operating Conditions for Class B Amplifiers," *Proc. IRE*, vol. 24, p. 305, February, 1936.

harmonic frequencies. Thus the load short circuits all plate current components to ground with the exception of the fundamental. The impedance presented to the fundamental is

$$R_L = \text{input impedance at antiresonance}$$

Thus only the fundamental appears in the output and only the fundamental component of current I_{p_1} causes any voltage drop in the load circuit. This contrasts sharply with the case of a resistance load in which *all* current components contributed voltage drops in the load circuit and variations in the plate voltage.

Under these assumed conditions, the plate voltage is

$$E_p = E_{p_1} = -I_{p_1} R_L \tag{12.34}$$

and so

$$E_1 = E_g + \frac{E_p}{\mu} = E_g - \frac{I_{p_1} R_L}{\mu} \tag{12.35}$$

It was previously shown that

$$I_{p_1} = \frac{g_m E_1}{\pi} (\theta_c - \sin\theta_c \cos\theta_c) \tag{12.36}$$

If the expression given for E_1 in equation (12.35) is substituted into equation (12.36), the result is

$$I_{p_1} = \frac{\mu E_g}{R_L + \beta r_p} \tag{12.37}$$

where

$$\beta = \frac{\pi}{\theta_c - \sin\theta_c \cos\theta_c} \tag{12.38}$$

The plate circuit efficiency was given in equation (12.30) and the values of I_b' and I_{p_1} in the same place. Thus

$$I_b' = \frac{g_m E_1}{\pi} (\sin\theta_c - \theta_c \cos\theta_c) \tag{12.39}$$

Thus

$$\frac{I_{p_1}}{I_b'} = \frac{\theta_c - \sin\theta_c \cos\theta_c}{\sin\theta_c - \theta_c \cos\theta_c} \tag{12.40}$$

Now substitute this ratio and the value of I_{p_1} given in equation (12.37) into the efficiency equation of (12.30). The result is

$$\%\eta_p = \frac{\mu E_g}{E_{bb}} \frac{R_L}{R_L + \beta r_p} \cdot \frac{\theta_c - \sin\theta_c \cos\theta_c}{\sin\theta_c - \theta_c \cos\theta_c} 50 \tag{12.41}$$

It will be noted that this equation is quite different from (12.31), which was obtained for the case of a resistance load.

That part of equation (12.41) that depends upon θ_c has been computed and plotted in figure (12.7). This quantity is called the *asymptotic efficiency* η_A of a tuned amplifier operating in the switching mode. That is,

$$\%\eta_A = \frac{\theta_c - \sin\theta_c \cos\theta_c}{\sin\theta_c - \theta_c \cos\theta_c} 50 \tag{12.42}$$

This factor approaches 100% for class C operation as shown by figure (12.7). It is evident that an actual amplifier is prevented from attaining 100% efficiency by the factor

$$\frac{\mu E_g}{E_{bb}} \frac{R_L}{R_L + \beta r_p}$$

It is also clear from equation (12.41) that the efficiency is a linear function of the grid voltage E_g as long as the tube conduction angle is

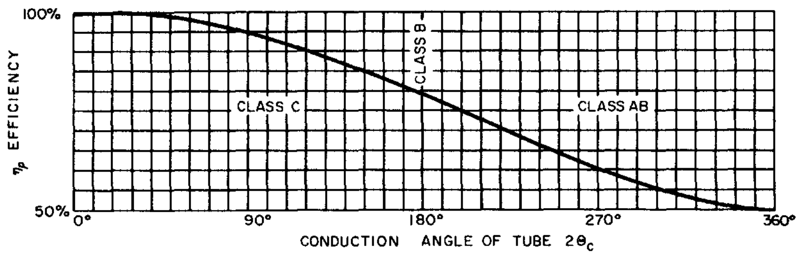


Fig. 12.7. Plate circuit efficiency vs. conduction angle, tuned load.

constant. This is an important factor in the design of a modulated class B amplifier and will be treated later. Of course, this linearity only holds up to the saturation region of the tube characteristics. This is not a serious limitation upon the use of the equation, however, because operation into the saturation region is usually avoided; it causes a flattening in the top of the plate current waveform.

It seems reasonably clear that *optimum* operation will be obtained when E_g has its maximum value and that this maximum value will drive the tube current just up to the saturation point. This analysis is undertaken in the next two sections.

12.9. Tuned Amplifiers at the Saturation Point

Just before operation at the saturation point occurs, the following terminology applies:

E_{cm} = maximum total grid voltage
E_{gm} = maximum grid signal amplitude
E_{bm} = minimum total plate voltage
I_{pm_1} = maximum fundamental component of plate current

If the load circuit is antiresonant at the fundamental frequency, the minimum plate voltage E_{bm} and maximum grid voltage E_{cm} occur at the same instant of time. Thus it is possible to write

$$E_{bm} = E_{bb} - I_{pm_1}R_L; \qquad E_{cm} = E_{gm} - E_{cc}$$

Saturation occurs when the total grid voltage is just equal to the total plate voltage, or $E_{cm} = E_{bm}$, or when

$$E_{gm} + I_{pm_1}R_L = E_{bb} + E_{cc} \tag{12.43}$$

The d-c and variational terms have been separated and combined in this equation. However, it was shown in equation (12.37) that

$$I_{p_1} = \frac{\mu E_g}{R_L + \beta r_p}$$

Hence it is also true that

$$I_{pm_1} = \frac{\mu E_{gm}}{R_L + \beta r_p} \tag{12.44}$$

Thus equation (12.43) can be written

$$E_{gm}\frac{(\mu + 1)R_L + \beta r_p}{R_L + \beta r_p} = E_{bb} + E_{cc} \tag{12.45}$$

It was shown earlier in the chapter in equations (12.6), (12.9), and (12.12) that

$$E_x = E_{cc} - \frac{E_b'}{\mu} = E_1 \cos \theta_c$$

$$E_1 = E_g + \frac{E_p}{\mu} = E_g - \frac{I_p R_L}{\mu}$$

In the present case shunt feed is assumed, so that $E_b' = E_{bb}'$, where $E_{bb}' = E_{bb} + E_0$ (reverse sign on E_0 for a pentode). Moreover, because only the fundamental produces a voltage drop in R_L, it is possible to write, for the saturation condition,

$$E_x = E_{cc} - \frac{E_{bb}'}{\mu} = E_1 \cos \theta_c$$

$$E_1 = E_{gm} - \frac{I_{pm_1}R_L}{\mu}$$

Combine these two equations and write

$$\left(E_{gm} - \frac{I_{pm_1}R_L}{\mu}\right) \cos \theta_c = E_{cc} - \frac{E_{bb}'}{\mu} \tag{12.46}$$

Substitute equation (12.44) for I_{pm_1} and solve for E_{cc}. The result is

$$E_{cc} = E_{gm} \frac{\beta r_p}{R_L + \beta r_p} \cos \theta_c + \frac{E_{bb}'}{\mu} \tag{12.47}$$

Now insert this value for E_{cc} into equation (12.45). Solve the result for E_{gm}. This yields

$$E_{gm} = \frac{\mu + 1}{\mu} \cdot \frac{R_L + \beta r_p}{(\mu + 1)R_L + \beta(1 - \cos \theta_c)r_p} E_{bb}' \tag{12.48}$$

Substitute this into equation (12.44) for the fundamental component of plate current; the eventual result is

$$I_{pm_1} = (\mu + 1) \frac{E_{bb}'}{(\mu + 1)R_L + \beta(1 - \cos \theta_c)r_p} \tag{12.49}$$

In both of these last two equations, divide numerator and denominator through by r_p and then define the following factors:

$$\alpha = (\mu + 1) \frac{R_L}{r_p} = \left(\frac{\mu + 1}{\mu}\right) g_m R_L \tag{12.50}$$

$$B = \beta(1 - \cos \theta_c) = \frac{\pi(1 - \cos \theta_c)}{\theta_c - \sin \theta_c \cos \theta_c} \tag{12.51}$$

Hence the maximum values of grid excitation and fundamental plate current can be written in terms of these parameters as follows

$$E_{gm} = \frac{\alpha + (\mu + 1)\beta}{\mu(\alpha + B)} E_{bb}' \tag{12.52}$$

$$I_{pm_1} = \frac{\alpha}{\alpha + B} \cdot \frac{E_{bb}'}{R_L} \tag{12.53}$$

It is now possible to compute the power output and efficiency at saturation. That is,

$$P_{ac} = \frac{1}{2} I_{pm_1}^2 R_L = \frac{E_{bb}'^2}{2R_L} \left(\frac{\alpha}{\alpha + B}\right)^2 \tag{12.54}$$

$$= \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{\alpha}{(\alpha + B)^2} \tag{12.55}$$

The plate circuit efficiency at the saturation point is

$$\eta_{pm} = \frac{P_{ac}}{P_{dc}} = \frac{1}{2} \frac{I_{pm_1}^2 R_L}{E_{bb} I_b'}$$

$$\% \eta_{pm} = 50 \frac{I_{pm_1} R_L}{E_{bb}} \cdot \frac{I_{pm_1}}{I_b'} \quad \text{per cent}$$

Substitution of the appropriate relationships reduces this to

$$\eta_{pm} = \frac{\alpha}{\alpha + B}\left[\frac{\theta_c - \sin\theta_c\cos\theta_c}{\sin\theta_c - \theta_c\cos\theta_c}\right]\frac{1}{2} \tag{12.56}$$

where the bracketed quantity is the asymptotic efficiency plotted in figure (12.7). Hence

$$\eta_{pm} = \frac{\alpha}{\alpha + B}\eta_A \tag{12.57}$$

The plate circuit efficiency of a shunt fed amplifier can be written

$$\eta_p = \frac{P_{ac}}{P_{ac} + P_p}$$

Solve this equation for the plate dissipation.

$$P_p = P_{ac}\left(\frac{1}{\eta_p} - 1\right)$$

Thus at saturation,

$$P_{pm} = P_{ac}\left(\frac{1}{\eta_{pm}} - 1\right)$$

Substitute for P_{ac} and η_{pm} and rearrange terms. This leads to

$$P_{pm} = \text{maximum allowable plate dissipation}$$

$$= \frac{\mu + 1}{\mu}\cdot\frac{g_m E_{bb}'^2}{2}\cdot\frac{\alpha(1 - \eta_A) + B}{\eta_A(\alpha + B)^2} \tag{12.58}$$

Now define another parameter,

$$\gamma = \frac{\alpha(1 - \eta_A) + B}{\eta_A(\alpha + B)^2} \tag{12.59}$$

so that the plate dissipation can also be written

$$P_{pm} = \frac{\mu + 1}{\mu}\cdot\frac{g_m E_{bb}'^2}{2}\gamma \tag{12.60}$$

or

$$\gamma = \frac{2\mu}{\mu + 1}\cdot\frac{P_p}{g_m E_{bb}'^2} \tag{12.61}$$

From this last equation it is apparent that the constant γ can be computed in terms of the constants of the power amplifier tube.

It is possible at this point that you feel a little bewildered by the maze of mathematical detail and you may have lost sight of the objectives of the analysis. The discussion in the next section should straighten things out.

12.10. Optimum Design of a Tuned Class C Amplifier

Optimum operation of a tuned class C amplifier is secured when operation is just at the saturation point and when the plate circuit efficiency has its maximum value; optimum operation includes the allowable loss condition. Although there are a number of related aspects to the problem, the main factor of interest is the tube conduction angle required for optimum operation. Once this angle is determined, equations have been derived that allow the computation of other factors of interest.

Suppose that an amplifier is to be designed and that a given tube has been selected. The tube manufacturer will specify the maximum allowable plate dissipation P_{pm} and the μ and g_m of the tube. The intercept voltage E_0 can be determined from the static plate characteristics. A power supply voltage can be selected or the recommendations of the tube manufacturer can be followed. Thus all the necessary factors are now known, and the constant γ can be computed from equation (12.61) or (12.59).

Equation (12.59) for γ is

$$\gamma = \frac{\alpha(1 - \eta_A) + B}{\eta_A(\alpha + B)^2}$$

Now γ is a constant in this equation, while η_A and B are functions of the tube conduction angle θ_c. Hence, if this equation is solved for α, the result will be a function of θ_c only. Following this procedure produces a quadratic equation for α having a solution of the following form:

$$\alpha = \frac{1}{\eta_A}\left[\frac{1 - \eta_A}{2\gamma} - B\eta_A + \sqrt{\left(\frac{1 - \eta_A}{2\gamma}\right)^2 + \frac{B\eta_A^2}{\gamma}}\right] \tag{12.62}$$

The plus sign is required in front of the radical to make α a positive number.

If this expression for α, which is a function of θ_c only, is substituted into the equation for the plate circuit efficiency,

$$\eta_{pm} = \frac{\alpha}{\alpha + B}\eta_A$$

the result will also be a function of θ_c only. The conduction angle required for maximum efficiency can now be obtained by differentiating η_{pm} with respect to θ_c and following the usual maximizing procedure. The resulting value for θ_c can be substituted into the appropriate factors in the equation for γ, so that under these optimum conditions

$$\gamma = \frac{B + (1 - \eta_A)K}{\eta_A(B + K)^2}, \tag{12.63}$$

where

$$B = \frac{\pi(1 - \cos\theta_c)}{\theta_c - \sin\theta_c \cos\theta_c} \tag{12.64}$$

$$\eta_A = \frac{1}{2}\frac{\theta_c - \sin\theta_c - \cos\theta_c}{\sin\theta_c - \theta_c\cos\theta_c} \tag{12.65}$$

$$K = \frac{\pi(\sin\theta_c - \theta_c)}{2\sin^2\theta_c - \theta_c^2 - \theta_c\sin\theta_c\cos\theta_c} \tag{12.66}$$

For each value of θ_c, equation (12.63) will yield one value for γ. Thus a graph of θ_c vs. γ can be computed and plotted as shown in figure (12.8). Also, the values of θ_c determine B and η_A, and therefore fix α and η_{pm}. Thus these quantities can also be plotted against γ as shown in figure (12.8).

Before concluding the discussion with an outline of the design procedure for optimum operation, two additional equations must be derived. The power output is

$$P_{ac} = \frac{I_{pm_1}^2 R_L}{2}$$

or

$$I_{pm_1}R_L = \sqrt{2R_LP_{ac}} \tag{12.67}$$

However, for optimum operation at the saturation point, equation (12.46) showed that

$$E_{gm} - E_{cc} = E_{bb} - I_{pm_1}R_L$$

Solve this for $I_{pm_1}R_L$, substitute the result back into equation (12.67), and compute E_{cc} to be

$$E_{cc} = E_{gm} + \sqrt{2R_LP_{ac}} - E_{bb} \tag{12.68}$$

It was previously shown in equation (12.29) that

$$I_{\max} = g_mE_1(1 - \cos\theta_c)$$

and from equation (12.21),

$$I_b' = \frac{g_mE_1}{\pi}(\sin\theta_c - \theta_c\cos\theta_c)$$

Thus the ratio of the maximum plate current to the average plate current is

$$\frac{I_{\max}}{I_b'} = \frac{\pi(1 - \cos\theta_c)}{\sin\theta_c - \theta_c\cos\theta_c} \tag{12.69}$$

This ratio, together with η_{pm}, is plotted as a function of γ in figure (12.9).

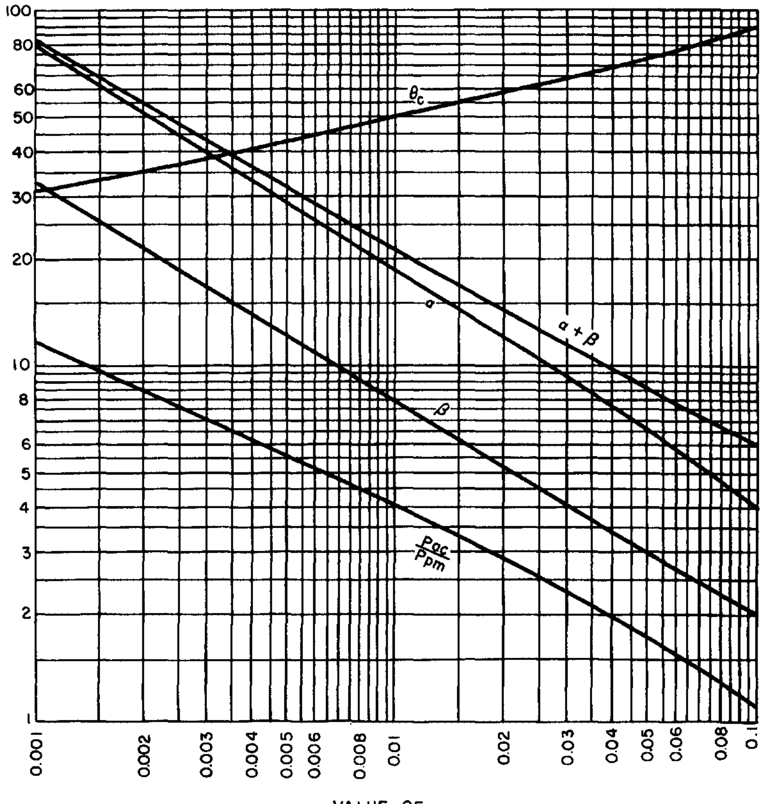


Fig. 12.8. Tuned class C amplifier optimum design curves. (From W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., 1937, p. 581.)

Basic Design Procedure Outline

(1) Select the tube and E_{bb}.

(2) Compute γ and E_{bb}'.

(3) For this value of γ, read off the corresponding values for α, β, and B from figure (12.8).

(4) Compute R_L from equation (12.50), E_{gm} from equation (12.52), and P_{ac} from equation (12.55).

(5) Compute E_{cc} from equation (12.67).

(6) For the value of γ calculated, look up η_{pm} and $I_{\max}/I_b'$ on figure (12.9).

(7) Compute I_b' from $\eta_{pm} = \dfrac{P_{ac}}{E_{bb}I_b'}$.

(8) Compute $I_{\max}$.

(9) Check to be sure that this $I_{\max}$ is within the capabilities of the tube selected. If not, repeat the design procedure for another tube or another value of E_{bb}.

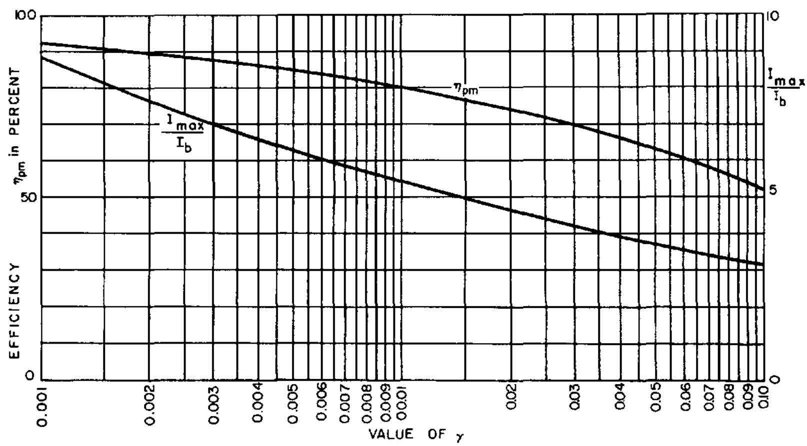


Fig. 12.9. Tuned class C amplifier optimum design curves. (From W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., 1937, p. 581.)

12.11. Neutralization

It was shown in chapter 3 that the input admittance of a vacuum tube amplifier involved both a conductance and a susceptance. The magnitude and sign of the conductance was shown to depend upon the impedance in the plate circuit. In the case of a grounded cathode amplifier this conductance becomes negative when the impedance in the plate circuit is inductive. When this happens the amplifier may burst into sustained oscillation because the feedback through C_{gp} is positive and of sufficient amplitude to create instability.

This is an acute problem in amplifiers having tuned load circuits. The effect is most frequently met with in triodes, because C_{gp} is comparatively large. It occurs occasionally with pentodes and beam power tubes where C_{gp} is much smaller. Regardless of the tube type used, whenever feedback is sufficient to cause oscillation, special circuits are added to neutralize the feedback caused by the interelectrode capacitance. Thus they are called *neutralizing circuits.*

The Rice and Hazeltine circuits are the most common methods of neutralizing amplifiers. The circuit connections are shown in figures (12.10a) and (12.10c). The diagrams have been redrawn in figures (12.10b) and (12.10d) in forms more suitable for discussion.

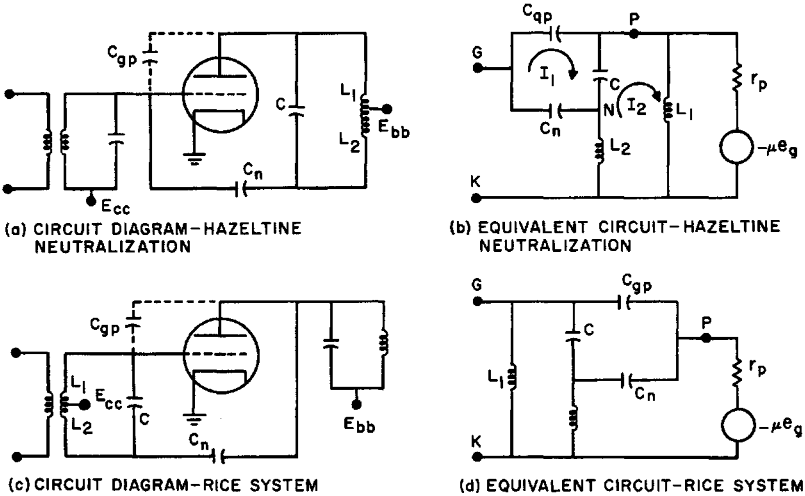


Fig. 12.10. Neutralization of single tube power amplifiers.

Consider the Hazeltine circuit. The plate power supply is connected to a tap, usually at the center, of the tank coil. The bottom end of the plate tank circuit is connected through a neutralizing capacitor C_n to the grid. The equivalent circuit given in figure (12.10b) shows that by proper adjustment of C_n, the applied voltage from the plate P to the cathode K will not produce any voltage from the grid G to the cathode. The condition for balance is easily written from this circuit as $E_{gk} = 0$ or because

$$E_{gk} = E_{C_n} + E_{L_2} = E_{C_{gp}} + E_{L_1} = 0$$

Then, for proper neutralization,

$$E_{C_n} = -E_{L_2} \quad \text{and} \quad E_{C_{gp}} = -E_{L_1}$$

If the plate tank circuit circulating current I_2 is much larger than the current I_3 the currents through L_1 and L_2 are virtually the same. This condition is generally true in high Q circuits. Therefore, in the steady state,

$$-\frac{I_1}{j\omega C_n} = I_2(j\omega L_2); \qquad \frac{I_1}{j\omega C_{gp}} = -I_2(j\omega L_1)$$

Solve these two equations simultaneously and eliminate $j\omega$ and I_1/I_2. The resulting condition required for neutralization is

$$C_n = \frac{L_1}{L_2} C_{gp}$$

Of course, if the high Q assumption is violated so that the coil

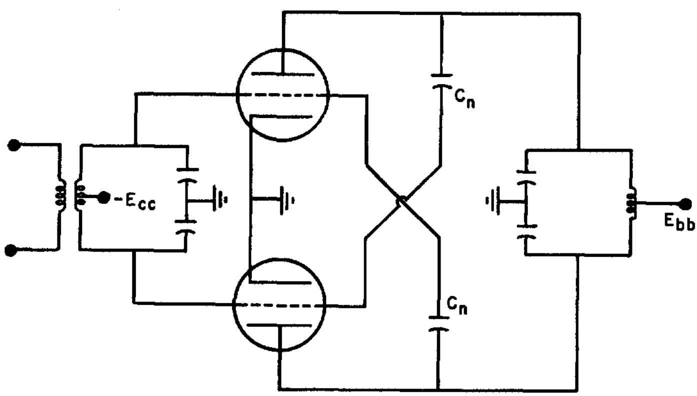


Fig. 12.11. Neutralization of a push-pull tuned power amplifier.

currents are not virtually equal, it may be impossible to neutralize the circuit by adjusting the neutralizing capacitor.

A similar analysis can be applied to the Rice circuit and it will be found that neutralization is secured when

$$L_1 = L_2 \quad \text{and} \quad C_n = C_{gp}$$

Push-pull circuits are widely used in power amplifiers because of the inherent characteristic of even harmonic suppression. Because the circuit is balanced to ground, the circuit is easily neutralized as shown in figure (12.11).

The conditions required for neutralization that were just derived make it possible to select the approximate value for the neutralizing capacitor in a given circuit. These capacitors are usually adjustable, because tube-to-tube variations in interelectrode capacitance always occur. Thus the circuit must be accurately neutralized after the amplifier has been constructed. This is most conveniently and safely done by turning the plate power supply off, but leaving the filaments heated. Grid excitation is supplied, and then the neutralizing capacitor is adjusted for minimum energy transfer to the plate circuit. If energy is transferred, it may be detected by a radio receiver connected to a loop of wire coupled to the plate coil. Any other sensitive type of detector can be used.

12.12. Power Supply Connections for Tuned Amplifiers

The plate power supply for power amplifiers is usually supplied from rectifier and filter circuits of the type covered in chapter 14. Similar circuits may also be used to provide the grid bias, but *grid leak bias*

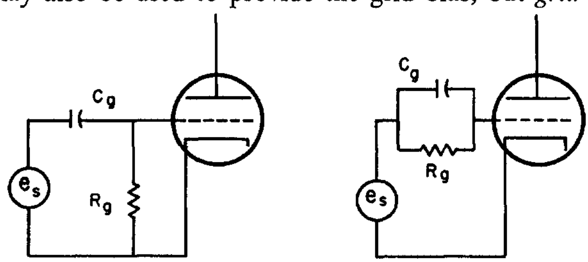


Fig. 12.12. Grid leak bias circuits.

circuits are more common. Cathode bias may be used as long as the switching operation of the tube does not exceed periods of 180°. Class C operation requires bias voltages greater than cutoff, and this cannot be obtained with conventional cathode bias circuits.

Grid leak bias circuits may assume either of two superficially different forms as shown in figure (12.12). Both circuits operate in essentially the same manner. When excitation is first supplied to the grid there is no bias, so that grid current flows on the positive swing of the grid signal. The grid-cathode part of the tube acts as a low resistance diode, so that C_g rapidly charges through the diode resistance r_g. The charging time constant is

$$T_{ch} = \frac{r_g R_g}{r_g + R_g} C_g \doteq r_g C_g$$

because r_g is nearly always very much less than R_g.

As the signal voltage swings negative, grid current ceases, and the accumulated charge on C_g leaks off through R_g with a discharging time constant of

$$T_{dch} = R_g C_g$$

Because of the large difference in size between r_g and R_g, the discharge time constant is much larger than the charging time constant and only a fraction of the accumulated charge leaks off. Clearly, after a few cycles of signal voltage, an equilibrium condition will be attained in which grid current flows for only the short period necessary to replenish

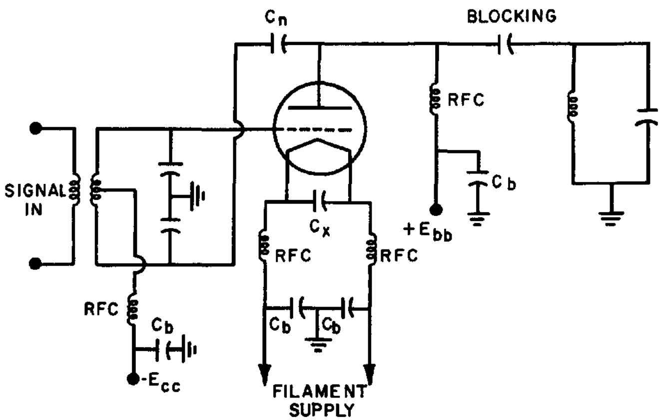


Fig. 12.13. High power, class C tuned amplifier, showing shunt feed, power supply connections, and a Rice system of neutralization. Although many ground connections are shown, they should be tied to a common point on the chassis; see section (10.17).

the charge that leaks off during discharge. As a result, a large negative voltage is developed and this constitutes the gridleak bias voltage. To a close approximation, $E_{cc} = I_g R_g$, where I_g = average or grid direct current. Therefore $R_g = E_{cc}/I_g$.

Tuned amplifiers are operated in the switching mode primarily for the purpose of obtaining a high efficiency of conversion of d-c power to signal power. Therefore steps should be taken to minimize losses throughout the circuit. In particular, it is desirable to avoid signal losses in the various power supplies connected to the amplifier. Thus steps are taken to isolate the plate, grid, and filament power supplies from the signal frequency currents and voltages. This requires the use of radio frequency (RF) chokes and bypass capacitors.

For example, RF chokes are nearly always placed in series with the plate power supply as shown in figure (12.13). An RF choke has negligible resistance to direct current but a high impedance at the signal frequency. Thus only a small signal current flows through the choke and internal impedance of the power supply. What little current does pass through the choke can be bypassed to ground by the capacitor C_b, so that the signal current through the power supply is virtually zero. A similar technique can be applied to the grid circuit where necessary.

Because high power tubes usually use directly heated cathodes, steps should be taken to prevent RF voltages from being impressed across the filament transformers. This is achieved by using all or part of the methods shown in figure (12.13). Radio frequency chokes are placed in series with the filament leads, and bypass capacitors to ground are placed at the transformer connections. To prevent RF voltages between filament leads, another bypass capacitor C_x is connected directly across the filament terminals.

12.13. Parasitics

A *parasitic* is an unwanted or spurious oscillation in an electronic circuit. They occur frequently in tuned power amplifiers. As noted in chapter 10, the most common cause of parasitics is inadvertent formation of a tuned grid, tuned plate oscillator operating at some frequency other than that for which the circuit was designed. A common cause of this circuit condition is the use of shunt feed in both the plate and grid circuits. This should be avoided if possible, but if it must be done, the chokes should be as dissimilar as possible. The ratio of choke inductances should be about 100.

Intertube parasitics often result when tubes are operated in parallel. The use of parasitic suppressing resistors of 10 to 50 ohms in series with the grid of each tube, or the use of chokes in the plate leads, will often remove this difficulty.

Parasitics often result through the use of ungrounded radio frequency tuning capacitors, excessively long leads to the neutralizing condenser, or through the use of multiple radio frequency grounds. Spurious oscillations also result from complex circuits formed when taps are placed on the tank coil for the purposes of loading or tuning. Long leads from tube connections to tank circuits will occasionally cause UHF parasitics.

12.14. Power Oscillators

It was noted in section (12.11) that tuned amplifiers will often oscillate because of feedback produced through C_{gp}. If this feedback is deliberately encouraged, the power amplifier is converted into a power oscillator. Thus power oscillators are designed as high efficiency tuned amplifiers operating in the switching mode, using the design procedure given in section (12.10). When the design is complete, steps are taken to determine the amount of feedback required to produce the necessary grid excitation, driving power, and so on.

Nearly all the considerations affecting power amplifiers also apply to power oscillators. The one notable exception is that fixed bias cannot be used with class C oscillators because the tube would always be cut off, plate current would never flow, and the oscillation would never start. Hence grid leak bias is nearly always used in power oscillators.

Another superficial difference between power amplifiers and oscillators is that crystals are often used in place of the grid tuned circuit to provide frequency stabilization.

12.15. Concluding Remarks

A few miscellaneous topics should be mentioned before closing the discussion. Because of the high harmonic content in the output of class C amplifiers, these circuits are frequently used as *doublers* and *triplers*. In these cases the plate tank circuit is tuned to the desired harmonic frequency. Because of the short conduction angles required for efficient operation as frequency multipliers, tubes operated in this manner require large grid bias voltages and large signal voltages. This difficulty can be partially overcome by using high μ triodes and beam power tubes and pentodes.

In connection with neutralizing problems, it should be remembered from chapter 3 that the grounded grid amplifier is less susceptible to oscillation than the grounded cathode circuit. Thus grounded grid amplifiers are widely used in class C power amplifiers to minimize neutralization problems. The general operation and analysis of such circuits proceeds along the same lines as those illustrated for the grounded cathode circuit. There are some differences in the computation of the power output.

The design of the output coupling circuits is not covered here because such matters are generally covered in detail in standard books of circuit theory.

PROBLEMS

12.1. Design a class C tuned amplifier for optimum operation at 20 megacycles using an RCA 833 tube for which $\mu = 35$; $E_b\,(\max) = 3000$ v; $r_p = 2400$ ohms; $P_p\,(\max) = 300$ w; $C_{gp} = 6.3\ \mu\mu$f; $I_g\,(\max) = 75$ ma; $C_{gk} = 12.3\ \mu\mu$f; $I_b\,(\max) = 500$ ma. The Q at the operating frequency should be 12. Determine all necessary factors in the design as well as the values of L and C in the tank circuit. Assume $E_0 = 750$ v.

12.2. Design a Hazeltine neutralizing circuit for the amplifier of problem (12.1).

12.3. Redesign the amplifier of problem (12.1) for optimum operation in class B.

12.4. Design an audio power amplifier to provide a minimum of 45 w of power at not more than 3% total harmonic distortion. There are many possible solutions to this problem. Work several such out and compare them on the basis of first cost, maintenance, reliability, and other factors.

Chapter 13

MODULATION AND MODULATORS

A continuous wave of constant amplitude, phase and frequency conveys no information whatever other than the fact of its existence. The *quantity* of information thereby transmitted is just one step above zero. Of course, such a signal occupies only a single frequency in the electromagnetic spectrum. There is a correlation between the bandwidth required for transmission and the maximum amount of information that can be transmitted. The treatment of this problem of *information theory* is beyond the scope of this book.

Nevertheless it should be clear that the requirements of information transmission necessitate an alteration of some type in the carrier wave. This alteration invariably increases the bandwidth required for transmission. The alteration is called *modulation* when some characteristic of the carrier wave is made to vary in accordance with the intelligence signal.

Although there are only two fundamentally different ways of modulating a wave, *amplitude modulation* and *angle modulation*, there are a number of important subdivisions of each that are virtually different methods of modulation. Both the principles and methods of modulating systems are briefly discussed in this chapter.

13.1. Principles of Amplitude Modulation

The principle involved in *amplitude modulation* (AM) is easily shown mathematically. If the carrier current is

$$i_c = I_c \cos \omega_c t \tag{13.1}$$

where i_c = instantaneous carrier current; I_c = carrier amplitude; ω_c = carrier frequency in radians/sec and if the intelligence or modulating signal is given by

$$i_m = I_m \cos \omega_m t \tag{13.2}$$

where i_m = instantaneous intelligence current; I_m = amplitude of the modulating signal; ω_m = modulating frequency in radians/sec; then amplitude modulation is produced by causing the carrier amplitude to vary about its own unmodulated value I_c by an amount proportional to the modulating signal amplitude and at the modulating frequency. In other words, the amplitude modulated carrier wave is

$$i_c = I_c(1 + m_a \cos \omega_m t) \cos \omega_c t \tag{13.3}$$

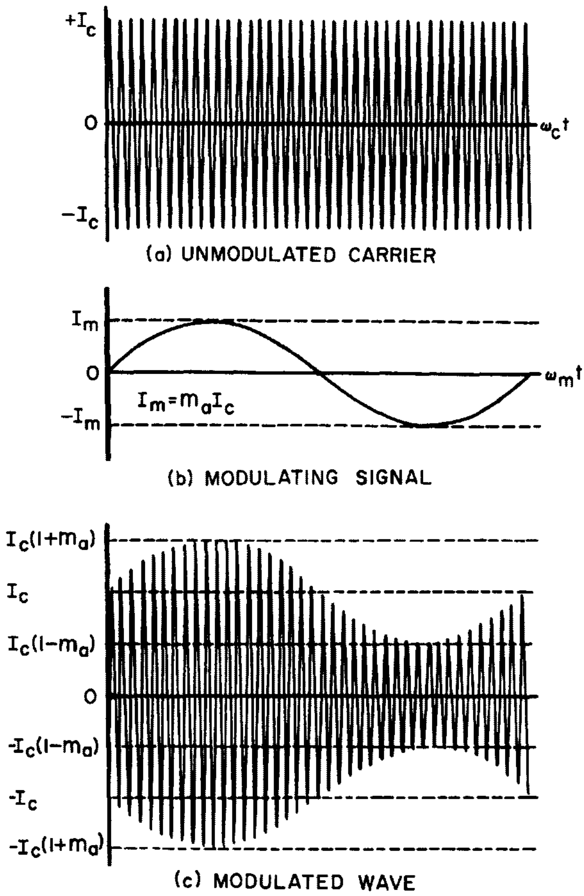


Fig. 13.1. Amplitude modulation.

where $$m_a = I_m/I_c = \text{modulation index} \tag{13.4}$$

The resulting effect and the meaning of the modulation index are illustrated in figure (13.1).

In equation (13.3), multiply through and separate terms; write the result as

$$i_c = I_c \cos \omega_c t + m_a I_c \cos \omega_c t \cos \omega_m t \tag{13.5}$$

The standard trigonometric identity for the product of two cosine functions of different frequency is

$$\cos \omega_c t \cos \omega_m t = \tfrac{1}{2} \cos (\omega_c - \omega_m)t + \tfrac{1}{2} \cos (\omega_c + \omega_m)t \tag{13.6}$$

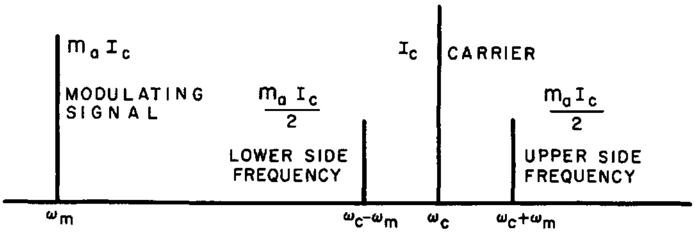


Fig. 13.2. Components produced when a carrier is amplitude modulated by a simple single frequency signal.

Thus the amplitude modulated wave can be expressed as

$$i_c = I_c \cos \omega_c t + I_c \frac{m_a}{2} \cos (\omega_c + \omega_m)t + I_c \frac{m_a}{2} \cos (\omega_c - \omega_m)t \tag{13.7}$$

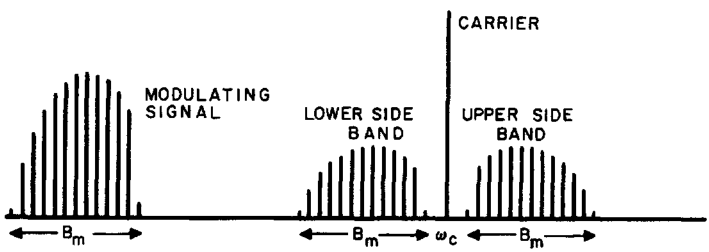


Fig. 13.3. Production of side bands by a complicated modulating signal.

Three terms are produced by the modulation process:

(1) $\cos \omega_c t =$ original unmodulated carrier

(2) $\cos (\omega_c + \omega_m)t =$ upper side frequency

(3) $\cos (\omega_c - \omega_m)t =$ lower side frequency

The relationships between these three components are shown in figure (13.2).

Ordinarily, the modulating signal will contain many terms of different frequency. Each such term will produce its own upper and lower side frequencies. Thus, when many modulating frequencies are involved, upper and lower side *bands* are created as shown in figure (13.3).

Suppose that the modulated current wave flows through a pure resistance circuit of R_L ohms. The total *carrier* power developed is

$$P_c = \tfrac{1}{2} I_c^2 R_L$$

The total power developed by the side frequencies is

$$P_m = \frac{1}{2}\left(\frac{m_a}{2} I_c\right)^2 R_L + \frac{1}{2}\left(\frac{m_a}{2} I_c\right)^2 R_L$$

$$= \frac{m_a^2}{4} I_c^2 R_L$$

or

$$P_m = \frac{m_a^2}{2} P_c \tag{13.8}$$

Thus the sideband power is always $m_a^2/2$ of that in the carrier. Because the modulation index never exceeds unity, the sideband power never exceeds 50% of the carrier power. The sideband power depends upon the square of the modulation index, so it will decrease rapidly as the modulation index drops below unity. Therefore, from the standpoint of energy conversion, best operation occurs when m_a is 1, or when 100% modulation is used.

If 100% modulation is employed so that $m_a = 1$, it is clear from figure (13.1) that the modulated wave will reach a peak value of $2I_c$. Hence the peak power output is

$$P_c\,(\text{peak}) = \frac{1}{2}(2I_c)^2 R_L = 2I_c^2 R_L \tag{13.9}$$

The average carrier power was shown to be

$$P_c = \frac{1}{2} I_c^2 R_L$$

Therefore

$$P_c\,(\text{peak}) = 4P_c \tag{13.10}$$

Hence, if the modulated signal is to be amplified, the amplifier must be capable of supplying four times the carrier power and withstanding twice the peak carrier current and voltage. This is important in the design of high power amplifiers for the transmission of amplitude modulated waves.

13.2. Principles of Angle Modulation

Angle modulation is produced by varying the angle of the carrier wave with respect to time and with respect to the angle of the unmodulated carrier wave. The angle and frequency of a wave are related to one another as follows: $\omega_c =$ carrier frequency in radians/sec; $\phi_c =$ relative phase angle of the carrier wave.

$$\omega_c = \frac{d\phi_c}{dt} \tag{13.11}$$

and

$$\phi_c = \int \omega_c \, dt \tag{13.12}$$

The relative phase angle of the carrier wave can be varied either by changing ϕ_c directly or by changing ω_c.

If the angle is varied directly with time in proportion to some intelligence signal, the carrier is said to be *phase modulated* (PM). If the

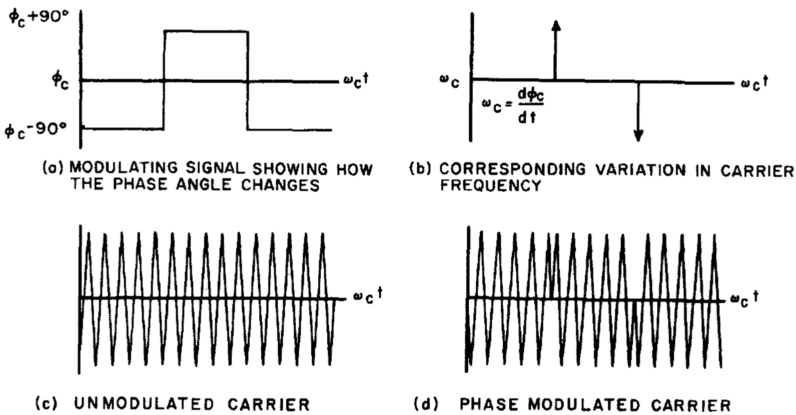


Fig. 13.4. Phase modulation with a square wave.

angle is varied by changing the frequency in proportion to an intelligence signal, the wave is said to be *frequency modulated* (FM). Because of the interrelationships between ω_c and ϕ_c, phase modulation also causes the frequency of the carrier to change during the time the phase is changing. Similarly, if the wave is frequency modulated, the relative phase angle of the wave also varies. This is best understood from figures (13.4) and (13.5), which show phase and frequency modulation of a carrier by a square wave.

When the *phase* of the wave changes abruptly in accordance with the square modulating signal in figure (13.4), there is an instantaneous change in frequency. However, if the frequency changes abruptly, as in figure (13.5), the phase angle of the wave continuously increases. Thus, while the two systems of modulation are different and have different characteristics, variation in frequency accompanies phase modulation and vice versa. This provides the basis for the Armstrong system of frequency modulation to be explained later.

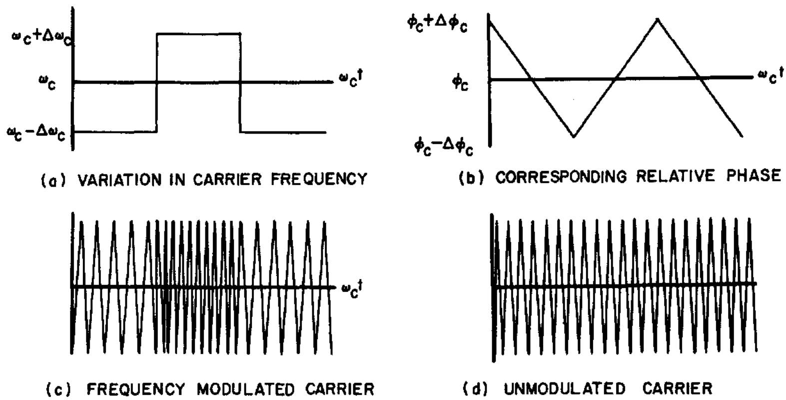


Fig. 13.5. Frequency modulation with a square wave.

13.3. Frequency Modulation (FM)

When frequency modulation is used, the actual carrier frequency deviates from the unmodulated carrier frequency ω_c in direct proportion to the amplitude of the modulating signal. Hence the carrier frequency can be written

$$\omega_c(t) = \omega_c + \Delta\omega_c \cos \omega_m t \tag{13.13}$$

where $\omega_c(t) =$ frequency of the carrier as a function of time when modulated; $\omega_c =$ frequency of the unmodulated carrier; $\Delta\omega_c =$ maximum frequency deviation of the carrier frequency on either side of ω_c; $\omega_m =$ frequency of the modulating signal. Define a new parameter k_f as follows:

$$k_f = \frac{\Delta\omega_c}{\omega_c} \tag{13.14}$$

so that the modulated carrier wave frequency is

$$\omega_c(t) = \omega_c(1 + k_f \cos \omega_m t) \tag{13.15}$$

The instantaneous value of the carrier phase angle can be computed now from equations (13.12) and (13.15). That is,

$$\phi_c(t) = \int \omega_c(t)\, dt$$

or

$$\phi_c(t) = \omega_c \int (1 + k_f \cos \omega_m t)\, dt$$

Evaluation of the integral yields

$$\phi_c(t) = \omega_c t + \frac{\omega_c k_f}{\omega_m} \sin \omega_m t \tag{13.16}$$

if the initial value of the relative phase angle is taken to be zero. The factor multiplying the $\sin \omega_m t$ term is

$$m_f = \frac{\omega_c k_f}{\omega_m} = \frac{\Delta\omega}{\omega_m} \tag{13.17}$$

where $m_f =$ modulation index or deviation ratio for the frequency modulated wave. Therefore the relative phase angle of the frequency modulated carrier current is

$$\phi_c(t) = \omega_c t + m_f \sin \omega_m t \tag{13.18}$$

and the carrier current is

$$\begin{aligned} i_c(t) &= I_c \cos \phi_c(t) \\ &= I_c \cos (\omega_c t + m_f \sin \omega_m t) \end{aligned} \tag{13.19}$$

This last equation can be expanded by a standard trigonometric identity:

$$\begin{aligned} i_c(t) = I_c[&\cos (\omega_c t) \cos (m_f \sin \omega_m t) \\ &- \sin (\omega_c t) \sin (m_f \sin \omega_m t)] \end{aligned} \tag{13.20}$$

These complicated terms can be expanded in an infinite series of Bessel functions,[1] as follows:

$$\begin{aligned} \cos (\omega_c t) \cos (m_f \sin \omega_m t) &= \cos \omega_c t \left[J_0(m_f) + 2 \sum_{n=1}^{\infty} J_{2n}(m_f) \cos (2n\omega_m t) \right] \\ &= J_0(m_f) \cos \omega_c t + 2 \sum_{n=1}^{\infty} J_{2n}(m_f) \cos (\omega_c t) \cos (2n\omega_m t) \end{aligned}$$

$$\begin{aligned} \sin (\omega_c t) \sin (m_f \sin \omega_m t) &= \sin (\omega_c t)\, 2 \sum_{n=1}^{\infty} J_{2n-1}(m_f) \sin (2n-1)\, \omega_m t \\ &= 2 \sum_{n=1}^{\infty} J_{2n-1}(m_f) \sin (\omega_c t) \sin (2n-1)\, \omega_m t \end{aligned}$$

The symbols $J_a(m_f)$ stand for Bessel functions of the first kind, of order a, with argument m_f. Values of these functions for various values of the argument may be found elsewhere.[2]

[1] See, for example, Ruel V. Churchill, *Fourier Series and Boundary Value Problems*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1941, p. 151.

[2] See, for example, E. Jahnke and F. Emde, *Tables of Functions*, Dover Publications Inc., New. York, 1945.

A simplification can be effected by the standard trigonometric identities shown below:

$$\cos\alpha\cos\beta = \tfrac{1}{2}\cos(\alpha-\beta) + \cos(\alpha+\beta)$$

$$\sin\alpha\sin\beta = \tfrac{1}{2}\cos(\alpha-\beta) - \cos(\alpha+\beta)$$

Therefore the preceding infinite series can be written as follows:

(1) First series

$$J_0(m_f)\cos\omega_c t + \sum_{n=1}^{\infty} J_{2n}(m_f)\cos(\omega_c - 2n\omega_m)t$$
$$+ \sum_{n=1}^{\infty} J_{2n}(m_f)\cos(\omega_c + 2n\omega_m)t$$

(2) Second series

$$\sum_{n=1}^{\infty} J_{2n-1}(m_f)\cos[\omega_c - (2n-1)\omega_m]t$$
$$+ \sum_{n=1}^{\infty} J_{2n-1}(m_f)\cos[\omega_c + (2n-1)\omega_m]t$$

Substitute these series back into equation (13.20) for the carrier current. Substitute values for n for a few terms and write

$$\begin{aligned} i_c(t) = I_c\{&J_0(m_f)\cos\omega_c t \\ &- J_1(m_f)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \\ &+ J_2(m_f)[\cos(\omega_c - 2\omega_m)t + \cos(\omega_c + 2\omega_m)t] \\ &- J_3(m_f)[\cos(\omega_c - 3\omega_m)t - \cos(\omega_c + 3\omega_m)t] \\ &+ J_4(m_f)[\cos(\omega_c - 4\omega_m)t + \cos(\omega_c + 4\omega_m)t] \\ &- J_5(m_f)[\cos(\omega_c - 5\omega_m)t - \cos(\omega_c + 5\omega_m)t] \\ &+ \ldots\} \end{aligned}$$

The nature of the frequency modulated wave is easily deduced now from the preceding equation. Note that there is always an infinite number of frequency components equally distributed on either side of the center frequency ω_c. The spacing between successive components is always equal to the modulating frequency ω_m. The amplitude of each term is governed by the modulation index m_f.

It is a relatively simple matter to make plots of the spectra for frequency modulated waves. Assume values for m_f, ω_c, and ω_m and then look up the corresponding values for the Bessel functions in footnote reference (2). Several sample spectra have been computed and plotted in figure (13.6).

Figure (13.6) shows that the properties of the Bessel functions cause the side frequency components to converge to zero rather rapidly. Thus for all *practical* purposes there is a definite bandwidth required for transmission even though it is theoretically infinite. It is also clear from figure (13.6) that the bandwidth required for transmission is *not* twice the frequency deviation. It is always greater.

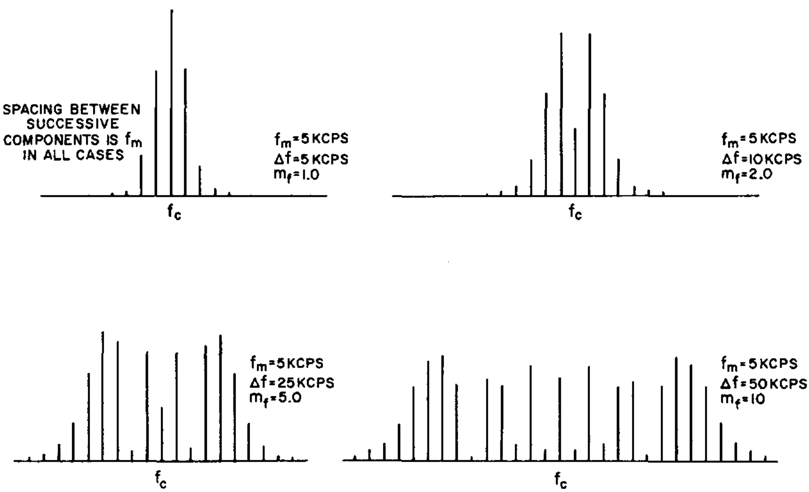


Fig. 13.6. Effect of variation in frequency deviation on the spectrum of an FM signal.

The center frequency term, corresponding to the frequency of the unmodulated carrier, will disappear completely when $J_0(m_f)$ is zero. This function is zero whenever[3]

$$m_f = 2.405, \quad 5.520, \quad 8.654, \quad 11.79, \quad 14.93, \quad \ldots$$

Under such conditions, *all* the power appears in the sidebands and is useful in signal transmission.

Another important characteristic of the spectrum of a frequency modulated wave is shown in figure (13.7). This shows the effect on the spectrum when the modulating frequency is varied and the frequency deviation is fixed; this is the usual case in frequency modulated transmission. Note that there are more side frequencies produced for a low

[3] Ruel V. Churchill, *Fourier Series and Boundary Value Problems*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1941, p. 157.

modulating frequency than are produced by a higher modulating frequency. However, the total bandwidth required for transmission is practically the same in both cases.

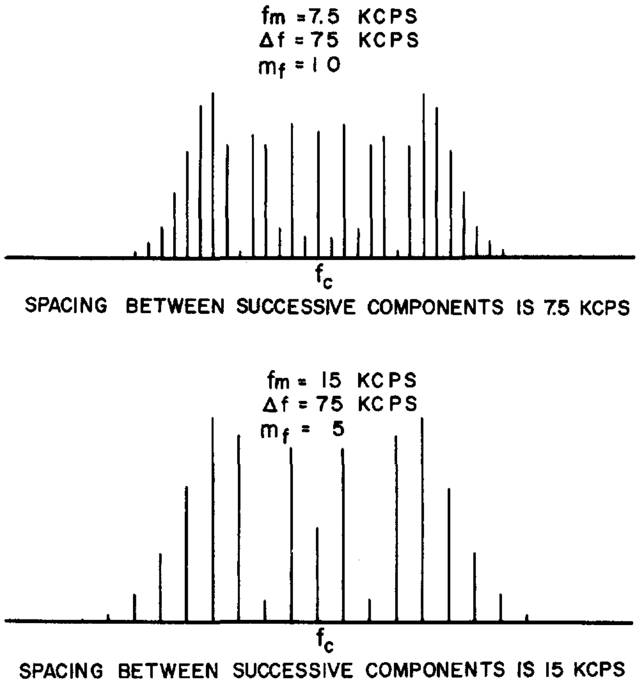


Fig. 13.7. Effect of variations in modulating frequency on the spectrum.

13.4. Phase Modulation

In the absence of modulation the carrier current

$$i_c(t) = I_c \cos \omega_c t \tag{13.21}$$

has an angle $\omega_c t$ that constantly increases with increasing time. If the carrier is phase modulated, this phase angle is advanced and retarded with respect to its unmodulated value in proportion to the modulating signal. Thus if m_p is used to denote the *phase modulation index*, or maximum phase deviation, the carrier angle is

$$\phi_c(t) = \omega_c t + m_p \cos \omega_m t \tag{13.22}$$

Consequently, the phase modulated carrier current is

$$i_c(t) = i_c \cos (\omega_c t + m_p \cos \omega_m t) \tag{13.23}$$

From a comparison of the phase modulated carrier current given in equation (13.23) and the frequency modulated carrier current given in equation (13.19), it is clear that both have essentially the same form, differing only in the time variation of the phase angle. This variation is cosinusoidal for the phase modulated wave and sinusoidal for the frequency modulated wave. Hence it would be a simple matter to expand equation (13.23) for the phase modulated wave into an infinite series of Bessel functions just as we did for the frequency modulated wave. Because of the similarity of form, phase modulated waves have the same spectra as frequency modulated waves if (1) the modulating signal is sinusoidal; (2) $m_p = m_f$. Thus the spectra plotted previously in figure (13.6) for various values of m_f can also apply to phase modulated waves under the two conditions stipulated above.

As in the case of frequency modulation, the total average power of the modulated wave is equal to that of the unmodulated carrier. Energy is simply diverted by the modulation process from the center frequency into the side bands.

Although the similarities between phase and frequency modulation are quite pronounced, there are important differences. The most important difference arises from the effect of modulating signals of different frequency on the bandwidth required for transmission. In FM transmission the frequency deviation Δf is held constant as the modulating frequency varies. This causes the modulation index, $m_f = \Delta f / f_m$, to decrease as f_m increases. In phase modulated transmission, the phase deviation m_p is held constant as the frequency of the modulating signal varies.

For a given value of m_p or m_f, the *number* and amplitude of the significant frequency components is fixed. However, these components are spaced ω_m radians apart. Hence the bandwidth required for transmitting a phase modulated signal is linearly related to the frequency of the modulating signal. If sufficient bandwidth is allowed for transmission of a high frequency modulation component, only a small fraction of this bandwidth is utilized when the modulating frequency is low. Thus a phase modulated wave does not make effective use of its allotted part of the frequency band.

For a frequency modulated wave, as the modulating frequency f_m is increased and the deviation is held constant, m_f decreases. Thus two partially counterbalancing effects occur simultaneously:

(1) The increase in the modulating frequency increases the *separation* between the frequencies.

(2) The decrease in m_f reduces the *number* of significant frequency components.

As a result, the bandwidth required for transmission is more nearly constant than that required for transmission of a phase modulated wave. In other words, FM signals make more efficient use of their allotted bandwidths than do phase modulated signals and are preferred for transmission for this reason.

Despite the preference for FM for transmission, it is generally easier to phase modulate an oscillator than to frequency modulate it. Thus phase modulation is often preferred in the transmitter.

13.5. Comparison of AM and FM; Interference and Noise

Most noises, such as static, ignition systems, and so on, are basically amplitude modulated signals and are received and amplified by an AM

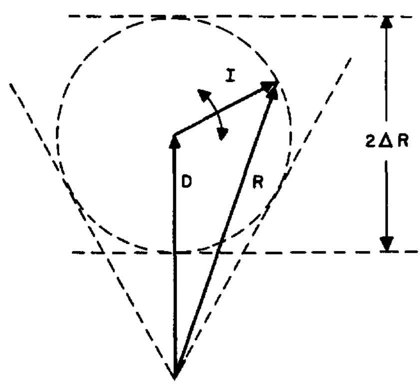


Fig. 13.8. Interference between two waves.

receiver just as though the noise was a part of the signal. Usually FM receivers are designed with clippers or limiters (see chapter 15) so that they do not respond to amplitude variations. Thus such noise is largely eliminated from the output of FM receivers.

Frequency modulation largely overcomes the effects of signal interference produced when an interfering signal of about the same frequency as the desired signal appears in the receiver input. Call the desired signal D and the interfering signal I. They add as phasors in the receiver input to produce a resultant R as shown in figure (13.8). Because of the slight frequency difference between the two signals, I

rotates about D. Clearly, the amplitude of the resultant will vary over a wide range of $2\Delta R$ and cause the amplitude of the received signal to vary. The interference is severe in this case.

Now suppose that D and I represent frequency modulated waves of slightly different center frequency. The rotation of I about D causes the phase of the resultant to swing over the range marked in figure (13.8). It is about ± 0.5 radian in this case. In a typical FM signal, the actual phase deviation may run anywhere from about ± 35 to $\pm 15{,}000$ or more radians. Thus the interference of about 0.5 radians is small.

In most *audio* signals the greater part of the signal power is contained in the low frequency components. The treble notes are not strong and may not greatly exceed noise components. Thus *pre-emphasis* is generally used in FM transmitters to accentuate the treble components. A reciprocal process of *de-emphasis* is used in the receiver to restore the true tonal balance. However, the noise level is greatly reduced by the de-emphasis circuit and adds another degree of superiority for the FM system over the AM.

13.6. Pulse-Time Multiplexing

The tremendous and ever increasing demand for more communication has led to the development of various *multiplexing* systems. Multiplexing refers to the simultaneous transmission of more than one signal on a common carrier wave. Two methods are in common use, as follows:

(1) Frequency division. Each signal is associated with a separate *subcarrier* frequency.

(2) Pulse-time division. Samples of each signal are transmitted in time sequence at a high sampling frequency on the same carrier frequency.

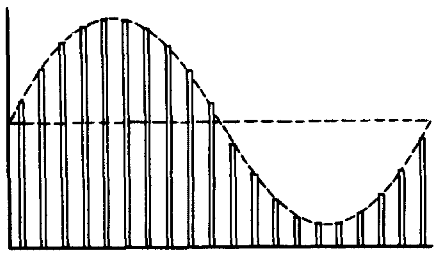
The time-division multiplexing systems ordinarily use pulses. In the sampling interval only one characteristic of the pulse can be made to vary in accordance with the signal being sampled. This has led to the development of a number of systems such as *pulse amplitude modulation* (PAM), *pulse width modulation* (PWM), *pulse position modulation* (PPM), *pulse frequency modulation* (PFM), and so on. The scope of this subject is entirely too broad to be covered in any detail here. Only the general features of PAM will be presented.

In these systems a series of equally spaced identical pulses is obtained from some sort of *commutating* device. This may be a mechanical or an electric contrivance, depending upon the pulse frequency required.

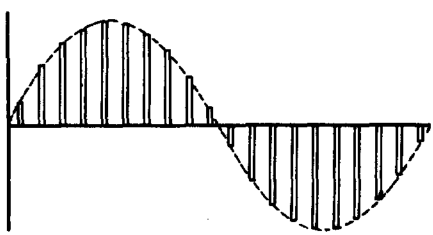
In any case, each such pulse obtained is varied in amplitude, width, position, or frequency in accordance with the signal being sampled.

For example, two types of pulse amplitude modulation are illustrated in figure (13.9) to show how the pulse amplitude is made variable in accordance with the amplitude of the signal being sampled. Similar sketches may be made for the other pulse systems.

Figure (13.9) illustrates the sampling of only one channel. Actually, many channels may be sampled by the pulse train. In an 8 channel



(a) UNIDIRECTIONAL



(b) BIDIRECTIONAL

Fig. 13.9. Pulse amplitude modulation (PAM).

system each channel is sampled by successive pulses. A synchronizing pulse is also required. Thus one complete cycle of an 8 channel PAM system would show 8 amplitude modulated pulses, one for each channel, and one synchronizing pulse. The sampling pulses shown in figure (13.9) correspond to every 9th pulse from the original pulse train. This is illustrated in figure (13.10).

The instantaneous amplitude of the signal in channel 1 is sampled, then channel 2, and so on through channel 8. Then the entire sampling process repeats after the transmission of the synchronizing pulse. The rate at which one channel is sampled is governed by the nature of the signal transmitted. 8000 pulses per second is the usual sampling rate for voice communication channels. This rate is so fast that the average listener cannot detect the difference between true and sampled signals. The detector or demodulating circuits are designed to help restore the sampled signal to its original continuous form.

Once the pulse time multiplexing has been achieved, the signal may be transmitted through space on a carrier. The carrier may be amplitude or frequency modulated by the pulse-time multiplexed signal.

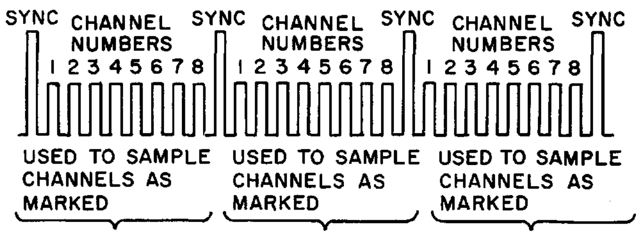


Fig. 13.10. Pulse train in an eight channel PAM system before modulation.

13.7. Methods of Amplitude Modulation

The discussion up to this point in this chapter was solely concerned with the *principles* of modulation, without regard to the methods of achieving it in practice. The remainder of this chapter presents practical modulation methods.

In chapter 11 it was shown that the square law characteristics of some electronic devices can be used to produce amplitude modulation. It was also shown that such systems are generally inefficient and unreliable. Most practical amplitude modulation systems require the electronic component to operate in the switching mode at high efficiency.

When large amounts of power are involved, high efficiency class C amplifiers are usually modulated. Any one of three methods may be used: (1) plate modulation; (2) cathode modulation; (3) grid modulation. The general operating principle involved is the same in all three cases, but the operating characteristics are quite different; each has particular attributes and disadvantages that render it suitable or impractical in various applications. This will be discussed later.

The general principle involved in all three methods is easily understood from the transfer characteristics of the amplifier. These are shown in figures (13.11) and (13.12) for class C operation.

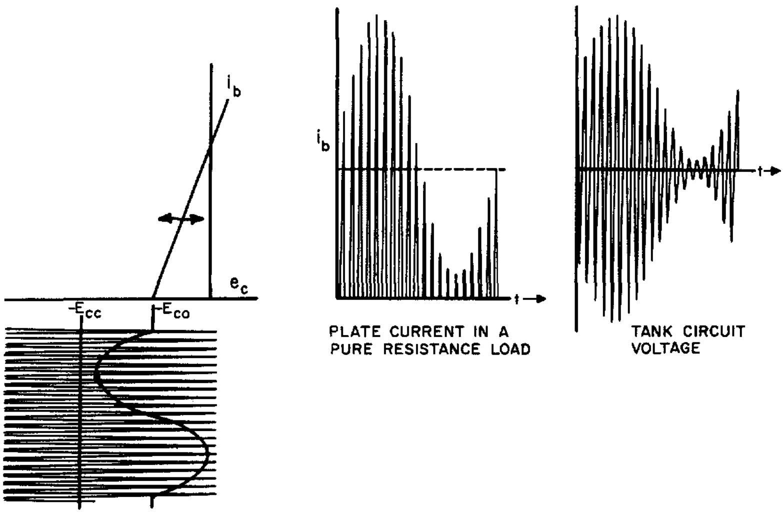


Fig. 13.11. Plate modulation of a class C amplifier. E_{cc} = constant, E_{co} = variable.

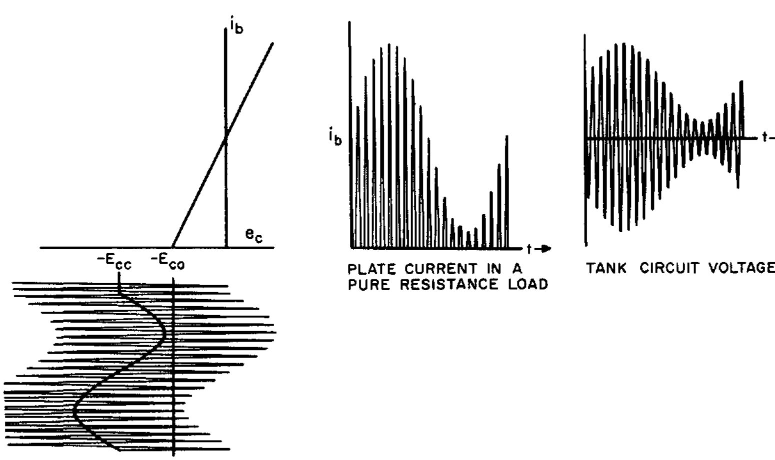


Fig. 13.12. Grid modulation of a class C amplifier. E_{cc} = variable, E_{co} = fixed.

When the amplifier is plate modulated the modulating signal is inserted in series with the plate power supply. Hence the total power supply voltage varies as a function of time in accordance with the modulating signal. This causes the transfer characteristic of the tube to slide parallel to itself, in the ideal case, as shown in figure (13.11a). This alters the cutoff voltage in accordance with the modulating signal because there is an almost linear relationship between cutoff voltage and plate supply voltage. With the bias held constant and the cutoff voltage varying in accordance with the modulating signal, the constant amplitude grid signal produces an amplitude modulated signal in the plate circuit. The effect is clear from figure (13.11a).

In the case of plate modulation, the required effect is produced by varying the relative spacing between the cutoff voltage and the bias voltage as a function of time. In grid modulation the same principle is used except that the grid bias is varied while the cutoff voltage is held constant. This is illustrated in figure (13.11b). The effect is accomplished by inserting the modulating signal in series with the grid carrier signal voltage.

It seems fairly obvious that cathode modulation is a combination of both plate and grid modulation. The modulating signal is inserted in series with the cathode of the tube. Therefore it causes both the grid bias and cutoff voltage to vary, but in opposite directions.

13.8. Plate Modulated Class C Amplifier Design[4]

This section is a mathematical analysis of a plate modulated class C amplifier, concluding with a practical design procedure. The design of high efficiency unmodulated class C amplifiers was covered in chapter 12. The material presented there will now be adapted to the case of the modulated amplifier.

The circuit diagram of a plate modulated class C amplifier is shown in figure (13.13). In this circuit the total effective plate power supply voltage is

$$e'_{bb} = E_{bb} + E_0 + E_m \cos \omega_m t \tag{13.24}$$

or

$$e'_{bb} = E'_{bb}(1 + m_a \cos \omega_m t)$$

where

$$m_a = E_m/E'_{bb}.$$

$$E'_{bb} = E_{bb} + E_0 \tag{13.25}$$

and $E_0 =$ equivalent intercept voltage from the equivalent circuit.

[4] W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., New York, 1937, chap. 17.

Evidently the plate supply voltage varies with time in accordance with the modulating signal $e_m = E_m \cos \omega_m t$.

For *optimum* operation of a class C amplifier it was shown in chapter 12 that the tube should be driven just up to the point of saturation. When this condition exists, the maximum amplitude I_{pm_1} of the

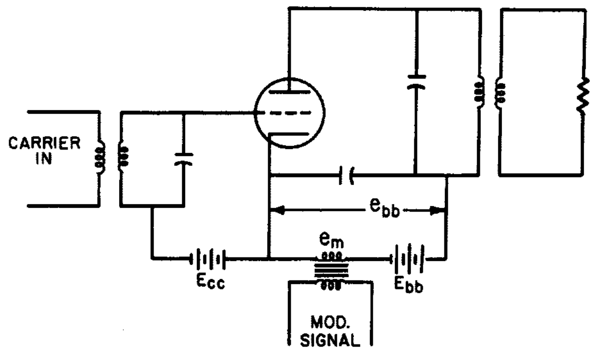


Fig. 13.13. Plate modulated class C amplifier; neutralizing circuit omitted.

fundamental component of plate current was derived and given in equation (12.53) as

$$I_{pm_1} = \left(\frac{\alpha}{\alpha + B}\right)\frac{E'_{bb}}{R_L} \tag{13.26}$$

where

$$\alpha = (\mu + 1)\frac{R_L}{r_p} \tag{13.27}$$

$$B = \beta(1 - \cos \theta_c) \tag{13.28}$$

$$\beta = \frac{\pi}{\theta_c - \sin \theta_c \cos \theta_c} \tag{13.29}$$

where $2\theta_c$ = tube conduction angle; R_L = input impedance of the tuned load circuit at resonance. According to equation (13.26), the amplitude of the fundamental component of plate current will be a linear function of the equivalent power supply voltage if α and B are constants. If this condition can be produced so that I_{pm_1} is a linear function of the e'_{bb} in equation (13.26), the circuit is called a *linear modulator*.

Unfortunately, as e'_{bb} varies with time in accordance with the modulating signal, the cutoff voltage E_{co} of the tube changes, and this causes the tube conduction angle to vary. Because B depends upon θ_c, it also changes, and linear modulation is apparently *not* achieved. Of course, the effect of variations in B could be minimized by making α much larger than B. This *is* helpful and can be accomplished through the use of large values for R_L. A more exact approach is developed in the following paragraphs.

From equation (12.34) of chapter 12,

$$E_1 = E_g - \frac{I_{p_1} R_L}{\mu}$$

and according to equation (12.37),

$$I_{p_1} = \frac{\mu E_g}{R_L + \beta r_p}$$

Hence
$$E_1 = E_g - \frac{R_L}{R_L + \beta r_p} E_g = \frac{\beta r_p}{R_L + \beta r_p} E_g \tag{13.30}$$

or, for operation at the saturation point,

$$E_1 = \frac{\beta r_p}{R_L + \beta r_p} E_{gm} \tag{13.31}$$

The maximum amplitude of the grid excitation E_{gm} is related to the effective power supply voltage e'_{bb} by equation (12.52):

$$E_{gm} = \frac{\alpha + (\mu + 1)\beta}{\mu(\alpha + B)} e'_{bb} \tag{13.32}$$

Therefore equation (13.31) can be written

$$E_1 = \frac{\beta r_p}{R_L + \beta r_p} \cdot \frac{\alpha + (\mu + 1)\beta}{\mu(\alpha + B)} e'_{bb} \tag{13.33}$$

Now substitute the value for α given in equation (13.27) and the equation for E_1 simplifies to

$$E_1 = \frac{\mu + 1}{\mu} \cdot \frac{\beta}{\alpha + B} e'_{bb} \tag{13.34}$$

According to equation (12.27), $E_x = E_1 \cos \theta_c$, or

$$E_x = \frac{\mu + 1}{\mu} \cdot \frac{\beta \cos \theta_c}{\alpha + B} e'_{bb} \tag{13.35}$$

Rearrange terms somewhat as follows:

$$\frac{\mu}{\mu + 1} \cdot \frac{E_x}{e'_{bb}} = \frac{\beta \cos \theta_c}{\alpha + B}$$

and define the quantity on the left as η. That is,

$$\eta = \frac{\mu}{\mu + 1} \cdot \frac{E_x}{e'_{bb}} \tag{13.36}$$

$$\eta = \frac{\beta \cos \theta_c}{\alpha + B} \tag{13.37}$$

According to equation (12.9),

$$E_x = E_{cc} - \frac{e'_{bb}}{\mu}$$

so that the equation for η in (13.36) becomes

$$\eta = \frac{\mu}{\mu + 1}\left(\frac{E_{cc}}{e'_{bb}} - \frac{1}{\mu}\right) \tag{13.38}$$

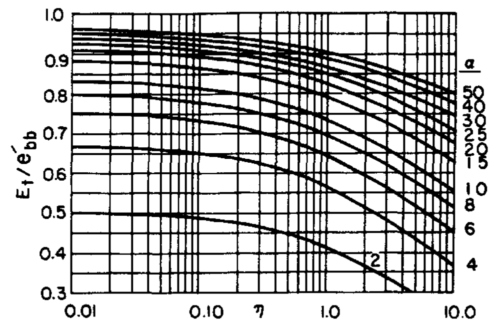


Fig. 13.14. (Tank voltage)/(supply voltage) as a function of η and α. (From W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., 1937, p. 592.)

For a given tube and bias voltage the parameter η can be computed from this formula as a function of the plate supply voltage e'_{bb}.

Substitute the expressions for β and B as functions of θ_c into equation (13.37). The result can eventually be written

$$\frac{\pi}{\alpha} \cdot \frac{\eta + 1}{\eta} = \frac{\theta_c}{\cos \theta_c} - \sin \theta_c \tag{13.39}$$

It is now a comparatively simple matter to make a plot of θ_c as a function of η with α as a family parameter. Simply assume values for θ_c and α and plot the resulting value for η. The curves so obtained can then be used to determine θ_c for any combination of values for α and η. These values for θ_c can then be used to compute B. The factor $\alpha/(\alpha + B)$ is then easily determined and plotted as a function of η with α as the parameter. The results of this determination are given in figure (13.14).

Figure (13.14) also gives the relationship between the amplitude of the tank circuit voltage E_T and the power supply voltage e'_{bb}. This is easily proved. The tank circuit voltage is

$$E_T = I_{pm_1} R_L = \frac{\mu R_L}{R_L + \beta r_p} E_{gm} \tag{13.40}$$

Substitute for E_{gm}.

$$E_T = \frac{\mu R_L}{R_L + \beta r_p} \cdot \frac{\mu + 1}{\mu} \cdot \frac{\beta}{\alpha + B} e'_{bb}$$

$$= \frac{\alpha}{\alpha + B} e'_{bb}$$

or

$$\frac{E_T}{e'_{bb}} = \frac{\alpha}{\alpha + B} \tag{13.41}$$

Thus figure (13.14) is a plot of the amplitude of the output voltage as a function of the plate supply voltage. It seems clear that linear modulation is *not* exactly achieved, but can be rather closely approximated. If linear modulation were produced, the curves in figure (13.14) would be linear.

Figure (13.14) is useful in computing the modulated output voltage from the amplifier for any tube and power supply combination.

The plate circuit efficiency can be determined as a function of the modulation. According to equation (12.57), the efficiency at the saturation point is

$$\eta_{pm} = \left(\frac{\alpha}{\alpha + B}\right) \eta_A \tag{13.42}$$

where $\eta_A =$ asymptotic efficiency $=$ function of θ_c. The factor in parentheses in equation (13.42) is plotted in figure (13.14) as a function of α and η, where η_A is a function of θ_c and θ_c is given by equation (13.39) in terms of η and α. Thus η_A can also be computed and a plot made of the plate circuit efficiency as a function of the parameters η and α. Such a graph is given in figure (13.15).

This figure shows that the efficiency is approximately constant as η varies because of modulation. It was shown in chapter 12 that

$$P_p = P_{ac} \frac{1 - \eta_p}{\eta_p} = \text{plate dissipation}$$

$$P_{ac} = \text{total signal power}$$

However, for an amplitude modulated amplifier it was shown at the beginning of this chapter that $P_{ac} = P_c(1 + m_a^2/2)$, where $P_c =$ power in the unmodulated carrier, so

$$P_p = P_c\left(1 + \frac{m_a^2}{2}\right)\left(\frac{1 - \eta_p}{\eta_p}\right)$$

Clearly, the plate dissipation increases as the modulation index is increased if the efficiency is constant. For 100% modulation the plate dissipation is 50% greater than that for the unmodulated amplifier. In other words, the plate dissipation in the absence of modulation is only 67% of the value with 100% modulation. This should be remembered when designing the amplifier. However, 100% modulation is produced

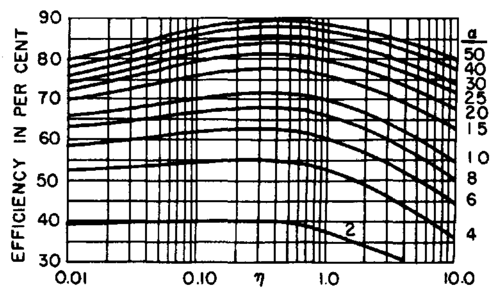


Fig. 13.15. Efficiency as a function of α and η. (From W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., 1937, p. 592.)

only momentarily in most cases, and temporary overloads can often be tolerated. Thus, in the design of the amplifier, the value of the parameter γ should be computed for the 67–100% value of the allowable plate dissipation.

13.9. Some Other Amplitude Modulation Circuits

As noted earlier, amplitude modulation of a class C amplifier can also be secured by inserting the modulating voltage in series with either the grid or cathode leads of the tube. Typical circuit diagrams are shown in figure (13.16).

In the ideal case the amplitude of the plate current in a class C amplifier can be made a linear function of the grid bias. Thus, in principle, linear grid modulation is possible. Assuming linear modulation, both the supply direct current and RF tank current are linear functions of the grid voltage. This being true, then because the *average* grid voltage is the same before and after the insertion of the sinusoidal grid modulating signal, the average plate current is unchanged by the modulation process. The d-c power input is also unchanged. However, the RF tank current and signal power output increase when modulation is applied. Therefore the plate circuit efficiency must increase. This reduces the plate dissipation, an effect that is almost the exact reverse of the situation with plate modulation.

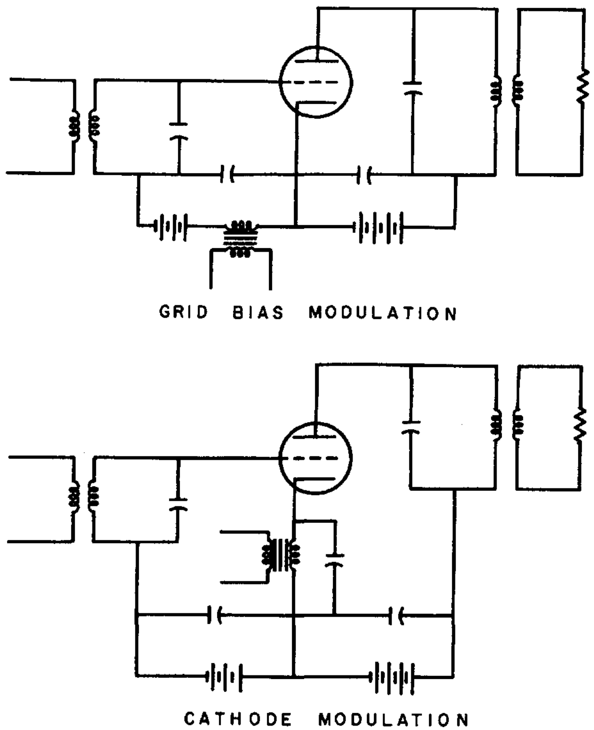


Fig. 13.16. Grid and cathode modulation of a class C amplifier.

The main advantage of grid modulation is the low modulating power required. This allows the use of small, lightweight, inexpensive modulation transformers and low power modulating amplifiers. However, it is difficult to achieve linear modulation and the highest obtainable efficiencies are quite low in comparison to the plate modulated case.

The characteristics of cathode modulated amplifiers fall about midway between grid and plate modulated circuits.

When the amplifier tube is a pentode, suppressor or screen modulation can be used. Suppressor modulation and the power relations

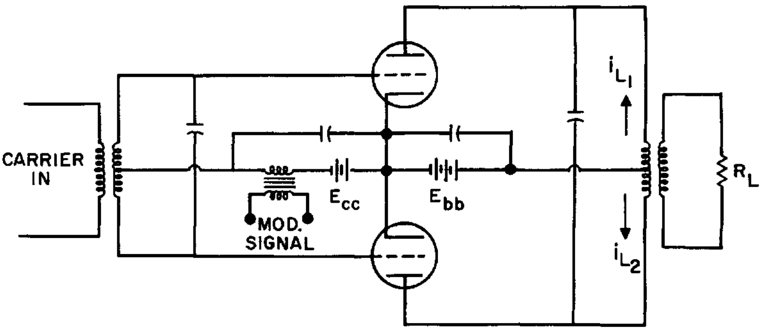


Fig. 13.17. Balanced modulator.

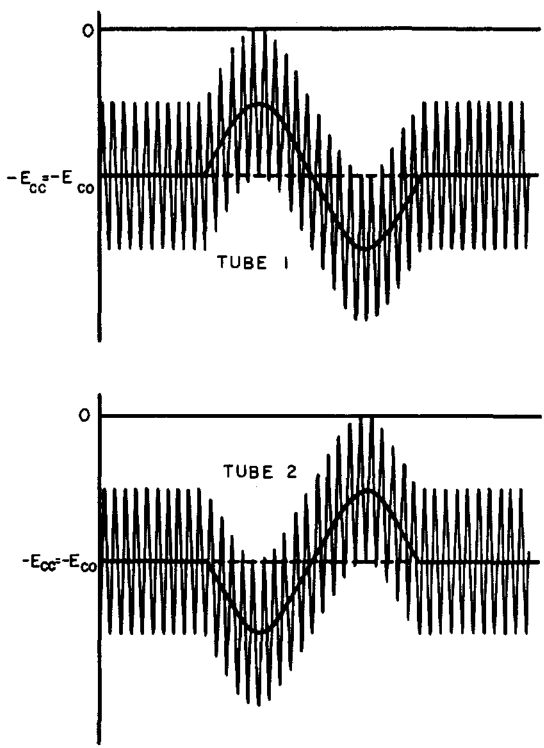


Fig. 13.18. Class B operation of a balanced modulator (100% modulation).

involved are similar to grid modulation. Screen grid modulation is only occasionally used in low power transmitters.

13.10. Balanced Modulator

The circuit diagram of a *balanced modulator* is shown in figure (13.17). This is essentially a push-pull carrier frequency input amplifier with

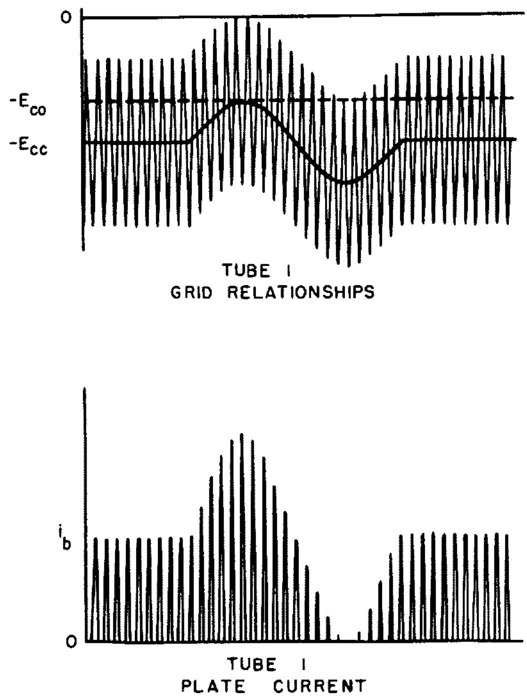


Fig. 13.19. Class C operation of a balanced modulator (100% modulation).

grid modulation. However, the modulating signal is so applied that the effect produced in the circuit of tube 1 is 180° out of phase with that produced in tube 2. The general operating conditions in the circuit are shown in figures (13.18) and (13.19).

If the load circuit has a high Q at the carrier frequency and the input impedance is essentially constant over the pass band of interest, it is clear that the currents in the tank coil are

$$i_{L_1} = I_L(1 + m_a \cos \omega_m t) \cos \omega_c t \tag{13.43}$$

$$i_{L_2} = I_L(1 - m_a \cos \omega_m t) \cos \omega_c t \tag{13.44}$$

As in the push-pull amplifier, the output voltage or current is proportional to the difference in primary currents because they set up opposing mmf's. Hence

$$e_0(t) = k(i_{L_1} - i_{L_2}) = 2kI_L m_a \cos \omega_m t \cos \omega_c t \tag{13.45}$$

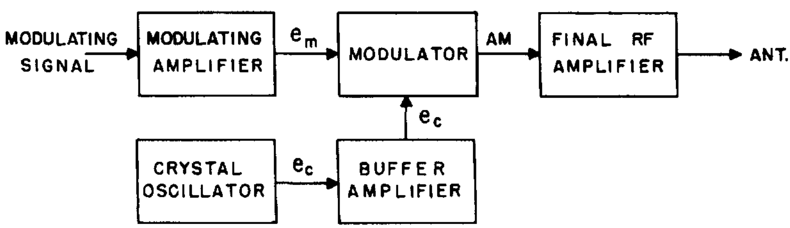
or, using the standard trigonometric identity for the product of two cosines,

$$e_0(t) = kI_L m_a [\cos (\omega_c - \omega_m)t + \cos (\omega_c + \omega_m)t] \tag{13.46}$$

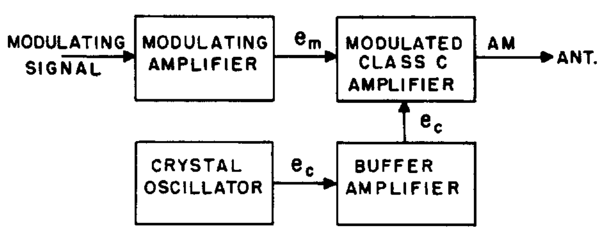
Thus the output has only the sideband terms. The carrier has been suppressed. This is often used in *carrier suppressed* transmission.

13.11. High Level and Low Level Modulation

Amplitude modulated transmitters may be modulated at a high power level or a low power level as shown by the system block diagrams



(a) LOW LEVEL MODULATION



(b) HIGH LEVEL MODULATION

Fig. 13.20. Two types of AM transmitters.

of figure (13.20). This has resulted in the designation of two types of modulation as *high level* and *low level*.

When low level modulation is used, the amplifier stages following the modulator must necessarily operate in class B if linear amplification at high efficiency is to be achieved. Class C amplifiers cannot be used to amplify amplitude modulated signals without introducing serious distortion or using trick circuitry. Thus the over-all efficiency of low level modulated transmitters is usually less than that of transmitters using high level modulation.

Plate modulation is nearly always used for high level modulation. This allows the final stage of the transmitter to operate at nearly constant efficiency, whereas the efficiency of a class B amplifier varies with the grid excitation. Although the initial cost of components for high level modulation is somewhat higher than for low level modulators, the operating cost is so much lower that high level modulation is most economical in fixed, high power, long term investment transmitters.

13.12. Design of Class B Amplifiers[5]

When low level modulation is used, class B linear amplifiers are nearly always required for high efficiency amplification of the amplitude modulated signal. The design procedure for such circuits can be developed rather quickly from the general material given in chapter 12.

The signal power developed in the load is

$$P_{ac} = \frac{1}{2} I_{p_1}^2 R_L \tag{13.47}$$

and the d-c power input to the amplifier is

$$P_{dc} = E_{bb} I_b' \tag{13.48}$$

The plate circuit efficiency is

$$\eta_p = \frac{P_{ac}}{P_{dc}} \tag{13.49}$$

while the plate dissipation is

$$P_p = P_{dc} - P_{ac} \tag{13.50}$$

The distinguishing characteristic of a class B amplifier is the linear relationship between the amplitude of the plate current I_{p_1} and the amplitude of the grid excitation voltage. The linear region extends from zero to the saturation point. If we let I_{p_1} denote the fundamental component of plate current for any value of grid excitation in the

[5] This is the method used by W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., New York, 1937, pp. 582–590.

absence of modulation, then I_{pm_1} will signify the fundamental component of plate current at the saturation point.

Now define a parameter K such that

$$K = \frac{I_{p_1}}{I_{pm_1}} \tag{13.51}$$

or

$$I_{p_1} = KI_{pm_1} \tag{13.52}$$

As a result, equation (13.47) for the signal power output is

$$P_{ac} = K^2 P_{ac} \text{ (at saturation)} \tag{13.53}$$

From equations (12.40) and (12.42) of the preceding chapter, $I_{p1}/I_b' = 2\eta_A$. Hence the d-c power input becomes

$$P_{ac} = KP_{dc} \text{ (at saturation)} \tag{13.54}$$

The plate circuit efficiency is thereby of the form

$$\eta_p = K\eta_{pm} \tag{13.55}$$

and the plate dissipation is

$$P_p = KP_{dc} \text{ (at saturation)} - K^2 P_{ac} \text{ (at saturation)} \tag{13.56}$$

The values of P_{ac} (at saturation) and η_{pm} were given in equations (12.57) and (12.55) of the preceding chapter. The value of P_{dc} (at saturation) is easily computed from these two relationships. As a result, the following four expressions are obtained:

$$\eta_p = \frac{\alpha}{\alpha + B} K\eta_A \tag{13.57}$$

$$P_{ac} = K^2 \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{\alpha}{(\alpha + B)^2} \tag{13.58}$$

$$P_{dc} = \frac{K}{\eta_A} \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{1}{\alpha + B} \tag{13.59}$$

$$P_p = K \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{1}{\alpha + B} \left(\frac{1}{\eta_A} - K \frac{\alpha}{\alpha + B}\right) \tag{13.60}$$

A class B amplifier operates with a fixed conduction angle of $2\theta_c = 180°$. As a result, all the factors dependent only upon θ_c become constant. Thus $B = \beta = 2$; $\eta_A = 0.785$, or 78.5%; $\eta_A = \pi/4$.

Thus equations (13.57) through (13.60) reduce to

$$\eta_p = \frac{\alpha}{\alpha + 2} \cdot \frac{\pi K}{4} \tag{13.61}$$

$$P_{ac} = K^2 \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{\alpha}{(\alpha + 2)^2} \tag{13.62}$$

$$P_{dc} = \frac{4K}{\pi} \cdot \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{1}{\alpha + 2} \tag{13.63}$$

$$P_p = K \frac{\mu + 1}{\mu} \cdot \frac{g_m E_{bb}'^2}{2} \cdot \frac{1}{\alpha + 2} \left(\frac{4}{\pi} - \frac{K\alpha}{\alpha + 2}\right) \tag{13.64}$$

The last equation for the plate dissipation is the key factor in the derivation of the design procedure. Assuming sinusoidal modulation, the average plate current I_b' is not affected by the modulation in a linear amplifier. Hence the power input P_{dc} is not affected by the presence or absence of a modulating signal. However, the signal power output is increased by the amount of power in the sidebands. Hence the plate dissipation must decrease when the signal is modulated. The circuit should be designed so that the plate dissipation is equal to the allowable value in the *absence* of modulation. Hence equation (13.64) for P_p can be used to compute the plate dissipation.

The parameter γ was defined in equation (12.60) of chapter 12, as follows:

$$\gamma = \frac{2\mu}{\mu + 1} \cdot \frac{P_p}{g_m E_{bb}'^2} \tag{13.65}$$

Substitute equation (13.64) for P_p in this equation, simplify, and express the result in the following form:

$$\gamma = \frac{K}{\alpha + 2} \left(\frac{4}{\pi} - \frac{K\alpha}{\alpha + 2}\right) \tag{13.66}$$

This equation is readily solved for α and the result written

$$\alpha = \frac{4K - \pi K + \sqrt{\pi^2 K^4 + 16K^2 - 8\pi K^3 + 8\pi^2 K^2 \gamma}}{2\pi\gamma} - 2 \tag{13.67}$$

For each value of γ and K there will correspond one value for α. This, in turn, determines R_L from $\alpha = (\mu + 1)R_L/r_p$. Ordinarily K is about 0.45 or 0.50, because the unmodulated carrier is adjusted to encompass about half of the linear region. Everitt (Reference 5) has computed α

as a function of γ for these two cases and the results are reproduced in figure (13.21).

The value of P_p is assumed known, because any given tube will have a specified maximum allowable plate dissipation. Then γ is calculated from equation (13.65) after a power supply voltage is assumed. Figure

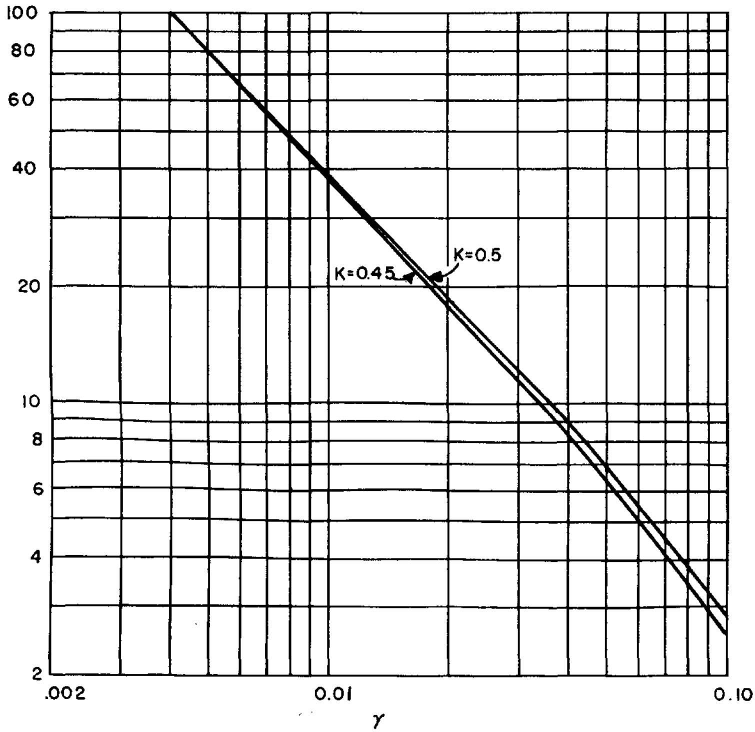


Fig. 13.21. Class **B** amplifier design chart. (From W. L. Everitt, *Communication Engineering*, 2d ed., McGraw-Hill Book Co., Inc., 1937, p. 585.)

(13.21) then determines α and the correct value for R_L for optimum operation. Because γ and α are now fixed, and because $B = \beta = 2$ and $\eta_A = \pi/4$, all the other factors of interest can be computed from the general equations and curves of chapter 12. The results are briefly summarized below.

(1) E_g = grid excitation without modulation

$$= K\left(1 + \frac{2\mu}{\alpha + 2}\right) E_{cc} \tag{13.68}$$

(2) $I_b' =$ d-c component of plate current

$$= K\frac{E_{bb}'}{R_L}\cdot\frac{2\alpha}{\alpha+2} \tag{13.69}$$

(3) $I_{p_1} =$ fundamental component of plate current

$$= \frac{\pi}{2}I_b' \tag{13.70}$$

(4) $I_{\text{emission}} = 2\pi I_b'$ (13.71)

For further discussion of the necessity of using the optimum operating conditions, refer to the original discussion by Everitt (Reference 5).

13.13. Reactance Tube Frequency Modulation

In a frequency modulated transmitter it is necessary to make the frequency deviation proportional to the amplitude of the modulating signal. The reactance tube is one of the simpler methods of accomplishing this result.

Reactance tube circuits are really nothing more than simple class A single stage amplifiers. Feedback through phase shifting circuits is used, together with remote cutoff tubes. Problems on such circuits will be found at the end of chapter 3. Some typical circuit configurations are given in figure (13.22).[6] The approximate values for the input capacitances and inductances are given on the circuits. The computation of these expressions is a straightforward application of the principles outlined in chapter 3.

The significant point to note is that the input capacitances or inductances are either directly or inversely proportional to the mutual transconductance of the tube.

In a remote cutoff tube this transconductance is a function of the grid bias voltage. Hence if the modulating signal is applied to the remote cutoff grid, the effective input inductance or capacitance will vary at the signal frequency.

Now suppose that this circuit is connected directly across the tank circuit of an oscillator as shown in figure (13.23). The oscillator frequency is approximately $\omega_0 = \sqrt{1/L_TC_T}$, where $L_T =$ total plate circuit inductance; $C_T =$ total plate circuit capacitance. Thus if one

[6] See August Hund, *Frequency Modulation*, McGraw-Hill Book Co., Inc., New York, 1942, pp. 155–174.

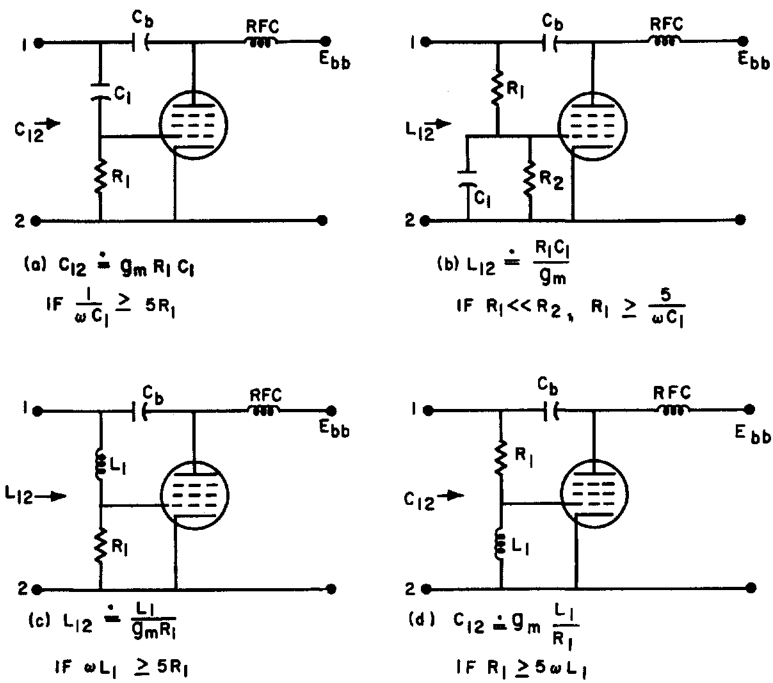


Fig. 13.22. Some reactance tube circuits; can be analyzed from their class A equivalent circuits. See chapter 3.

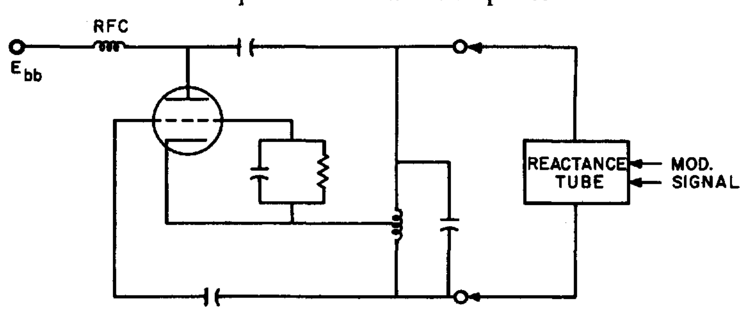


Fig. 13.23. Reactance tube modulation of a Hartley oscillator.

of these two factors is varied at the modulating frequency by the reactance tube, the oscillator frequency will also vary. Ordinarily, small frequency deviations of about 0.2% of the center frequency are used, because the oscillator frequency is linearly dependent upon L_T and C_T only over a confined range near the center frequency. Feedback is generally used to improve the linearity of the frequency-voltage characteristic.

The block diagram of the essential elements of a reactance tube modulated FM transmitter is given in figure (13.24). Such a system is inherently impractical because the center frequency is unstable. This is

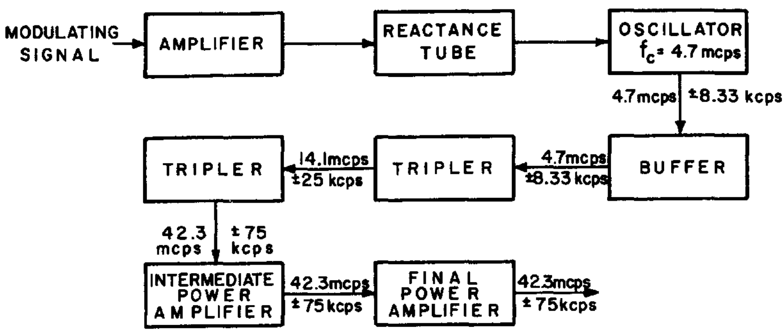


Fig. 13.24. Block diagram of a reactance tube FM transmitter; impractical because of center frequency instability.

directly counter to Federal Communications Commission requirements, which stipulate that the center frequency stability shall be as follows: (1) ± 2 kcps of the assigned center frequency for broadcast transmitters; (2) $\pm 0.02\%$ of the assigned center frequency for mobile communications; (3) $\pm 0.01\%$ of the assigned center frequency for fixed communication stations.

It is impossible to achieve this degree of frequency stability with the system shown in figure (13.24). However, two solutions to this problem are covered in the next section.

13.14. Center Frequency Stabilization[7]

The stabilization of the center frequency of a reactance tube modulated FM transmitter is an important problem, and many solutions

[7] Discussion based upon Milton B. Sleeper, *Standard FM Handbook*, 1st ed., 1946, FM Co., Great Barrington, Mass.

have been developed. Only two representative cases, the Crosby and the Western Electric (or Morrison) systems, are covered here. Both of these systems use circuits such as converters, discriminators, and frequency dividers; these circuits are treated in later chapters. At this point it is necessary only to know that such circuits exist and that they perform the following services:

(1) A converter or mixer translates frequency.

(2) A frequency divider divides frequency by some integral factor.

(3) A discriminator converts frequency variations into variations in direct voltage.

In the Crosby system shown in figure (13.25) a crystal oscillator

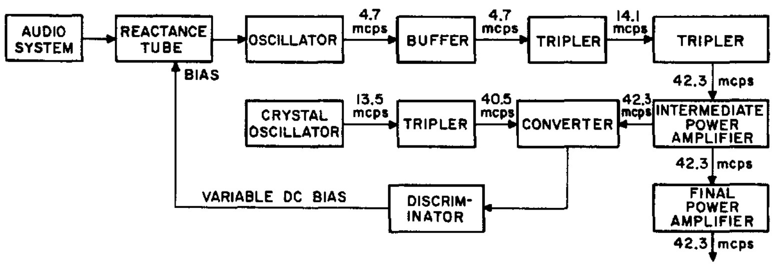


Fig. 13.25. Crosby system of center frequency stabilization in a reactance tube FM transmitter.

operates at a frequency that differs from the correct center frequency of the reactance tube modulated oscillator and by a known amount. This *reference* frequency is tripled and applied to a frequency converter. The output from the main oscillator is tripled twice and also applied to the converter. The *difference frequency* is extracted from the converter and connected to the discriminator. The d-c output from the discriminator, which is proportional to the frequency difference, is used to change the bias on the reactance tube and thereby change the oscillator frequency until the difference frequency from the converter and the discriminator output are zero. Thus the center frequency of the modulated oscillator is held constant by the feedback loop.

The discriminator is made unresponsive to *fast* changes in oscillator frequency caused by modulation. It responds only to fairly slow changes.

In the Western Electric system shown in figure (13.26), the drift of the center frequency is automatically corrected by a small reversible motor that readjusts the settings of variable capacitors in the oscillator tank circuit.

The motor is stationary when the two inputs to the motor control circuit have the same frequency. Under any condition of inequality an output develops, the motor rotates and adjusts the oscillator trimming capacitors to provide the correct center frequency. Large variations in the *amplitude* of the center frequency component caused by changes in the modulation index are removed by the frequency-dividing network and do not affect the operation of the motor control circuits. The inertia and friction of the motor are sufficient to damp any response to

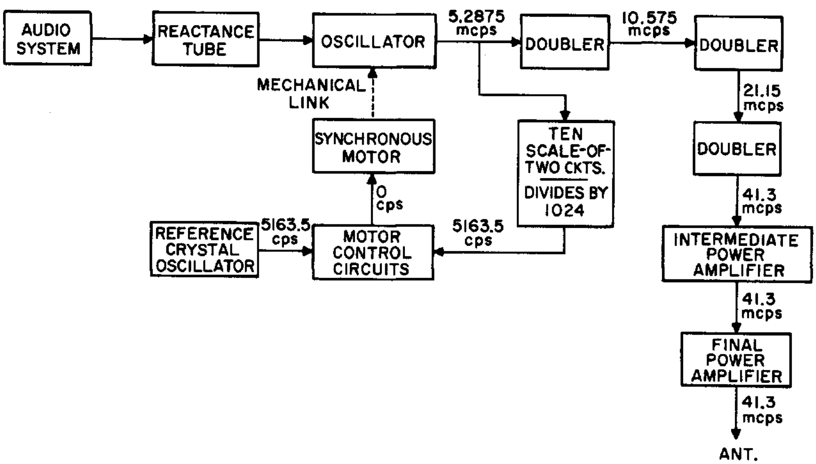


Fig. 13.26. Western Electric system for stabilizing the center frequency in a reactance tube FM transmitter.

control signals arising from frequency variations caused by modulation. The motor responds only to slow variations in frequency.

13.15. The Armstrong System

When a carrier is amplitude modulated by a single frequency it was shown in section (13.1) that the modulated wave has three components: (1) a carrier or center frequency term; (2) two side frequencies. The amplitude of the side frequencies depends upon the modulation index. The sum of the two side frequencies is called the *double sideband* and it is always at 0° time displacement with respect to the carrier.

In the discussion of phase modulation in section (13.4), the following points were noted:

(1) A center frequency term is produced though its amplitude may be zero for certain values of the modulation index.

(2) A number of side frequencies are produced, the number depending upon the modulation index if a single sinusoidal modulating frequency is assumed.

It can be shown that when $m_p = 0.2$ radian, only one pair of side frequencies is large enough to be important. Under this condition, the frequencies of all three terms in the phase modulated output are the

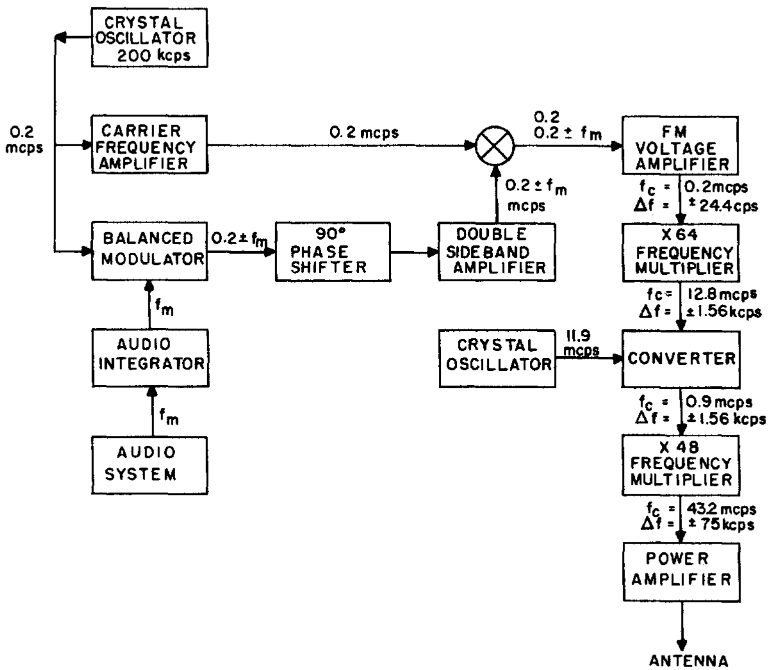


Fig. 13.27. Original Armstrong phase shift modulator and FM transmitter.

same as the three terms produced by amplitude modulation. However, the vector sum of the side frequency components yields a double sideband that is 90° ahead of the center frequency.

Now suppose that we amplitude modulate a carrier with $m_a = 0.201$ and phase modulate a carrier of the same frequency with $m_p = 0.2$. If the modulating frequencies are the same, the outputs from both modulators will have the following terms: (1) a center frequency term at ω_c of relative magnitude 0.99. (2) two side frequencies at $(\omega_c \pm \omega_m)$

of relative magnitude 0.0995. In other words, the two waves are identical except that the double sideband is in phase with the carrier for AM and 90° out of phase for the PM case. This indicates that a wave could be amplitude modulated, the double sideband shifted by 90° and recombined with the carrier to produce a phase modulated carrier. This is the essential principle of the Armstrong modulator.

The output from an Armstrong modulator is phase modulated, but *frequency* modulation is desired for transmission. This is achieved by *integrating* the modulating signal before modulation. Thus phase modulation proportional to the integrated modulated signal is produced, and this corresponds exactly to frequency modulation by the original signal.

With the foregoing points in mind, the block diagram of the original Armstrong phase shift modulator in figure (13.27) is readily understood. A balanced modulator is used to provide a carrier suppressed amplitude modulated output, amplitude modulated by the integral of the original modulating signal. The output from the balanced modulator, which is the double sideband, has its phase shifted by 90° and is then recombined with the original carrier. The resulting wave is phase modulated by the integral of the modulating signal, or frequency modulated by the signal itself.

The modulated signal is passed through a number of frequency multipliers, then through a converter, and finally through more multipliers. This tactic is necessary to achieve the required frequency deviation about the correct center frequency.

13.16. The Phasitron

The *phasitron*[8] is a special phase modulator tube for use in FM transmitters. The construction of the tube is not covered here because the unit is more of a device or component than a circuit.

However, the tube will deliver a phase modulated signal at low distortion with a frequency deviation of 175 c from a 220 kc input. Only 50 mw of audio power are required for modulation and the tube will operate at frequencies up to 500 kc.

The particularly advantageous characteristic of the phasitron is that it permits direct crystal control over the center frequency with comparatively large phase deviations. This reduces the number of frequency multiplications required and further simplifies the transmitter.

[8] General Electric Co., *Application Data, ETX*-109, pp. 1–4, 5–46.

PROBLEMS

13.1. A current $i = 10(1 + 0.8 \cos \omega_m t) \cos \omega_c t$ amp flows through a 72 ohm resistor. Compute the total power dissipated in the resistance. If the coefficient of $\cos \omega_m t$ is zero, what power will be dissipated?

13.2. A modulated carrier current is given by

$$i = 10(1 + 0.8 \cos \omega_m t) \cos \omega_c t$$

where $f_m = 50$ c and $f_c = 1000$ c. Compute and plot the current i as a function of time. Show one complete cycle of the modulating frequency.

13.3. An amplitude modulated wave has the following form:

$$i = 10(1 + 0.4 \cos 628t + 0.6 \cos 1256t + 0.5 \cos 2512t$$
$$+ 0.2 \cos 3140t) \cos (6.28)(10^6)t$$

Make a spectrum plot of the components of the wave, showing the frequency and amplitude of each component,

13.4. In a typical frequency modulation system, the frequency deviation is ± 75 kcps and the modulating frequency ranges from 30 cps to as high as 15 kcps. Make spectrum plots of the transmitter output in these two extreme cases. See footnote reference 2 for values of Bessel functions. Include all terms having values of 0.01 or more.

13.5. From problem (13.4), what do you conclude about (a) the utilization of the allotted bandwidth as a function of the modulating frequency? (b) the bandwidth relative to the frequency deviation?

13.6. A plate modulated amplifier is to be designed using a tube for which the data in problem (12.1) apply. One hundred per cent modulation is desired. Compute all necessary information.

13.7. For the amplifier designed in problem (13.6), compute the carrier power and the sideband power. Where does the sideband power come from? How much voltage is required to produce 100% modulation?

13.8. If the amplifier of problem (13.6) is modulated by a class A power amplifier capable of supplying the necessary power, compute the turns ratio on the modulating transformer if the class A amplifier load is to be 1800 ohms.

13.9. Design a class B power amplifier for optimum operation in a low level modulated transmitter. The data given for the tube in problem (12.1) may be used. Assume the signal input is 100% modulated. Compute all factors of interest.

13.10. Derive the equation for the input inductance of the reactance tube in figure (13.22b). Be sure to state clearly and justify all approximations you use.

13.11. Suppose the reactance tube of problem (13.10) is connected to a Hartley oscillator having a center frequency of 10 mcps when disconnected from the reactance tube. The g_m of the reactance tube can be varied from 250 to 1250 μmhos. A 20 kcps deviation is desired. Determine the values of R_1, R_2, and C_1 required in the reactance tube curcuit.

Chapter 14

RECTIFIERS AND POWER FILTERS

Chapters 12 and 13 treat various functions that can be fulfilled by electronic components operating in the switching mode. The purpose of this chapter is to present an operation most effectively provided by diodes operating as switches.

A diode operating in the switching mode is called a *diode switch* and it might be a vacuum or gas tube or varistor. Diode switches are used to provide two important services, as follows:

(1) Conversion of a-c power to d-c power; in this application the diodes are usually called *rectifiers*.

(2) Conversion of amplitude modulated radio frequency signals to audio or video frequency voltages; these circuits are called *detectors*. Although the theory of operation and method of analysis are the same for both cases, they are covered separately in two chapters. Fundamental principles are largely covered in this chapter, so it is a prerequisite for chapter 15.

Diodes, as well as some multigrid vacuum tubes, can also be used as *mixers* or frequency *converters*. Because such circuits are essentially diode detectors operating under special conditions, they are included in chapter 15 as a logical extension of the discussion of detectors.

The method of analysis follows the same path as that used in other chapters. The equivalent circuits are constructed and the analysis is made by the principles of linear circuit theory.

14.1. The Diode Switch and Equivalent Circuits

In section (1.3) of chapter 1 it was shown that any diode could be represented by the idealized current-voltage characteristic and equivalent circuit shown in figure (14.1). Of course, the values for $E_0 =$ intercept voltage, $r_p =$ forward slope resistance, $R_b =$ reverse slope resistance will be different for different electronic components. In nearly all cases R_b may be assumed infinite (notable exceptions occur when varistors are used in magnetic amplifier circuits). Also, in gas tubes, r_p is virtually zero.

It is clear from figure (14.1) that any diode has the properties of a switch when the voltage applied across it alternates as a function of time. The switch will be in position 1 as long as the voltage across the diode is positive, while it moves to position 2 when this voltage falls below E_0 or becomes negative.

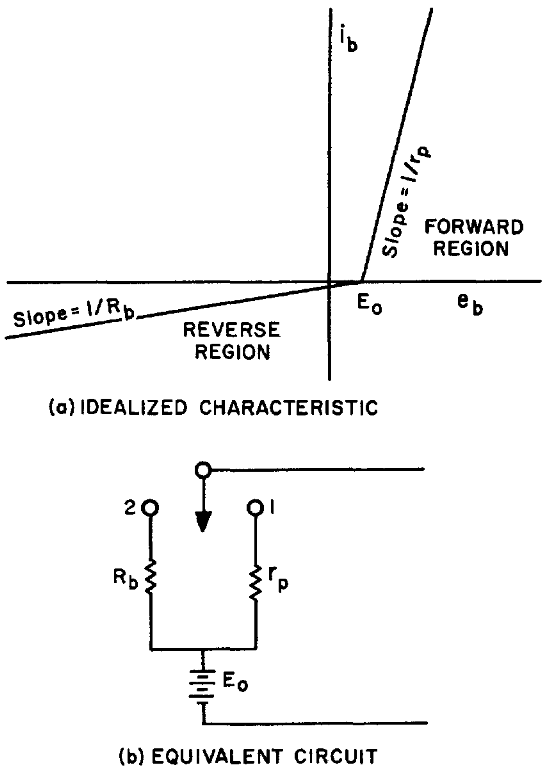
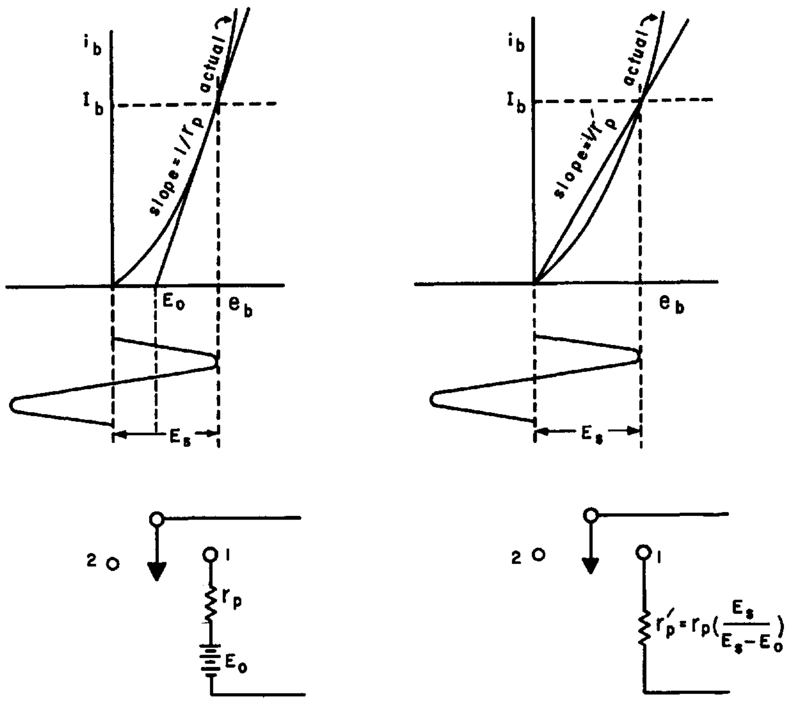


Fig. 14.1. Idealized and equivalent representation of diode characteristics.

The presence of the E_0 term together with r_p in the equivalent circuit hampers the analysis. Indeed, the calculations are inherently so tedious in some cases that it is desirable to simplify the equivalent circuit to the point where it involves just a single resistance. The applied signal in cases of this type is usually sinusoidal, so that the operation of the diode can be shown schematically as in figure (14.2a). This circuit is to be replaced by that shown in figure (14.2b), which also shows the assumed characteristic. Thus a single resistor r_p' is used to replace the series combination of E_0 and r_p. If the peak value of the applied signal voltage is E_s, the two circuits will have essentially the same terminal characteristics if $I_b = E_s/r_p$ in figure (14.2b); $I_b = (E_s - E_0)/r_p$ in figure (14.2a). Hence, for equivalence at I_b,

$$r_p' = r_p \frac{E_s}{E_s - E_0} \tag{14.1}$$



(a) ORIGINAL CHARACTERISTIC AND EQUIVALENT CIRCUIT

(b) PROPOSED ALTERNATE CHARACTERISTIC AND EQUIVALENT CIRCUIT

Fig. 14.2. Relationship between original and alternate equivalent circuit of a diode.

This single resistor equivalent circuit is most advantageously used for vacuum diodes and varistors when the load is a parallel *RC* circuit.

14.2. Rectifier and Filter Circuits, Single Phase Supply

In all circuits in this chapter the diode switch operates with some sort of load circuit connected in series with the cathode or base lead of the electronic device. The simplest such circuit, composed of a resistor in series with a diode, is shown in figure (14.3a). With an alternating voltage applied across the combination, the diode conducts whenever the voltage across it is positive with respect to ground. Conversely, it is an open circuit when this voltage is negative. Thus current flows through the load resistance R_L only during the positive half cycles of the applied

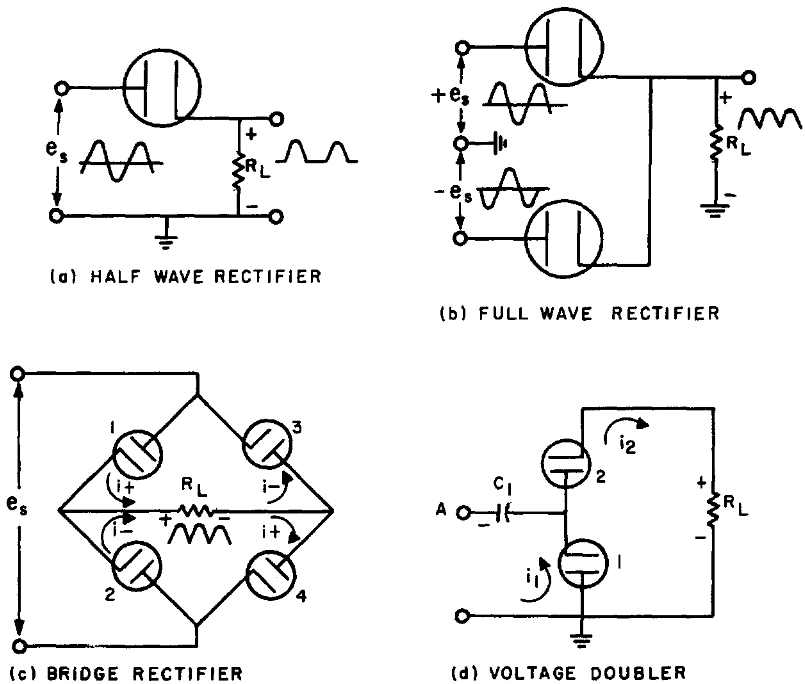


Fig. 14.3. Rectifier circuits.

voltage, and the voltage developed across R_L will consist of a series of half cycles of a sine wave. Such a circuit is called a *half wave rectifier* with resistance load.

If two half wave rectifiers are combined with a common load resistance and the input voltage is split into two voltages 180 degrees out of phase with each other, the output waveform and circuit appear as shown in figure (14.3b). This is called a *full wave* rectifier.

Full wave rectification can be obtained from a single phase supply by using the *bridge rectifier* circuit of figure (14.3c). When terminal A is positive with respect to B, diodes 1 and 4 conduct, thereby providing a closed conducting path through R_L and the supply voltage source. Current flows from left to right through R_L. When terminal A is negative with respect to B, diodes 2 and 3 conduct, providing a closed path with current again flowing from left to right through R_L. Thus a full wave rectified output appears across R_L as shown in figure (14.3c).

Voltage doublers, triplers, and *quadruplers* are other rectifier circuits that find frequent application. A typical doubler circuit is shown in

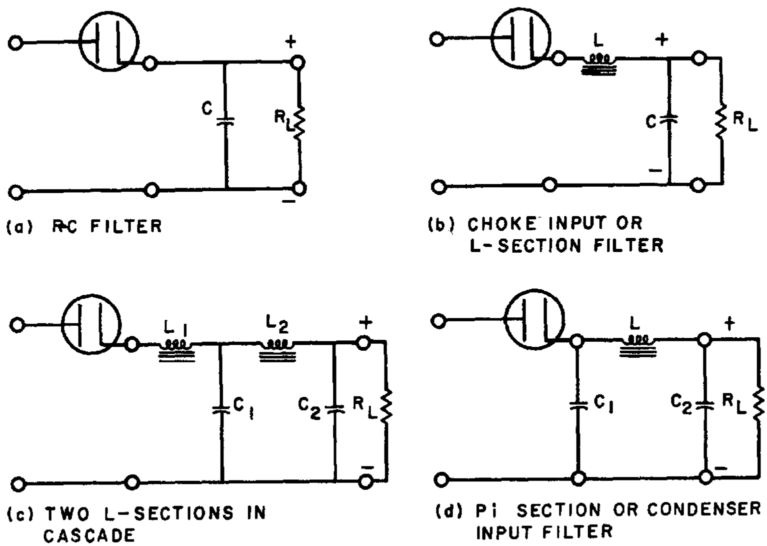


Fig. 14.4. Power filters for use with the rectifiers of figure (14.3).

figure (14.3d). Suppose that terminal A is negative with respect to ground. This causes diode 1 to conduct, and C_1 charges rapidly to nearly the peak value of the input signal voltage, and with the polarity shown. When terminal A is positive with respect to ground, diode 2 conducts. However, the voltage across the diode is now the sum of the voltage across C_1 and the applied voltage. Thus the peak value of the output will be nearly twice the peak value of the input.

Only rarely are any of the foregoing circuits operated with pure resistance loads. The variations in output voltage with respect to time are too large to be tolerated in most applications. Thus *filter* circuits are generally used to smooth the time variations and make the output voltage more nearly time invariant. Typical filter circuits are shown in figure (14.4). They can be used with any of the rectifier circuits previously given, though a half wave rectifier is shown in figure (14.4) for simplicity.

14.3. Rectifier Performance Criteria

The main purpose of a rectifier circuit is to convert a-c power input to d-c power output, as shown diagrammatically in figure (14.5). In such cases, one of the most important factors will be the power transfer function of the rectifier circuit. This is usually called the *rectification efficiency* η_R and is expressed in per cent as

$$\eta_R = \frac{P_{dc}}{P_{ac}} \times 100 \quad \text{per cent} \tag{14.2}$$

The signal input to a rectifier is usually sinusoidal, while the output

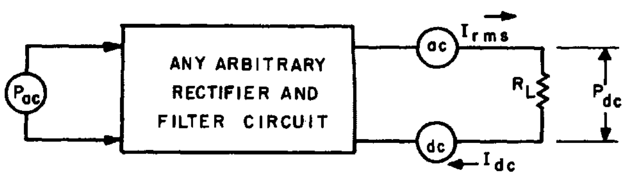


Fig. 14.5. The rectifier as a four terminal network.

is definitely nonsinusoidal. If a Fourier analysis is made of the load current i_L, the result will show the presence of a d-c component, a fundamental, and an infinite series of harmonics. It is desirable to have the d-c term predominate, but some of the a-c components will always be present and will produce fluctuations in the output voltage measured across the load resistance. This fluctuation is called the *ripple.* The amount of ripple is generally expressed by means of a *ripple factor* as follows:

$$\text{ripple factor} = \gamma = \frac{I_{ac}}{I_{dc}} \tag{14.3}$$

where I_{dc} = direct current in the load; I_{ac} = effective value of all alternating components of load current.

If a d-c ammeter is connected in series with the load, it will read the constant term in the Fourier series of the load current. In other words, it will read the *average* value of the load current. This can be expressed mathematically as

$$I_{dc} = \frac{1}{2\pi}\int_0^{2\pi} i_L \, d\omega t \tag{14.4}$$

where i_L = load current expressed as a function of ωt over an interval of 2π radians. This is the defining equation for the constant or d-c term in the Fourier series for a nonsinusoidal waveform.

If an a-c ammeter is connected in series with the load resistance it will read the *effective*, or *root mean square*, value of the total load current. The rms value of a current i_L is defined as

$$I_{rms} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} i_L^2 \, d\omega t} \tag{14.5}$$

Because the load current involves a d-c term as well as a series of alternating terms, the rms current can also be expressed as

$$I_{rms} = \sqrt{I_{ac}^2 + I_{dc}^2} \tag{14.6}$$

Hence the effective value of the a-c components alone is

$$I_{ac} = \sqrt{I_{rms}^2 - I_{dc}^2}$$

The previous general equation for the ripple factor can now be expressed in the following form:

$$\gamma = \text{ripple factor} = \sqrt{\left(\frac{I_{rms}}{I_{dc}}\right)^2 - 1} \tag{14.7}$$

$$\gamma = \sqrt{F^2 - 1}$$

where $F = I_{rms}/I_{dc}$ = *form factor* of the load current. The form factor is simply a useful parameter in the discussion of nonsinusoidal waveforms because it is fairly easily determined experimentally from ammeter readings or theoretically from equations (14.4) and (14.5).

The *terminal* characteristics of a rectifier circuit are also important. That is, the graph showing the load voltage E_{dc} as a function of the direct load current I_{dc} is of considerable importance, because loads on rectifiers are seldom constant. A typical plot, which is called the *regulation characteristic*, is shown in figure (14.6). In most cases, rectifier circuits contain some internal resistance, and this causes the output voltage to decrease as the load current increases. In many cases the output voltage should be independent of load current changes. The departure of the circuit behavior from this proposed ideal is expressed in terms of a percentage factor called the *voltage regulation*. The term is defined as

$$\% \text{ voltage regulation} = \frac{E\,(\text{no load}) - E\,(\text{rated load})}{E\,(\text{rated load})} \times 100$$

$$= \left(\frac{E_{NL}}{E_{RL}} - 1\right) \times 100 \tag{14.8}$$

Current regulation may be defined in a similar manner.

When the diode switch in the rectifier circuit is conducting, there is little voltage across the diode. However, when the diode is not conducting, it is possible that the peak value of the applied alternating voltage plus the load voltage may appear across the diode. Clearly, the diode must be capable of withstanding a large inverse voltage. Thus any given circuit and rectifier combination is characterized by some *peak inverse voltage* (PIV) that the diode must be capable of withstanding. Hence this figure is an aid in selecting the proper diode for use in a given application.

The discussion can be summarized by listing the four performance criteria of particular interest in rectifier design; that is, (1) rectification

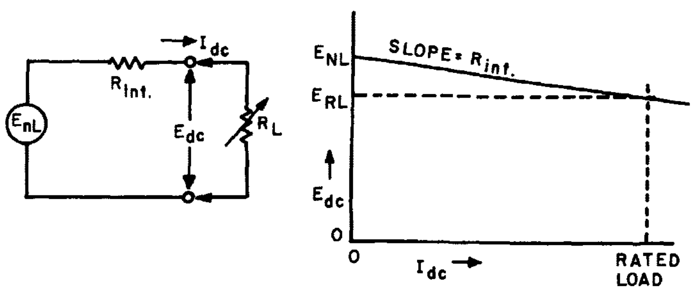


Fig. 14.6. Output or regulation characteristic of a rectifier circuit.

efficiency, (2) voltage regulation, (3) ripple factor, (4) peak inverse voltage.

It seems clear at this point that the discussion of rectifier and filter circuits reduces to the determination of these factors so that all the various circuit combinations can be compared with one another

14.4. Gas Triode Rectifiers

Two types of triode gas tubes, the *thyratron* and *ignitron*, are widely used as rectifiers because the direct voltage developed can be controlled with comparative ease by variations in the control electrode potential or current. Only representative circuits are treated here.

For thyratrons it will be recalled that the operation of each tube type is controlled by its *firing characteristic*. This curve, shown in figure (14.7), gives the value of plate voltage that will make the tube fire for any given value of grid voltage.

Now suppose that a thyratron is connected in series with a load resistance and an alternating voltage source as shown in figure (14.8).

The bias voltage is adjustable as shown. When the plate voltage is negative the tube does not conduct. As the plate voltage swings positive, the tube will not conduct until the plate voltage exceeds the

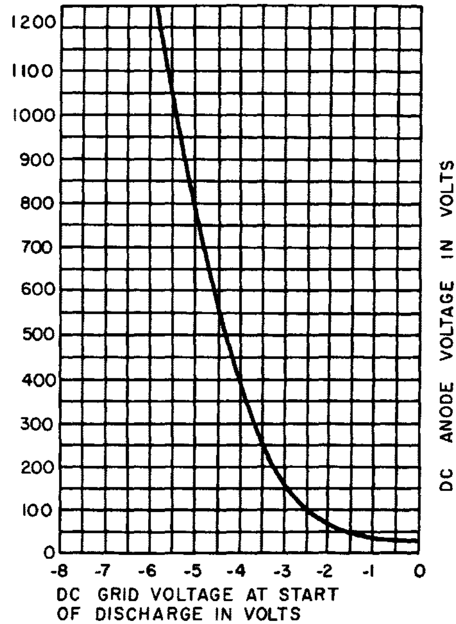


Fig. 14.7. Typical thyratron firing characteristic.

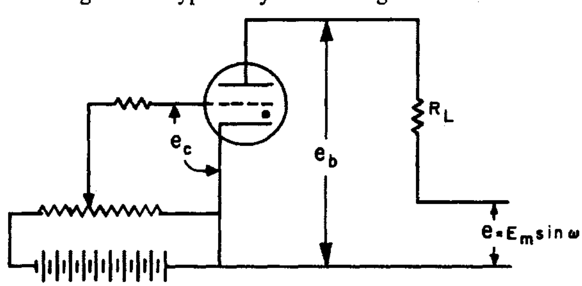


Fig. 14.8. Simple thyratron circuit showing bias control.

critical value corresponding to a particular value of grid bias. As soon as the tube conducts, a heavy current flows and the tube voltage drops to E_0. The grid loses control, and this control is not regained until the plate voltage goes negative long enough for the gas in the tube to deionize.

Clearly, variations in bias will produce variations in the tube conduction angle by changing the firing point θ_f.

During the conduction period the instantaneous plate current is

$$i_b = \frac{E_m \sin \omega t - E_0}{R_L} \tag{14.9}$$

The plate voltage before firing is

$$e_b = E_m \sin \omega t \tag{14.10}$$

The point at which firing occurs is found from the firing characteristic of the tube. If this is designated as E_{bf}, then

$$E_{bf} = E_m \sin \theta_f \tag{14.11}$$

Therefore the firing angle of the tube is

$$\theta_f = \sin^{-1}\left(\frac{E_{bf}}{E_m}\right) \tag{14.12}$$

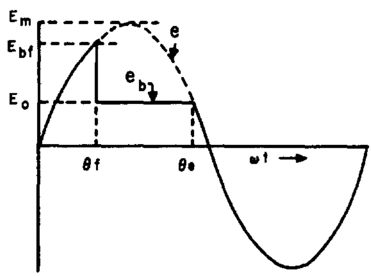


Fig. 14.9. Voltage relationships in the thyratron circuit of figure (14.8).

The tube is extinguished when the alternating voltage falls below the tube drop E_0. Hence the extinction angle θ_e is such that

$$E_0 = E_m \sin (180^\circ - \theta_e) \tag{14.13}$$

or
$$\theta_e = 180^\circ - \sin^{-1}\left(\frac{E_0}{E_m}\right) \tag{14.14}$$

Thus the total angle of tube conduction is

$$\theta_c = \text{conduction angle} = \theta_e - \theta_f \tag{14.15}$$

The average load current through the resistance R_L is now easily computed from

$$I_{dc} = \frac{1}{2\pi} \int_0^{2\pi} i_b \, d(\omega t) \tag{14.16}$$

$$= \frac{1}{2\pi} \int_{\theta_f}^{\theta_e} \frac{E_m \sin \omega t - E_0}{R_L} \, d(\omega t) \tag{14.17}$$

If this integral is evaluated the result can be expressed as

$$I_{dc} = \frac{E_m}{2\pi R_L}\left[\cos \theta_f + \sqrt{1 - \left(\frac{E_0}{E_m}\right)^2} - \frac{E_0}{E_m}(\theta_e - \theta_f)\right] \tag{14.18}$$

When E_0 is much less than E_m, and this is often the case, the direct load current is approximately

$$I_{dc} \doteq \frac{E_m}{2\pi R_L}(1 + \cos\theta_f) \tag{14.19}$$

This last equation shows that the direct component of the load current can be smoothly and continuously varied by changing the firing angle θ_f of the tube. This is easily accomplished by grid bias variations as shown in figure (14.9).

One difficulty with simple bias control is that the firing angle can be varied only over the limited range from 0 to 90 degrees. By using a different method, called *phase control*, the firing angle can be made to have any value from 0° to 180°. To obtain phase control, an alternating

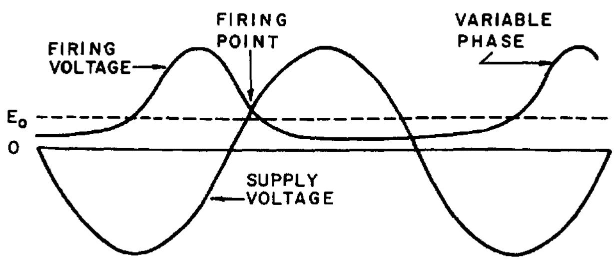


Fig. 14.10. Phase control of a thyratron rectifier.

voltage of the same frequency as the plate supply voltage is applied to the grid of the thyratron. The phase of this grid voltage is adjustable and *lagging* the plate voltage. The sinusoidal variation in grid voltage causes the critical plate voltage to also vary with time. This is shown by the *firing voltage* curve of figure (14.10). The data required to plot this curve were obtained from the firing characteristic of the tube.

The thyratron will conduct only when the plate voltage is equal to the firing voltage. From the sketch in figure (14.10) it is clear that the firing point can be varied over a range of 180 degrees by changing the relative phase of the grid voltage with respect to the plate voltage, because this changes the relative phase of the firing voltage curve.

One of the simplest phase controlled circuits is shown in figure (14.11). The phase angle is variable over a 180° range by changing R and C. In any case, the phase angle is

$$\theta = 2\tan^{-1}\omega RC \tag{14.20}$$

There are many practical variations of this basic circuit.

Thyratrons are generally used to control the firing time of ignitron rectifiers. It will be recalled that the main arc will not start in an ignitron until a cathode spot is formed by the igniter electrode. Thus control of the firing angle of an ignitron is accomplished by firing the igniter circuit at the desired instant. This action is usually accomplished by discharging a fully charged capacitor through the igniter circuit. A thyratron is used as a synchronous switch to control the time of discharge. A representative circuit is shown in figure (14.12).

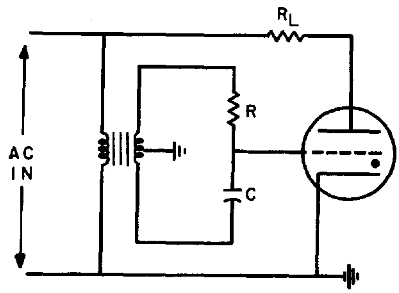


Fig. 14.11. Circuit diagram showing phase control of a thyratron.

In the circuit of figure (14.12), the inductance L is used to limit the thyratron current to a safe value during the capacitor discharge. The phanotron is operated as a half wave rectifier and is used to charge the

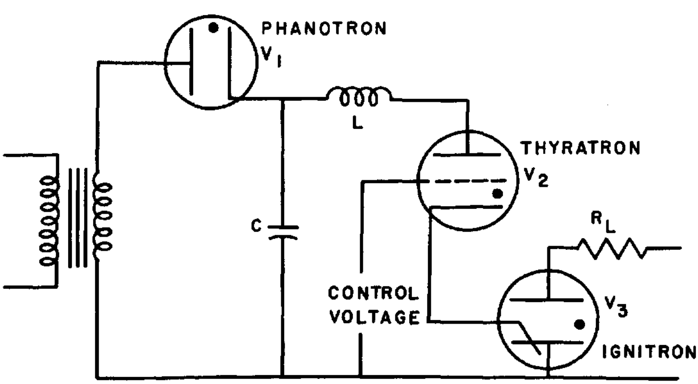
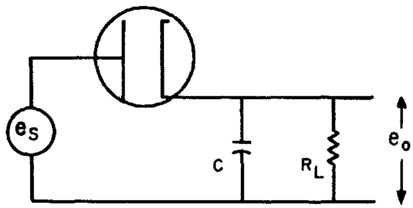


Fig. 14.12. Ignitron rectifier controlled by a thyratron.

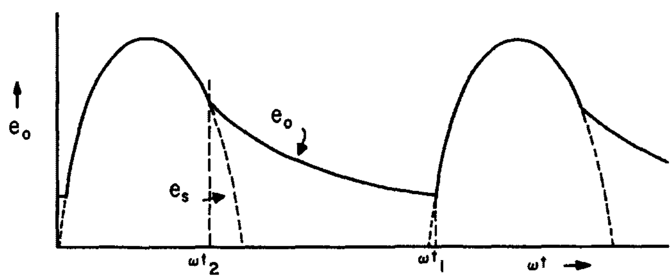
capacitor C. The capacitor discharge occurs when the thyratron conducts and this is determined by the control voltage on the thyratron grid. When the thyratron fires, C discharges quickly through V_2 and L, causing a large current to flow in the igniter circuit, thereby producing the required cathode spot. The thyratron rapidly extinguishes because its plate voltage drops rapidly and the grid regains control.

14.5. Rectifier with *RC* Filter, Introduction

The half wave diode rectifier with *RC* filter is a common electronic circuit and one of the simplest in appearance. The appearance is misleading because it has proved to be an exceptionally difficult circuit to treat analytically. Indeed, as far as I know, there never has been a successful analysis of a practical circuit. Of course, certain hypothetical cases have been treated in nearly every text book, but the results



(a) HALF WAVE RECTIFIER WITH RC FILTER CIRCUIT



(b) WAVEFORM OF THE OUTPUT VOLTAGE

Fig. 14.13. Rectifier with *RC* filter when source and diode resistance are zero.

obtained are of little practical use because the simplifying assumptions are not valid in practical circuits. The difficulty in treating the circuit is not one of technique, because it will be apparent later that the operation of the circuit can be outlined rather accurately. Instead, the problem is primarily one of tedium in making graphical solutions of transcendental equations, performing numerous integrations, and so on.

A simplified circuit diagram of a half wave rectifier with *RC* filter is shown in figure (14.13a). If the diode and source resistances are zero, the waveform of the voltage across the load resistance appears as shown in figure (14.13b). This waveform may be explained rather simply.

Assume that the applied voltage is sinusoidal and of constant amplitude. Whenever the voltage e_b across the diode is positive in the sense indicated, the diode conducts, and the capacitor charges through the diode from the applied voltage source. No transients are involved because the circuit has zero resistance during the charging interval. Therefore the output voltage is exactly equal to the applied voltage during this period.

When the applied voltage begins decreasing, the capacitor voltage can decrease only by releasing some of its charge through the resistor R_L. Eventually a point is reached at ωt_2 where the diode voltage is falling more rapidly than C can discharge through R_L, and the voltage across the diode becomes negative. The diode ceases conduction and the capacitor discharges exponentially through R_L. Later, at $2\pi + \omega t_1$, the voltage across the diode is zero and increasing in the positive direction. The diode again conducts and the entire cycle repeats.

When the assumed conditions of zero diode and source resistance are justified, and this is rare in practice, the circuit can be formulated rather easily from the preceding discussion. Although a transcendental equation results, the circuit design equations can be formulated without too much difficulty.

The problem is greatly complicated in practical cases because the simplifying assumptions are invalid. Suppose that a practical circuit, such as that shown in figure (14.14a) is to be analyzed. The equivalent circuit shown in figure (14.14b) and the corresponding Thevenin equivalent in (14.14c) provide the basis for the analysis. Here R_s is assumed to include the source resistance R_g and the diode resistance r_p'. If the applied voltage is

$$e_s = E_s \sin \omega t \tag{14.21}$$

then the Thevenin generator during the conducting period is

$$e_e = E_s \frac{R_L}{R_L + R_s} \sin \omega t = E_s \frac{R_e}{R_L} \sin \omega t \tag{14.22}$$

$$= E_e \sin \omega t \tag{14.23}$$

Thus, during tube conduction when all transients are virtually complete, the magnitude of the capacitor voltage is

$$E_c = E_e \frac{-jX_c}{R_e - jX_c} \tag{14.24}$$

Hence the instantaneous capacitor voltage is

$$e_c = E_c \sin \omega t \tag{14.25}$$

The current through the load resistance R_L is

$$i_{R_L} = \frac{e_c}{R_L} = \frac{E_c}{R_L} \sin \omega t \tag{14.26}$$

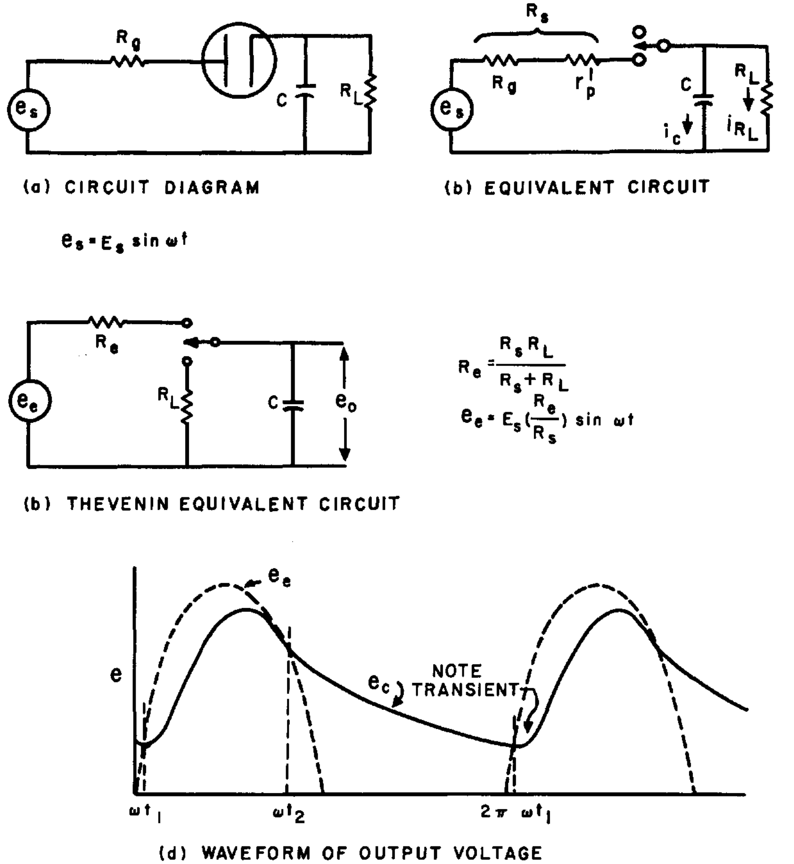


Fig. 14.14. Relationships in a half wave rectifier with *RC* filter when both the source and diode resistance are considered.

The current through the filter capacitor C is

$$i_c = C\frac{de_c}{dt} = \omega C E_c \cos \omega t \tag{14.27}$$

At the instant that the tube becomes nonconducting, as at ωt_2 in figure (14.19d), the tube current

$$i_b = i_c + iR_L \tag{14.28}$$

must become zero, or $-iR_L = i_c$. Hence

$$\frac{1}{R_L}\sin \omega t_2 = -\omega C \cos \omega t_2$$

or

$$\omega t_2 = \tan^{-1}(-\omega R_L C) \tag{14.29}$$

Thus the *cutout angle* ωt_2 is easily computed for any value of the dimensionless constant $\omega R_L C$.

The voltage across the capacitor at the cutout instant is $e_c(\omega t_2) = E_c \sin \omega t_2$, or

$$e_c(\omega t_2) = E_s \frac{R_e}{R_s} \cdot \frac{-jX_c}{R_e - jX_c} \sin \omega t_2$$

During the nonconducting period the capacitor voltage decays exponentially from this value, approaching zero in the limit. Thus, during this period, the capacitor voltage is

$$e_c = e_c(\omega t_2)\varepsilon^{-(t-t_2)/R_L C} \quad \text{for} \quad t > t_2$$

or

$$e_c = E_s \frac{R_e}{R_s} \cdot \frac{-jX_c}{R_e - jX_c} (\sin \omega t_2)\varepsilon^{-(\omega t - \omega t_2)/\omega R_L C} \tag{14.30}$$

Tube conduction commences again when the voltage e_c given by equation (14.30) decays to a point where it is equal to the applied voltage $E_s \sin \omega t$. This occurs at a point designated as $(\omega t_1 + 2\pi)$. At this instant.

$$E_s \sin(\omega t_1 + 2\pi) = E_s \frac{R_e}{R_s} \cdot \frac{jX_c}{R_e - jX_c} (\sin \omega t_2)\varepsilon^{-(\omega t_1 - \omega t_2 + 2\pi)/\omega R_L C}$$

After some algebraic manipulation and rearrangement of terms, this can also be written

$$\sin \omega t_1 = \frac{1}{\omega R_s C} \cdot \frac{\sin \omega t_2}{\sqrt{1+\beta^2}}\,\varepsilon^{-[\omega t_1 - \omega t_2 + (3\pi/2) + \tan^{-1}\beta]/\omega R_L C} \tag{14.31}$$

where

$$\beta = \frac{1 + R_L/R_s}{\omega R_L C} \quad \text{and} \quad \omega t_2 = \tan^{-1}(\omega R_L C)$$

Equation (14.31) is transcendental, and it cannot be solved explicitly for the *cut-in angle* ωt_1. It can be solved graphically, and if this procedure is followed a value will be obtained for each value of the dimensionless parameters $\omega R_L C$ and R_L/R_s. As you might imagine, the procedure is tedious and time consuming, but it can be done. Once the cut-in and cutout angles are known, all factors of interest can be computed in a straightforward manner.

The only assumption made in this development was that all transients were virtually complete after switching in before the switching out time ωt_2 is reached. This approximation is valid except where R_s is large compared with R_L. In these cases the transient current resulting from diode switching *in* must be evaluated, and the analysis is further complicated.

Once the cut-in and cutout angles are determined, the direct output voltage can be computed from

$$E_{dc} = \frac{1}{2\pi}\int_{\omega t_1}^{\omega t_2} e_c \, d(\omega t) + \frac{1}{2\pi}\int_{\omega t_2}^{2\pi+\omega t_1} e_c' \, d(\omega t)$$

where e_c is given by equation (14.25); e_c' is given by equation (14.30). The rms voltage can be computed in a similar manner and the ripple factor evaluated.

A similar analysis can be made for the full wave circuit and for the voltage doubler with *RC* filter.

14.6. Characteristics of Rectifiers with *RC* Filters

It is quite clear from the preceding analysis that it is *possible* to compute the various factors of interest in rectifiers with *RC* filters, but it would be an extremely time consuming proposition. Fortunately, a careful set of experimental determinations has been made by Schade[1] and the results tabulated in a universal system of dimensionless parameters. These charts are reproduced in figures (14.15) through (14.19). Although the curves were obtained experimentally, theoretical calculations following the method given in the preceding section should check these curves within a few per cent, certainly within the tolerance variations allowed the circuit components.

In all the design charts the value of R_s is assumed to include the equivalent diode resistance r_p' and the internal resistance R_g of the generator. Schade used a different scheme for evaluating the diode resistance, but the end result is closely approximated by the value used here.

Aside from the utility of these figures for design purposes, which is treated in the next section, they are informative about certain operating characteristics of the circuit. For example, figure (14.18) shows a plot of the ratio of the peak diode current to the diode direct current as a function of the $\omega R_L C$ and R_s/R_L parameters. For any given value of

[1] O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, July, 1943, pp. 341–361.

R_s/R_L, the peak current increases as $\omega R_L C$ increases. Also, as R_s/R_L decreases, the peak current increases. Clearly, the peak current has a maximum value for any given tube that cannot be safely exceeded. As a result, for any given tube there are two possibilities:

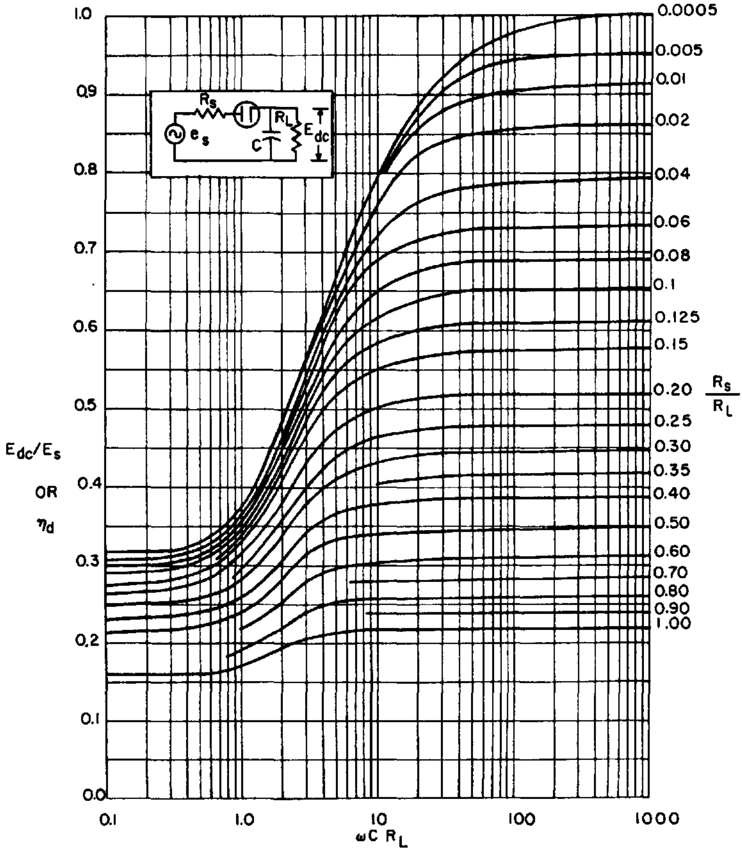


Fig. 14.15. Characteristic of half wave rectifier with *RC* filter. (From O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, no. 7, July, 1943, pp. 341–361.)

(1) For a fixed R_s/R_L there will be a maximum value for $\omega R_L C$ that cannot be exceeded without damaging the tube.

(2) For a fixed value of $\omega R_L C$ there will be a minimum R_s/R_L below which tube damage will result.

Thus in many cases in which the value of $\omega R_L C$ will be fixed by other design considerations, it may be necessary actually to add some resistance to R_g to hold the peak current through the tube within safe limits.

Figure (14.19) shows the ripple voltage as a function of the factor $\omega R_L C$ for various values of the parameter R_s/R_L. Apparently the ripple is not affected much by relatively large changes in R_s/R_L.

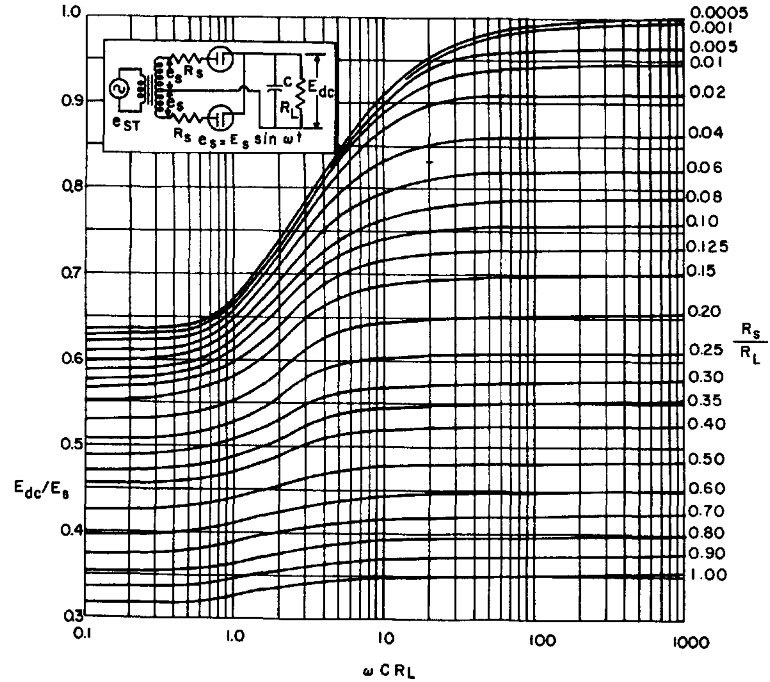


Fig. 14.16. Characteristics of full wave rectifier with *RC* filter. (From O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, no. 7, July, 1943, pp. 341–361.)

On all these graphs the dimensionless parameter $\omega R_L C$ is used as the independent variable. It can be made to change through variations in ω, C, or R_L. However, ω is generally fixed for a given design. Also, certain limiting values of R_L are specified by the design requirements.

Actually, R_L may not appear physically in the circuit, but may simply represent the loading of the circuit. That is, if E_{dc} = direct output voltage, I_L = specified direct load current, then $R_L = E_{dc}/I_L$. Clearly, as the load current changes, R_L will vary and cause $\omega R_L C$ to change.

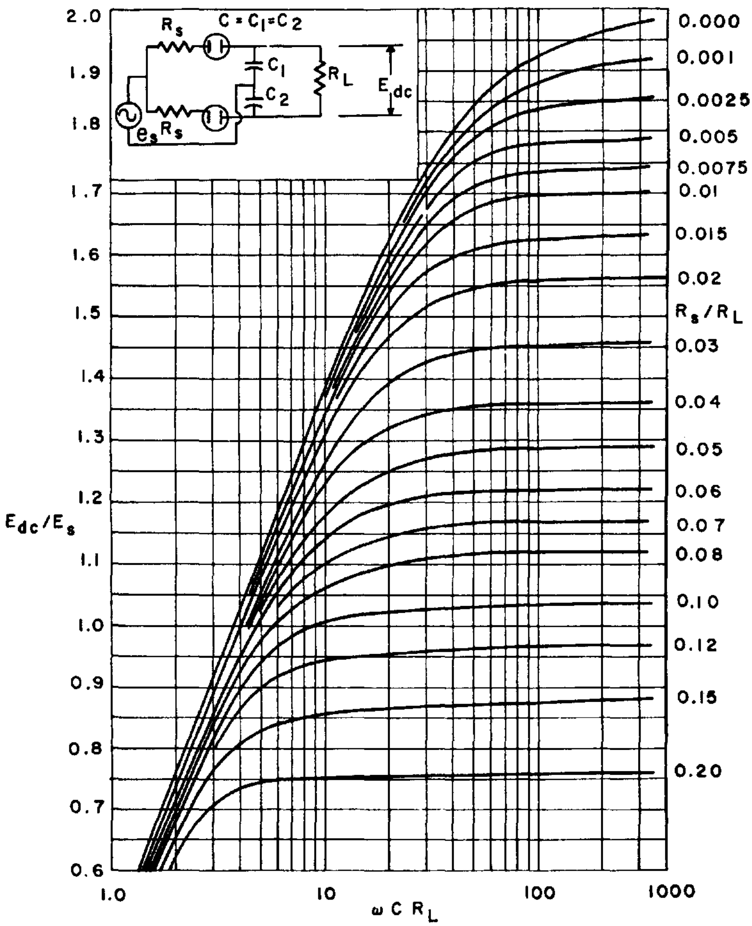


Fig. 14.17. Characteristics of voltage doubler with *RC* filter. (From O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, no. 7, July, 1943, pp. 341–361.)

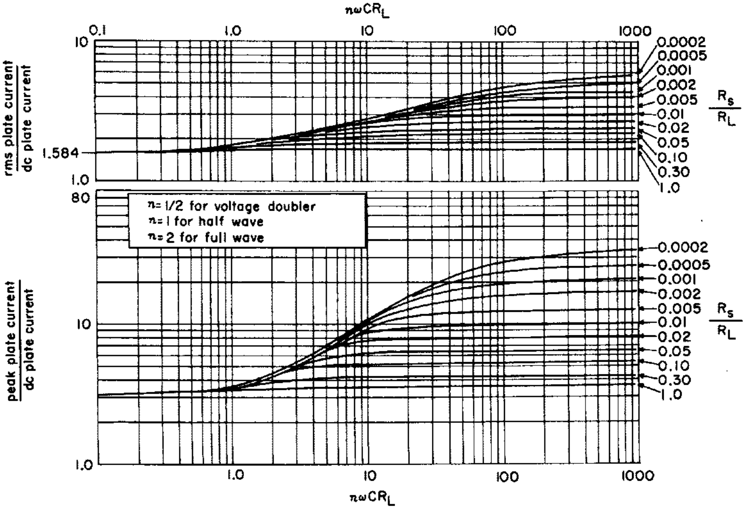


Fig. 14.18. Current relationships for a single plate of a rectifier with *RC* filter. (From O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, no. 7, July, 1943, pp. 341–361.)

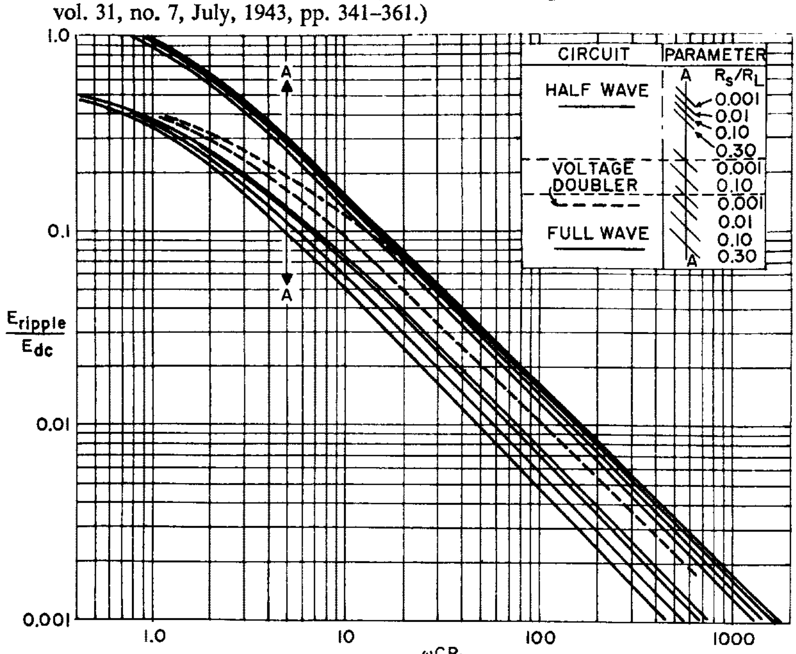


Fig. 14.19. Ripple voltage characteristic of rectifiers with *RC* filters. (From O. H. Schade, "Analysis of Rectifier Operation," *Proc. IRE*, vol. 31, no. 7, July, 1943, pp. 341–361.)

14.7. *RC* Filter Design, Sample Problem

Rectifier and filter design are not clear-cut, straightforward processes yielding a single answer to a single problem. As in any design, there are multitudes of solutions. There may be several that are equally suitable from certain standpoints, but from cost, life, or component availability considerations are not suitable at all. Thus a design is really just an exploration of the various possibilities, and then an evaluation of important parameters is made to restrict the choice of available circuits. Such a process cannot be illustrated well. It is possible only to demonstrate the general methods of using the preceding five sets of graphs.

Suppose that a full wave rectifier with RC filter is to be designed to the following requirements:

$$f = 60 \text{ cps}; \qquad E_{dc}(\text{full load}) = 300 \text{ v}; \qquad \text{ripple} = 1\%$$

$$\omega = 377 \text{ rps}; \qquad I_{dc}(\text{full load}) = 50 \text{ ma}; \qquad \text{volt. reg.} = 20\%$$

The design is essentially a cut-and-try process. First, assume a value for R_s. The value chosen is based upon previous experience with the characteristics of transformers and rectifier diodes. For the purposes of this problem, let $R_s = 500$ ohms.

Because the operating characteristics of the circuit are governed by the ratio of R_s to R_L, assume a value for this ratio under no load conditions. As a first try let $R_s/R_L = 0.01$ at no load. Thus, under no load conditions, the load resistance has its maximum value of

$$R_L\,(\max) = \frac{R_s}{0.01} = \frac{500}{0.01} = 50{,}000 \text{ ohms}$$

From figure (14.19), it can be seen that ripple requirements will be met under this condition if $\omega R_L C = 70$.

At full load the current demand is represented by a resistance R (load) where

$$R\,(\text{load}) = \frac{E_{dc}\,(\text{full load})}{I_{dc}\,(\text{full load})}$$

$$= \frac{300}{50}\,(10{,}000) = 6000 \text{ ohms}$$

Therefore under full load conditions, the total load on the rectifier circuit is the parallel combination of R (load) and R_L (max). Hence

$$R_L\,(\text{min}) = \frac{R\,(\text{load})\; R_L\,(\text{max})}{R\,(\text{load}) + R_L\,(\text{max})}$$

$$= \frac{300}{56}\,(10{,}000) = 5350 \text{ ohms}$$

Therefore, at full load the R_s/R_L parameter has a value

$$\frac{R_s}{R_L\,(\text{min})} = \frac{500}{5350} = 0.0933$$

From this and the 1% ripple requirement, the required value for $\omega R_L C$ can be found from figure (14.19) to be about 60.

The necessary capacitances can now be computed for both the no load and full load cases because the corresponding values of $\omega R_L C$, ω, and R_L are known. Thus:

(1) at no load, $\omega R_L C = 70$, so $C\,(\text{min}) = 3.72\ \mu\text{f}$

(2) at full load, $\omega R_L C = 60$, so $C\,(\text{max}) = 29.7\ \mu\text{f}$

To make sure that the ripple does not exceed 1% under all conditions of load, the filter capacitor should have a value not less than 29.7 μf. For practical convenience, use two parallel 16 μf sections for a total capacitance of 32 μf. Using this value of C, the value of $\omega R_L C$ can be computed for each of the two cases and the direct load voltage and peak plate current can then be obtained from figures (14.16) and (14.18). The following results are obtained:

(1) at no load

$$\omega R_L C = 603; \quad \frac{R_s}{R_L} = 0.01; \quad \frac{E_{dc}}{E_s} = 0.945; \quad \frac{I_{\text{peak}}}{I_{dc}} = 10$$

(2) at full load

$$\omega R_L C = 64.5; \quad \frac{R_s}{R_L} = 0.093; \quad \frac{E_{dc}}{E_s} = 0.770; \quad \frac{I_{\text{peak}}}{I_{dc}} = 5.2$$

Now see if this design satisfies the voltage regulation requirements

$$\%\ \text{volt. reg.} = \left(\frac{E_{NL}}{E_{FL}} - 1\right) \times 100$$

$$= \left(\frac{0.945}{0.770} - 1\right) 100\%$$

$$= (1.23 - 1)\, 100\% = 23\%$$

The 20% regulation requirement is not met, so a redesign is required.

For the second design assume that $R_s = 500$ ohms as before, but that $R_s/R_L = 0.05$ at no load. Therefore

$$R_L\,(\text{max}) = \frac{500}{0.05} = 10{,}000 \text{ ohms}$$

$$R_L\,(\text{min}) = \frac{(10)\,(6)}{10+6}\,(1000) = \frac{60}{16}\,(1000)$$

$$= 3750 \text{ ohms}$$

Thus to satisfy the 1% ripple requirements under all load conditions, from figure (14.19),

(1) at no load, $\omega R_L C = 65$, so $C\,(\text{min}) = 17.3\ \mu\text{f}$

(2) at full load, $\omega R_L C = 55$, so $C\,(\text{max}) = 38.9\ \mu\text{f}$

For practical convenience, let $C = 40\ \mu$f. Therefore

(1) at no load

$$\omega R_L C = 150; \quad \frac{R_s}{R_L} = 0.05; \quad \frac{E_{dc}}{E_s} = 0.94; \quad \frac{I_{\text{peak}}}{I_{dc}} = 6.5$$

(2) at full load

$$\omega R_L C = 56.5; \quad \frac{R_s}{R_L} = 0.133; \quad \frac{E_{dc}}{E_s} = 0.725; \quad \frac{I_{\text{peak}}}{I_{dc}} = 5.0$$

Therefore

$$\%\ \text{volt. reg.} = \left(\frac{0.840}{0.725} - 1\right) 100\% = 15.6\%$$

Evidently this design will satisfy all requirements, though we have no assurance that it is the *optimum* design for this problem.

The peak plate currents can be computed from the ratios given in the two cases. That is,

(1) at no load

$$I_{dc} = \frac{E_{dc}\,(\text{full load})(1 + \text{volt. reg.})}{R_L\,(\text{max})} = 34.7 \quad \text{ma}$$

$$= \frac{(300)(1.156)}{10{,}000}$$

$$I_{\text{peak}} = 6.5\, I_{dc} = 226 \quad \text{ma}$$

(2) at full load

$$I_{dc} = \frac{E_{dc}\,(\text{full load})}{R_L\,(\text{min})} = \frac{300}{3750} = 80 \quad \text{ma}$$

$$I_{\text{peak}} = 5.0\, I_{dc} = 400 \quad \text{ma}$$

Hence the rectifier element must be capable of supplying a peak plate current of 400 ma and a direct plate current of 80 ma. This will assist in selecting the rectifier.

The maximum inverse voltage on the rectifier is

$$\begin{aligned} \text{P.I.V.} &= E_{dc}\,(\text{max}) + E_s = E_{dc}\,(\text{full load})(1 + \text{reg}) + E_s \\ &= E_{dc}\,(\text{full load})(1 + \text{reg}) + E_{dc}\,(\text{full load})\frac{1}{0.725} \\ &= 761 \quad \text{volts} \end{aligned}$$

The required secondary voltage on the transformer is

$$E_s = \frac{E_{dc}}{0.725} = \frac{300}{0.725} = 441 \text{ volts}$$

Furthermore R_s was assumed to be 500 ohms, so that

$$R_s = R_g + r_p' = R_g + r_p \frac{E_s}{E_s - E_0}$$

The tube must be selected so that the values of r_p and E_0 together with the transformer constants R_g and E_s give the required value for R_s. Otherwise, series resistance must be added or different tubes or transformers must be selected.

14.8. Rectifier with *L*-Section Filter

The rectifier with *RC* filter suffers from two notable disadvantages:

(1) Excessive peak current demand if the ripple is low.

(2) Poor regulation characteristics.

There are a number of possible filter circuits using inductances that might be considered in an effort to achieve improved operation. Several of these possibilities are shown in figure (14.20). The circuit using a series inductor is easily analyzed and is not treated here. The half wave rectifier with *L*-section filter is mainly of academic interest. Thus, interest here centers upon the full wave rectifier with *L*-section filter because it has a wide area of practical application.

For the unloaded condition of operation of the full wave rectifier with *L*-section filter, R_L is infinite. Thus the filter capacitor will charge up to the peak value of the input voltage. Because the applied voltage is $E_m \sin \omega t$, the no load output voltage will be E_m.

As the load resistance decreases so that load current flows, each diode is alternately connected to the applied voltage for a few moments, and the capacitor charges to E_m. Between these short conduction periods, the capacitor discharges somewhat through R_L. Thus the

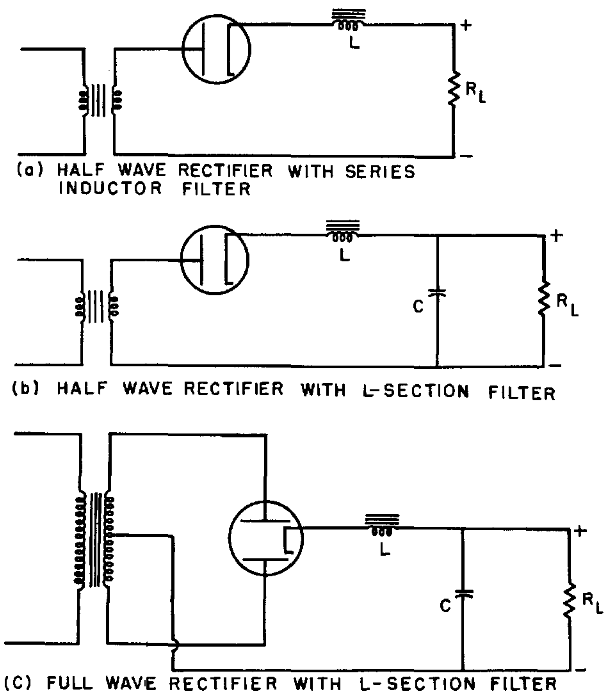


Fig. 14.20. Some rectifier and filter circuits using inductances.

average value of the output voltage will be somewhat less than E_m. Because the current is quite small, the energy stored in the filter inductance is small and the inductor has little effect upon the operation of the circuit. Hence, for small load currents, the circuit behaves essentially like a full wave rectifier with RC filter; this accounts for the appearance of the output characteristic in the low current region as shown in figure (14.21).

As the load current is gradually increased by decreasing R_L, the conduction angles of the diodes lengthen. Eventually each tube conducts for exactly 180°. This situation exists at the point I_k on figure (14.21). At this point the voltage applied to the filter section consists of positive half cycles of the applied sine wave as shown in figure (14.22) and the current through the inductor becomes continuous for the first time. If

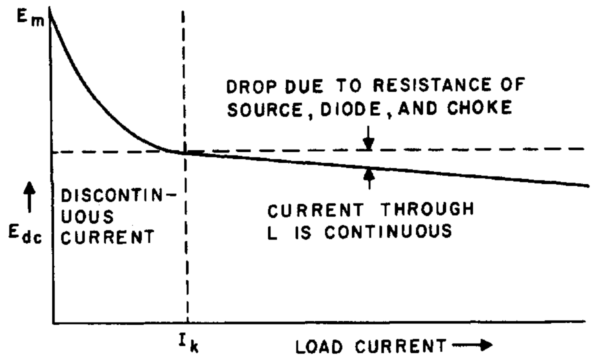


Fig. 14.21. Terminal characteristic for a full wave rectifier with L-section filter.

the load current is increased beyond I_k, the current is large enough so that the magnetic energy storage in the inductor has an appreciable effect upon the operation of the circuit. Thus, as the load current

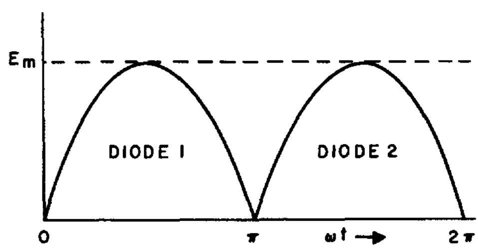


Fig. 14.22. Voltage supplied to the filter by a full wave rectifier when each tube conducts for 180°.

increases, each tube conducts for one half cycle, and the current through the inductor is never zero. Under such conditions the diodes operate as synchronous switches and the load current commutates smoothly between the two tubes. For load currents in excess of I_k, the drop in load voltage is largely determined by the source resistance, diode resistance, and resistance of the filter inductor.

The voltage applied to the circuit for load currents equal to or greater than I_k has the form shown in figure (14.22). The Fourier series of this waveform was given previously as

$$e = \frac{E_m}{\pi}\left(2 - \frac{4}{3}\cos 2\omega t - \frac{4}{15}\cos 4\omega t - \dots\right) \tag{14.32}$$

The ripple components in the output are clearly the even harmonics. The constants of the filter are selected to make these terms small. Therefore L should have a high reactance at the second harmonic frequency, while the reactance of C should be small. If these conditions prevail at the second harmonic, they will be even more extreme at the fourth harmonic. Thus all ripple components except the second harmonic will be neglected.

Because C is to have negligible reactance at the second harmonic, then

$$R_L \gg \frac{1}{2\omega C}$$

The input impedance of the filter at the second harmonic frequency is

$$Z_{in} = 2\,j\omega L - \frac{jR_L/2\omega C}{R_L - j/2\omega C} \tag{14.33}$$

When the preceding inequality is used, the input impedance is approximately

$$Z_{in} \doteq 2\,j\omega L - \frac{j}{2\omega C} = j\left(2\omega L - \frac{1}{2\omega C}\right)$$

$$\doteq \frac{j}{2\omega C}(4\omega^2 LC - 1) \tag{14.34}$$

Because the magnitude of the second harmonic voltage is

$$E_2 = \frac{4E_m}{3\pi} \tag{14.35}$$

the magnitude of the coil current is

$$I_{L_2} = \frac{E_2}{Z_{in}} = \frac{4E_m}{3\pi}\cdot\frac{2\omega C}{4\omega^2 LC - 1} \tag{14.36}$$

This current divides through R_L and C and the load current is

$$I_{R_{L_2}} = I_{L_2}\frac{1/2\,j\omega C}{R_L + 1/2\,j\omega C} \doteq I_{L_2}\frac{1}{2\,j\omega R_L C} \tag{14.37}$$

$$\doteq \frac{4E_m}{3\pi R_L}\cdot\frac{1}{4\omega^2 LC - 1} \tag{14.38}$$

Thus the rms ripple current is

$$I_{R_{L_2}}(\text{rms}) = \frac{4E_m}{\sqrt{2}\,3\pi R_L} \cdot \frac{1}{4\omega^2 LC - 1} \tag{14.39}$$

The direct component of current is

$$I_{dc} = \frac{2E_m}{\pi R_L} \tag{14.40}$$

Hence the ripple factor is easily computed from

$$\gamma = \frac{I_{ac}}{I_{dc}} = \frac{I_{R_{L_2}}}{I_{dc}}$$

or

$$\gamma = \frac{2}{3\sqrt{2}} \cdot \frac{1}{4\omega^2 LC - 1} = \frac{0.471}{4\omega^2 LC - 1} \tag{14.41}$$

This simple relationship for the ripple factor was obtained on the

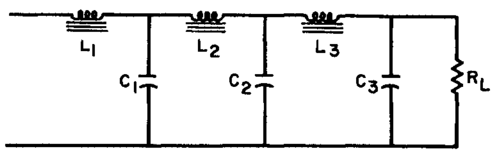


Fig. 14.23. Cascade of L-section filters.

strength of two assumptions: (1) continuous current flow through L, or a load current greater than I_k; (2) R_L is much larger than the reactance of C at the second harmonic frequency. As long as these conditions are valid the ripple factor is independent of the load because R_L does not enter into equation (14.41).

If several L-sections are connected in tandem as shown in figure (14.23), and if the reactance of each filter capacitor is much less than the resistance of R_L at the second harmonic frequency, it is fairly easy to show that the ripple factor has the general form

$$\gamma = \frac{0.471}{(4\omega^2 L_1 C_1 - 1)(4\omega^2 L_2 C_2 - 1)\ldots(4\omega^2 L_n C_n - 1)} \tag{14.42}$$

14.9. Critical Values for L-Section Filter Components

The current-voltage characteristic for a full wave rectifier with L-section filter was shown in figure (14.21). Because of the rapidly changing voltage at low load currents, in the interests of good regulation it is desirable to maintain current through the choke coil at a level equal to or greater than I_k.

In the analysis of the filter circuit it was assumed that the circuit constants were adjusted so that only the d-c component and second harmonic need be considered. Thus the total current through the inductance L will be the sum of these two terms or $i_L = I_{dc} + I_2 \cos 2\omega t$. This current will become continuous when the peak value of the second harmonic component is just equal to I_{dc} as shown in figure (14.24). Thus, at the critical value of current, $I_{dc} = I_2$.

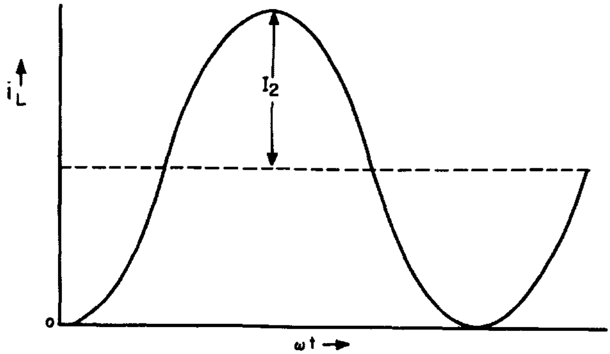


Fig. 14.24. Current through the choke at the point where the current flow becomes continuous.

The direct and second harmonic voltages were previously given as

$$E_{dc} = \frac{2E_m}{\pi}; \qquad E_2 = \frac{4E_m}{3\pi}$$

The input impedance of the filter to direct current is R_L, while its value at the second harmonic is approximately

$$Z_{in}(2\omega) \doteq \frac{j}{2\omega C}(4\omega^2 LC - 1) \tag{14.43}$$

Thus the magnitudes of the two currents in the inductor are

$$I_{dc} = \frac{2E_m}{\pi R_L} \tag{14.44}$$

$$I_2 \doteq \frac{8\omega C E_m}{3\pi(4\omega^2 LC - 1)} \tag{14.45}$$

When the direct load current is I_k, so that $I_{dc} = I_k$, this load current corresponds to a particular value of load resistance. Solve the equality for R_L and designate it R_k. That is,

$$R_k \doteq \frac{3(4\omega^2 LC - 1)}{4\omega C} \tag{14.46}$$

Solve for C and write

$$C \doteq \frac{3(4\omega^2LC - 1)}{4\omega R_k} \tag{14.47}$$

However, the ripple factor was shown to be

$$\gamma \doteq \frac{0.471}{4\omega^2LC - 1}$$

or

$$(4\omega^2LC - 1) = \frac{0.471}{\gamma} \tag{14.48}$$

The equation for the filter capacitance is therefore

$$C \doteq \frac{3(0.471)}{4\omega\gamma R_k} = \frac{0.0562}{f\gamma R_k} \tag{14.49}$$

where $f =$ frequency in cycles per second. Now solve equation (14.48) for the filter inductance L. The result is

$$L \doteq \frac{0.471 + \gamma}{4\omega^2C\gamma} = \frac{R_k}{3\omega}\left(1 + \frac{\pi}{0.471}\right) \tag{14.50}$$

Finally,

$$R_k = \frac{E_{dc}}{I_k} \tag{14.51}$$

and

$$E_m = \frac{\pi}{2}E_{dc} \tag{14.52}$$

These last four equations are the desired design relationships.

It seems clear that there will be an arbitrariness in the design unless some design factor other than the supply frequency, ripple factor, and direct load voltage is specified. As long as this arbitrariness exists, the designer is free to make an initial selection of either R_k, C or L. Once the selection is made the design formulas lead to specific values for the other elements.

The resistor R_k is called the *bleeder* resistance. The power loss in the bleeder should be kept as small as possible. Thus R_k should be large. However, from equation (14.50) it is clear that the filter inductance must be increased in direct proportion to the bleeder resistance. This raises the cost considerably. One way of solving the problem is to use a *swinging choke*. This type of inductance is designed to have a variable inductance depending upon the current flowing through the coil. The effect is produced by operating the choke core near saturation at full load so that the coil is not saturated at light loads and the inductance is much larger. Thus the inductance of the choke *swings* with the load current and can, by proper design, be adjusted to assure a continuous current flow through L even at very low current levels.

Although the derivations in this section neglected the diode resistance, Schade's work (Reference 1) indicates that this does not materially affect the results.

14.10. Pi-Section Filter

In many cases a capacitor is added to the L-section filter as shown in figure (14.25). The resulting circuit is called a pi-section filter and finds frequent application with full wave rectifier circuits.

The mathematical formulation of the circuit is exceedingly difficult and will not be attempted here. The circuit operation may be understood in a qualitative way. When one of the diodes in figure (14.25)

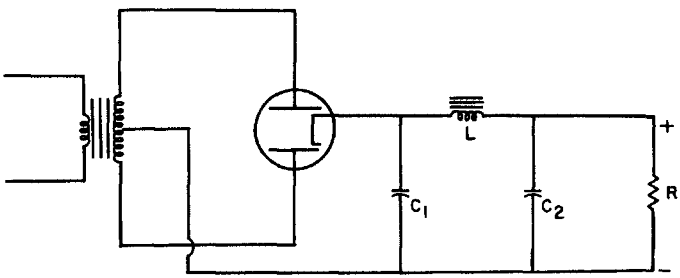


Fig. 14.25. Full wave rectifier with pi-section filter.

conducts, capacitor C_1 rapidly charges to a voltage E_m before being disconnected by the diode. Then C_1 discharges through L and the parallel combination of R_L and C_2 until the other diode conducts.

Relative to the L-section filter, this circuit has a lower ripple factor, higher output voltage, and a poorer voltage regulation characteristic. High peak currents flow through the diodes when the circuit is first turned on, because C_1 is a virtual short circuit. This can cause serious tube damage in some cases, especially when gas tubes are used.

The circuit is generally designed by treating it as a two-section filter. The first section is taken to be R_LC_1 and is designed as a standard RC filter. The second section is taken to be LC_2 and is designed as an L-section filter. In other words, figure (14.19) is used to determine the ripple from the first section and equation (14.41) is used to find the ripple from the second section. The procedure is fairly accurate as long as the ripple from the first section does not exceed 10%.

14.11. Voltage Regulated Power Supplies

It frequently develops that the rectifier and filter circuits discussed in the preceding sections cannot provide the required degree of voltage regulation and ripple without exceeding cost or size limitations. Other devices and circuits are then combined with rectifier and filter circuits to improve the over-all characteristics of the system. The system is then called a *voltage regulated* power supply. Similar techniques can be used to produce *current regulated* power supplies.

The simplest type of voltage regulation of a power supply is achieved through the use of glow discharge tubes. A typical circuit is shown in figure (14.26). The voltage regulator tube is selected so that the voltage drop across it is equal to the desired direct output voltage. By combining several such tubes of various ratings in series, many direct voltages are available.

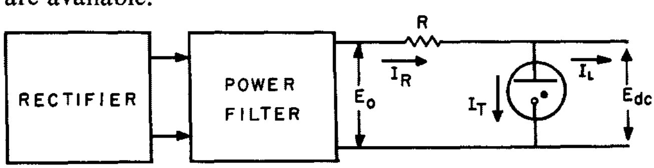


Fig. 14.26. Use of a glow discharge tube in a voltage regulated power supply.

The current limiting resistor R shown in figure (14.26) must be selected so that the current through the glow tube never exceeds the value specified by the manufacturer. Clearly, maximum current will flow through this tube when the load current is zero. Then if $I_T(\text{max}) =$ max. rated tube current, the value of R must be such that $E_0 = I_T(\text{max}) R + E_b$, where $E_b =$ rated voltage drop across the glow tube. Hence $R = (E_0 - E_b)/I_T(\text{max})$. If the value of E_0 is subject to variation because of line voltage changes or any other foreseeable factor, the fact should be recognized and the value of R adjusted accordingly.

Vacuum tubes are widely used in regulating circuits. When so employed they generally provide two functions: (1) class A amplification; (2) act as a variable resistance to direct current. One tube is usually inserted in series with the d-c lead from the power supply, and this tube is used as a variable resistance. For example, in a typical circuit such as that shown in figure (14.27), V_1 is used as the variable resistance. The value of the tube resistance changes with the negative grid voltage, increasing as the grid voltage is made more negative.

A class A amplifier V_2 is connected into the circuit so that the signal input to the amplifier is a fraction of the variation in direct output voltage. The output from the amplifier tube changes the bias and resistance of the regulator tube V_1. The glow tube V_3 and its associated circuit are used to provide the proper bias for the amplifier tube.

From the circuit diagram it is clear that R_1, R_2, R_4, and R_5 are adjusted to provide the proper bias and signal input to V_2 and R_3 is the current limiting resistor for the glow tube.

The operation of the circuit is readily understood. Suppose that E_{dc} momentarily increases because of an increase in rectifier output or because of the ripple component. A fraction of this change is coupled

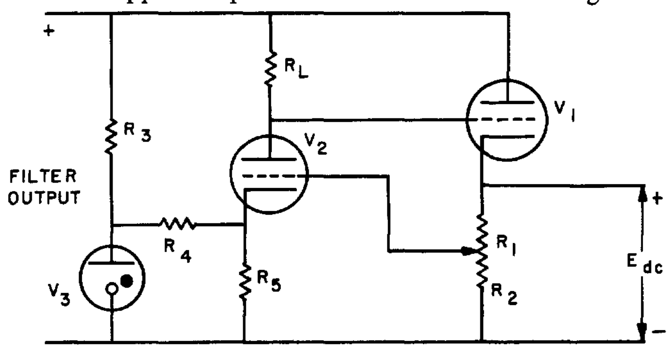


Fig. 14.27. Voltage regulator using vacuum tubes.

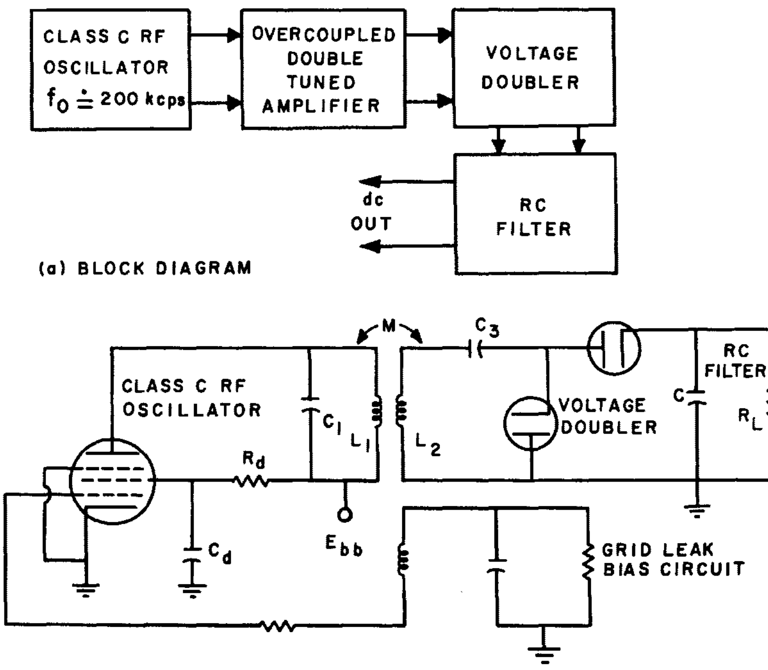
into the amplifier, and if it is an increase in voltage, causes the plate voltage of the amplifier tube to drop. This drop reduces the grid voltage on the regulator tube V_1, increasing its resistance, increasing the drop across the tube, and thereby holding the output voltage nearly constant. The plate load resistance R_L of the amplifier tube is determined partly by the gain requirements for the amplifier and partly from bias considerations of the regulating tube.

The necessary design formulas can be derived using the methods outlined for class A circuits in Part II.

14.12. Radio Frequency Power Supplies

There are many applications in which high voltage, low current power supplies are required. A typical case is the power supply for the second anode of a cathode ray tube. The transformers required to step up a 60 cps 115 v line to the requisite levels are expensive. Also, the cost of iron core inductors is higher at low frequencies than at high frequencies. Thus in some cases it may be more economical to develop an RF voltage to supply a rectifier and filter circuit, because inexpensive air core transformers can be used at a saving in cost that offsets the cost of the RF oscillator circuit.

A typical RF power supply can deliver from 5000 to 30,000 volts



(b) CIRCUIT DIAGRAM

Fig. 14.28. Typical radio frequency power supply.

with a power rating of 2 to 20 watts.[2] Most of this power is dissipated in a bleeder resistance.

The block diagram and circuit diagram of a typical radio frequency power supply are given in figure (14.28). The general nature of the circuit

[2] O. H. Schade, "Radio-Frequency Operated High Voltage Supplies for Cathode Ray Tubes," *Proc. IRE,* vol. 31, April, 1943, pp. 158–163; R. S. Mautner and O. H. Schade, "Television High Voltage RF Supplies," *RCA Rev.*, vol. 8, March, 1947, pp. 43–81; H. C. Bauman, "Television Receiver R.F. Power Supply Design," *Comm.*, vol. 26, March, 1946, pp. 26–27; George W. C. Mathers, "Some Additions to the Theory of Radio Frequency High Voltage Supplies," *Proc. IRE,* vol. 37, February, 1949, pp. 199–206.

operation is self-evident and does not require further explanation. The design of the voltage doubler *RC* filter combination is standard and the material previously presented can be used. The main factor is then the design of the class C oscillator. This is performed by the methods given in chapter 12. It is usually assumed, and the accuracy is quite good, that the coefficient of coupling is so small that load variations in the rectifier circuit are not reflected into the plate circuit of the oscillator.

An analysis based upon equivalent circuits has been made and interested readers should consult Mathers, Reference 2.

Another slightly related circuit is the *flyback* type of high voltage power supply in general use in television receivers having magnetically deflected cathode ray tubes. In this circuit another winding is placed on the horizontal deflecting coil. The current through the deflecting coil is a 15,750 cps sawtooth. The rapid change in current during the flyback period (see chap. 17) causes the added winding to oscillate at its own natural frequency of about 100 kcps. The oscillation is damped out at the end of the first half cycle by a damper tube. The 100 kcps pulses are applied to a diode rectifier with *RC* filter, and a high voltage is thereby secured.

PROBLEMS

14.1. A half wave rectifier with resistance load uses a diode for which $r_p' = 100$ ohms. The input voltage is $250 \sin 377t$ and $R_L = 2000$ ohms. Compute I_{dc}, E_{dc}, rectification efficiency, ripple factor, and the power lost in the tube.

14.2. Repeat problem (14.1) for a full wave rectifier.

14.3. A full wave rectifier with *RC* filter is to be designed to have the following characteristics: E_{dc} (full load) = 250 v; I_{dc} (full load) = 50 ma; ripple = 2%; volt. reg. = 15%. The supply frequency is 60 c. Design the circuit and discuss the possible choices of rectifier elements and the properties required of the transformer assuming a 110 v supply.

14.4. Design a voltage doubler to the same requirements as those given in problem (14.3). Discuss this circuit, relative to the full wave design, from the standpoint of commercial practicality and feasibility.

14.5. Design a full wave rectifier with *L*-section filter for use with a 60 cps supply to provide E_{dc} (full load) = 300 v; I_{dc} (full load) = 100 ma; ripple = 1%; volt. reg. = 5%. Discuss some of the various design alternatives and comment upon relative feasibility and cost of some of the alternatives.

14.6. Redesign the rectifier of problem (14.5), using a two-section filter with only 0.1% ripple factor allowed.

14.7. Analyze the performance of a full wave rectifier with series inductor filter. Derive equations for E_{dc} and the cutout angle.

14.8. Outline a design procedure for the approximate design of a radio frequency power supply.

14.9. Design a voltage regulated power supply using glow discharge tubes to provide a regulated output of 210 v. Over what range of input voltages and load currents will regulation be obtained? What is the maximum allowable load current? How can it be increased?

14.10. Design a voltage regulated power supply using vacuum tubes connected as shown in the text to provide 250 v at 100 ma regulated from a 500 v unregulated supply. The output is to be held constant to $\pm 0.1\%$ for a $\pm 20\%$ variation of unregulated input and over the full range of load current from 0 to 100 ma.

This is an extensive design problem in which the first step is the derivation of appropriate design formulas. Then the problem of tube selection arises and should be discussed at some length. Thereafter the problem becomes routine.

Chapter 15

DETECTORS AND MIXERS

A *detector* or *demodulator* is used to separate the original intelligence being transmitted by a modulated wave. *Linear diode detectors* are generally used to demodulate amplitude modulated waves. Square law detectors can be used for the same purpose, but they operate in class A and are treated in chapter 11.

Mixers, or *frequency converters*, are circuits capable of performing frequency translation. Such circuits are found, along with linear diode detectors, in all superheterodyne radio receivers. The same effect can be achieved by square law action, but it is not as efficient nor as distortionless as operation in the switching mode. Thus most practical mixers operate according to the principle outlined in this chapter.

It will be shown here that linear detectors are simply rectifiers operating under certain special conditions, so that some of the design procedures and charts developed in the preceding chapter can be used here.

An equivalent circuit will be developed for mixers that will make it directly comparable to amplifiers, so that nearly all the amplifier theory developed in earlier chapters can be applied to mixers.

15.1. Linear Diode Detector

Two circuit diagrams of linear detectors are shown in figure (15.1). Although a half wave rectifier is shown, full wave or bridge rectifier circuits can also be used. It should be clear from this figure that a linear diode detector is really nothing more than a rectifier with *RC* filter. Thus the generalized design curves for this rectifier-filter circuit given in figures (14.15) through (14.19) can be used to design diode detectors with certain additional considerations to be covered later.

The generalized curves of the rectifier with *RC* filter given in chapter 14 show the operation of the circuit when the input is a constant amplitude sinusoid of fixed frequency. The resulting output voltage is constant with a superimposed ripple component. The ratio of the direct output voltage to the peak value of the signal input is called the *detection efficiency* and denoted by η_d. That is, $\eta_d = E_{dc}/E_s$. Thus figures (14.15) through (14.17) can be used to determine the detection efficiency for any circuit and combination of ω, R_L, R_s, and C.

When these circuits are used as detectors, the input signal is amplitude modulated rather than being a continuous wave. Therefore $e_s = E_s(t) \cos \omega_c t$, where ω_c = carrier frequency. Because amplitude modulation

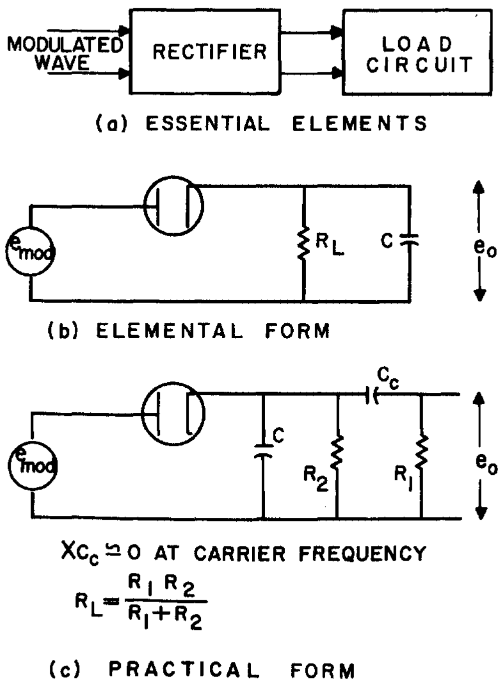


Fig. 15.1. Linear diode detector.

is assumed, $E_s(t)$ varies with time in accordance with the modulating signal of frequency ω_m. As long as this variation is slow compared with the carrier frequency, the output from the *RC* filter will vary along with $E_s(t)$, and the results obtained for the rectifier with *RC* filter are not altered. Thus the ω appearing in the dimensionless parameter $\omega R_L C$ and appearing in figures (14.15) through (14.19), should be replaced by ω_c, the carrier frequency.

For an amplitude modulated carrier,

$$E_s(t) = E_s(1 + m_a \cos \omega_m t)$$

where E_s = unmodulated carrier amplitude

$$m_a = \frac{E_m}{E_s} = \frac{\text{amplitude of modulating wave}}{\text{amplitude of unmodulated carrier}}$$

= modulation index

Therefore the detector output will be

$$E_0(t) = E_s\eta_d(1 + m_a \cos \omega_m t)$$

The detector output contains two terms: (1) a constant voltage $= \eta_d E_s$; (2) the original modulating signal $= \eta_d E_s m_a \cos \omega_m t$. Because the original modulating signal has been recovered from the modulated carrier, detection or demodulation has been achieved.

In general the constant term is undesired and a blocking capacitor C_c is ordinarily used as shown in figure (15.1c). The reactance of this capacitor is negligible at the carrier frequency, so that the effective load resistance on the diode is the parallel combination of R_1 and R_2 as shown.

15.2. Maximum Permissible Capacitance

The detector output voltage will follow the modulating signal exactly as long as the filter capacitor is permitted to discharge as rapidly as the amplitude of the carrier wave decreases. If the modulated carrier input drops too rapidly, the plate voltage of the diode decreases more rapidly than the cathode potential, so that the plate-to-cathode voltage becomes negative; diode conduction then ceases. If this occurs, the output voltage is governed by the rate at which C discharges through R_L. The output voltage during such an interval is

$$e_0(t) = E\varepsilon^{-t/R_L C}$$

where E = value of the output voltage when the rate of discharge of C was too slow to follow the modulation envelope. Hence, if the output voltage is to follow the modulation envelope, it is necessary that the rate of change of capacitor voltage be equal to or greater than the rate of change of the modulation envelope at the highest modulating frequency. That is,

$$\frac{dE_0(t)}{dt} \geq \frac{dE_s(t)}{dt} \tag{15.1}$$

However,

$$E_0(t) = E\varepsilon^{-t/R_L C}$$

$$E_s(t) = E_s\,(1 + m_a \cos \omega_m t) \tag{15.2}$$

Differentiating each expression and setting up the specified inequality yields

$$\omega_m R_L C \leq \frac{E}{E_s} \cdot \frac{\varepsilon^{-t'/R_L C}}{m_a \sin \omega_m t'}$$

where t' = time when the diode stops conducting; $E\varepsilon^{-t'/R_L C}$ = capacitor voltage at the instant the diode stops conducting. Hence

$$E\varepsilon^{-t'/R_L C} = E_s(1 + m_a \cos \omega_m t')$$

Consequently, the preceding inequality can be written

$$\omega_m R_L C \leq \frac{1 + m_a \cos \omega_m t'}{m_a \sin \omega_m t'} \tag{15.3}$$

The maximum value of the factor $\omega_m R_L C$ can be found by differentiating the preceding expression with respect to t', setting the result equal to zero, and solving for t'. This value of t' is then substituted back into equation (15.3) to obtain the maximum value of $\omega_m R_L C$. Proceeding in this manner yields $m_a = -\cos \omega_m t'$, so that $\sqrt{1 - 2m_a^2} = \sin \omega_m t$. Therefore

$$(\omega_m R_L C)\text{max} = \frac{\sqrt{1 - m_a^2}}{m_a} = \sqrt{\frac{1}{m_a^2} - 1}$$

Terms can be rearranged and the equation expressed as

$$C\text{max} = \frac{1}{\omega_m \text{ (max) } R_L} \sqrt{\frac{1}{m_a^2} - 1} \tag{15.4}$$

The maximum value of capacitance that can be used successfully in the detector load circuit can be determined as soon as the modulation index, highest modulating frequency, and detector load resistance are known. Of these three factors, the first two, ω_m (max) and m_a, will be determined by other considerations; circuits are usually designed for $m_a = 1$. Only the detector load resistance is unspecified. It is usually governed by input impedance considerations, a problem explored in some detail in the next section.

Equation (15.3) is plotted in a convenient form for design purposes in figure (15.2). The proper value for $C_{\max}$ is easily determined from this figure by the following method:

(1) Assume that R_L, ω_m(max), and m_a are known.

(2) Locate the specified value of R_L on the chart. Find where the vertical line of constant R_L intersects the proper line of constant m_a.

(3) From this intersection, project horizontally to the specified line of constant ω_m(max).

(4) From this intersection project vertically down and read the value of $C_{\max}$.

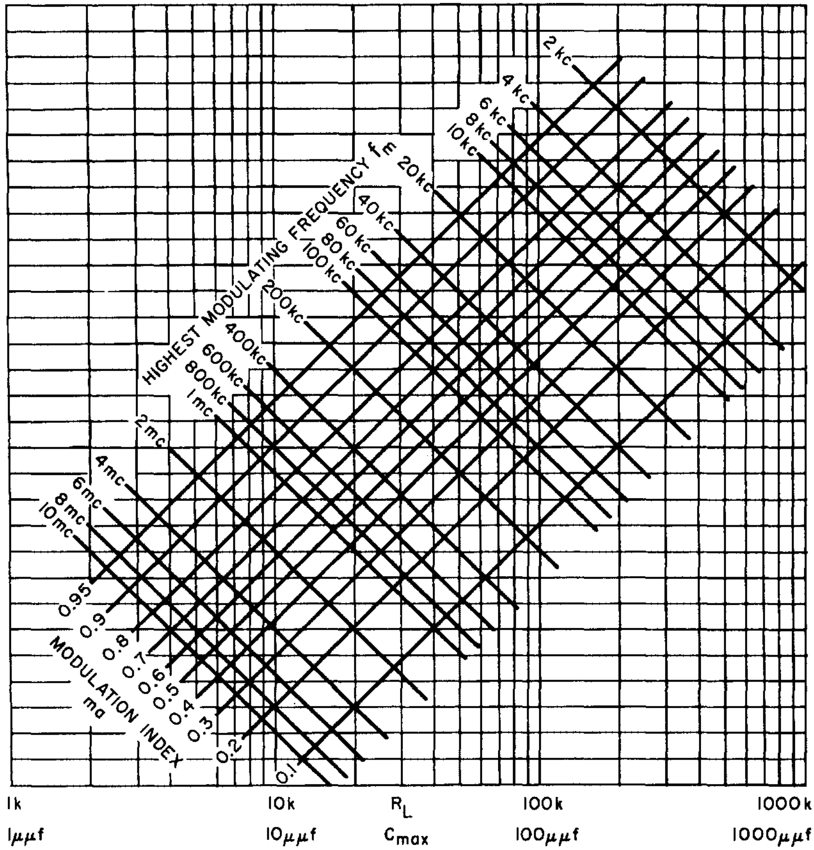


Fig. 15.2. Linear diode detector: determination of $C_{\max}$.

(5) For example, if $f_m = 5000$ c, $m_a = 0.5$, $R_L = 100{,}000$ ohms, $C_{\max} = 730\ \mu\mu$f.

15.3. Equivalent Representation of a Linear Detector

The operation and some of the performance characteristics of a linear detector can be conveniently represented through the use of a block diagram and its associated transfer function.

The detector has three separate and distinct components in the output:

(1) a constant term

$$E_{dc} = \eta_d E_s$$

(2) a signal component

$$e = \eta_d m_a E_s \cos \omega_m t$$

(3) a ripple component

$$e_r = \eta_d \gamma E_s \cos n\omega_c t$$

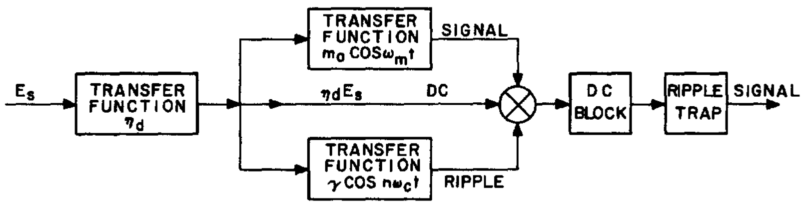


Fig. 15.3. Block diagram representation of a linear diode detector.

where γ = ripple factor; $n = 1$ for a single diode, or 2 in other rectifier circuits. Thus the detector can be thought of as a single input system with three outputs as shown in figure (15.3). Also, most circuits are designed to block the d-c component and trap the ripple so that the over-all system has the appearance of figure (15.3).

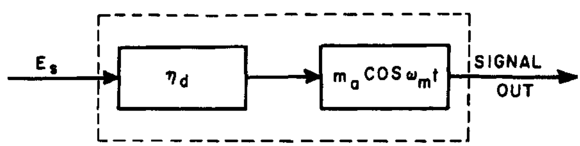


Fig. 15.4. Simplified block diagram of a linear diode detector.

If the circuit is actually designed to block the direct current and trap or otherwise remove the ripple component, the over-all operation of the detector is conveniently represented by the block diagram of figure (15.4). Transfer functions are shown inside the blocks.

The response characteristics of the detector can be computed from the proposed equivalent circuit shown in figure (15.5). Here r_d is a hypothetical resistance whose value must be computed to validate the equivalent circuit. If the circuit is to represent truly the actual circuit behavior, the output voltage must be

$$\eta_d E_m = \eta_d m_a E_s$$

Assume the reactance of the capacitor C to be high at the frequency of the modulating signal. Hence

$$E_m = I_m r_d + I_m R_L = I_m r_d + \eta_d E_m$$

so

$$E_m(1 - \eta_d) = I_m r_d$$

But, from the equivalent circuit in figure (15.5),

$$I_m = \frac{E_m}{r_d + R_L}$$

It is then easily shown that

$$r_d = R_L \frac{1 - \eta_d}{\eta_d} \tag{15.5}$$

Hence the equivalent detector resistance is easily computed when the rectification efficiency is known.

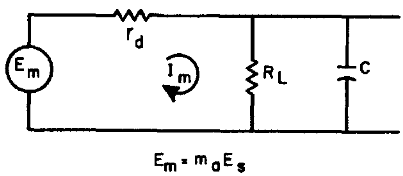


Fig. 15.5. Proposed equivalent circuit for a linear diode detector.

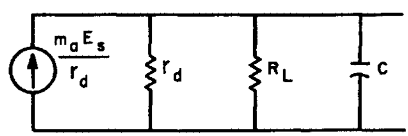


Fig. 15.6. Current source equivalent circuit for a linear diode detector.

The constant current form of the equivalent circuit is given in figure (15.6). Because this has precisely the same form as the equivalent plate circuits of class A vacuum tube amplifiers, all data on rise times, overshoots, and cutoff frequencies derived there can be applied to detectors. The mid-band, or reference, gain of the detector is clearly

$$A_r = \eta_d m_a \tag{15.6}$$

15.4. Input Impedance of a Diode Detector

The diode detector in a receiving system is usually driven by an amplifier stage. The detector loads the amplifier and thereby affects its amplification characteristics. Thus it is necessary to have a method of determining the amount of loading introduced by the detector.

Designate the effective input resistance of the detector circuit by R_e. Then the average power input to the circuit will be

$$P_{in} = \frac{E_s^2}{2R_e} \tag{15.7}$$

or the equivalent input resistance is

$$R_e = \frac{E_s^2}{P_{in}} \tag{15.8}$$

where P_{in} is the sum of the power lost in the equivalent detector resistance r_d, plus that delivered to the load resistor R_L. Hence

$$P_{in} = I_{dc}^2 R_L + I_{dc}^2 r_d \tag{15.9}$$

However, $E_{dc} = I_{dc} R_L$, so that

$$P_{in} = \frac{E_{dc}^2}{R_L^2}(r_d + R_L) = \frac{E_{dc}^2}{R_L}\left(1 + \frac{r_d}{R_L}\right) \tag{15.10}$$

Therefore the equivalent input resistance becomes

$$R_e = \frac{R_L}{2}\left(\frac{E_s}{E_{dc}}\right)^2 \cdot \frac{1}{1 + r_d/R_L} \tag{15.11}$$

However, because

$$\eta_d = \frac{E_{dc}}{E_s} \quad \text{and} \quad r_d = R_L \frac{1 - \eta_d}{\eta_d}$$

then

$$R_e = \frac{R_L}{2\eta_d^2} \cdot \frac{1}{1 + r_d/R_L} = \frac{R_L}{2\eta_d} \tag{15.12}$$

The equivalent input resistance of the detector can be determined from equation (15.12) after $\omega_c R_L C$ and R_s/R_L are known, because they determine the rectification efficiency. Then the value of η_d can be determined from figures (14.15) through (14.17). Resistor R_L is selected to provide the desired value for R_e based upon consideration of the driving amplifier.

In brief, the design of a linear diode detector may proceed according to the following outline:

(1) The value of R_e will be specified by the requirements of the driving amplifier stage.

(2) The value of R_s will be set by the circuit elements and the diode characteristics.

(3) By trial and error, the values of R_L and C will be determined from figures (14.15) through (14.17) and (15.2).

The characteristic curves of a diode rectifier are shown in figure (15.7). These curves show the relationship between the output current and voltage with the carrier amplitude as the parameter. They are directly analogous to the plate characteristics of a triode.

Once the load resistance for the detector has been determined, the quiescent or d-c load line can be drawn on these characteristics as shown in figure (15.7). The quiescent condition corresponds to operation with an unmodulated carrier input. The operating point Q is located under this quiescent condition; it can be determined only if the amplitude of the unmodulated carrier input is known.

It was shown in chapter 11 that amplifier tubes have different load resistances for quiescent operation than for dynamic or a-c operation. This was the origin of the a-c and d-c load line terminology. Exactly the same situation exists with detectors except that the signal frequency

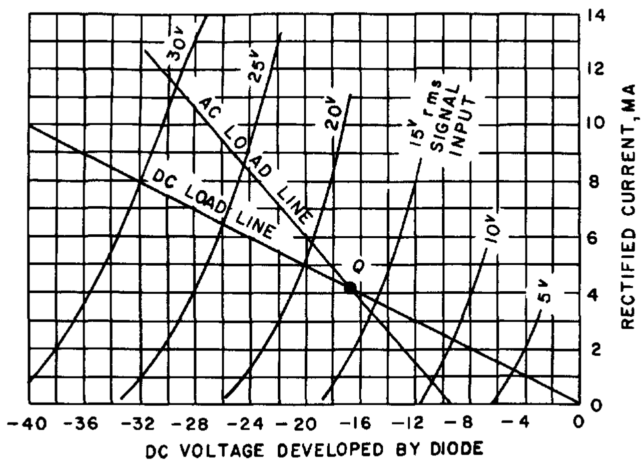


Fig. 15.7. Detection characteristics of a receiving type of diode.

is the frequency of the modulating signal, not the carrier. Thus an a-c load line must be drawn through the Q point as shown in figure (15.7). This line defines the path of operation of the circuit on the detector characteristics. This path becomes elliptical if the load contains reactive components.

It is clear from figure (15.7) that serious distortion will be produced when the carrier is modulated near the 100% point. This is so because the diode will be made nonconducting on the negative swing of the modulating signal for a period exceeding one cycle of the carrier. This acts to flatten the bottom of the detector output signal. It is called *clipping*. Negative peak clipping can be prevented by reducing the modulation index of the incoming wave or by making the a-c and d-c load lines as nearly coincident as possible. This requires that the a-c and d-c load resistances be as nearly equal as possible.

15.5. Some Other Detector Circuits

In addition to diodes, triodes and pentodes are occasionally used as detectors, usually operating in one of the following two modes: (1) grid leak detection; (2) plate detection.

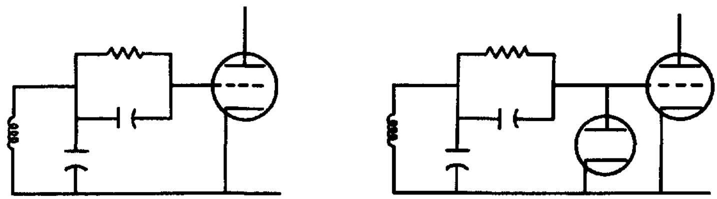


Fig. 15.8. Grid leak detector.

The circuit diagram of a grid leak detector is shown in figure (15.8). In principle, the grid-cathode circuit simply operates as the rectifier element, replacing the diode in the conventional detector; the grid leak bias circuit corresponds to the detector load. Thus the circuit can be designed in exactly the same manner as the conventional diode detector. The main claimed advantage for the circuit is that the detected signal is amplified in the tube and the functions of detection and amplification are thereby combined. However, when semiconductor diodes are used in conventional detectors and followed by an amplifier stage, the same effect is achieved. However, detection with diodes is more linear than with triodes and pentodes and is usually preferred.

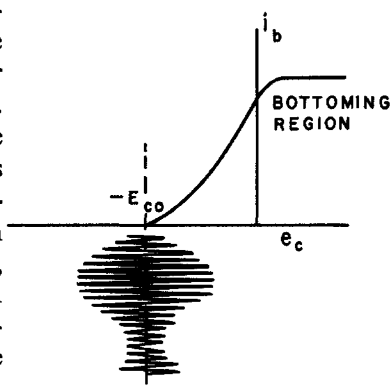


Fig. 15.9. Plate detector operation.

Plate detection is accomplished by operating the tube in class B so that the tube is biased to cutoff. Rectification is then produced by nonconduction in the plate circuit. This is illustrated in figure (15.9). This form of detection is more nonlinear than that obtainable with the conventional diode because of curvature in transfer characteristic near cutoff. However, this type of detector does not appreciably load the amplifier driving stage, because grid current generally does not flow. It is often called an *infinite impedance* detector for this reason.

Pentodes can also be used to provide rectification by *bottoming* (see section 1.5). This is not usually desirable because the grid voltage required is quite positive and the input impedance of the circuit is fairly low.

15.6. Automatic Gain Control

The signal appearing at the antenna of a receiver often varies in strength from time-to-time and from station-to-station. Despite this variation it is desirable to keep the receiver output constant. This output can be held fixed if the detector output does not vary. Circuits

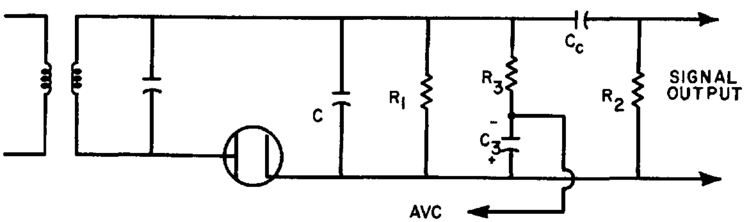


Fig. 15.10. Detector circuit with simple AVC; AVC voltage is negative with respect to ground and is used to bias the super control IF amplifier tubes.

accomplishing this are called automatic volume control (AVC) or automatic gain control (AGC) circuits. They are used in virtually all radio receivers.

The principle involved is rather simple. It will be recalled from section (15.1) that the linear diode detector develops a component of direct voltage $\eta_d E_s$ in the output that is proportional to the strength of the carrier E_s. If the detector circuit is grounded so that this constant voltage is negative, it can be used to control the bias, and hence the gain, of the remote cutoff tubes used in the IF amplifier stages. This is called *simple AVC,* and a typical circuit is shown in figure (15.10).

In practice it is necessary to separate the d-c component from the signal component. This action is provided by the R_3–C_3 circuit in figure (15.10). The time constant of this circuit is made long compared with the period of the lowest modulating frequency.

The primary disadvantage of simple AGC is that it is always operative and always reduces the receiver gain to some extent even when the signal input is weak. This is illustrated in figure (15.11). An ideal AGC system would allow the receiver to operate at maximum gain for weak signals and then hold the output constant when the input reaches and exceeds a certain value. The ideal AGC system would have the characteristic shown in figure (15.11).

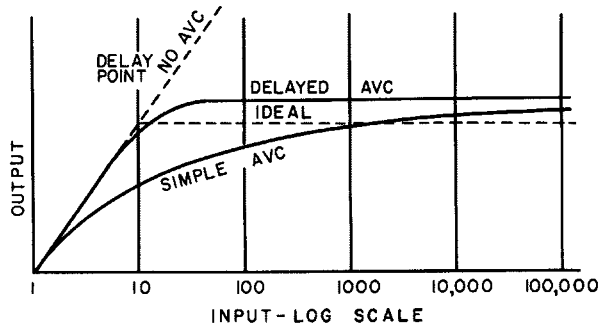


Fig. 15.11. Input-output characteristics of receivers with various types of AGC.

In effect it is desirable to delay the development of AVC voltage until the input signals exceed a certain specified value. Circuits providing this service are called *delayed AGC* circuits, and a typical arrangement is shown in figure (15.12).

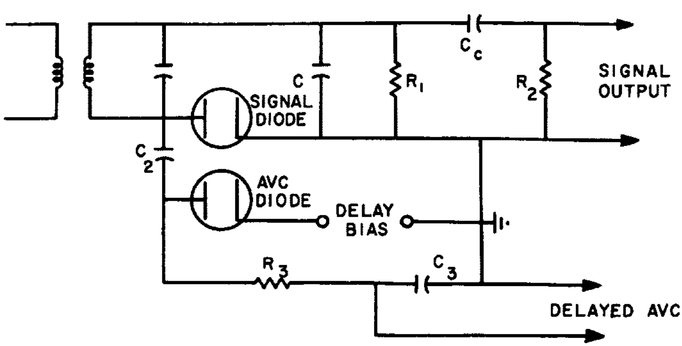


Fig. 15.12. Delayed AGC circuit.

This circuit requires a second diode that is biased so that it does not rectify and develop AGC voltage until the input signal reaches a certain level. The resulting circuit characteristic closely approximates the ideal form as shown in figure (15.11).

15.7. Problems in FM Detection

At the point of reception, a frequency modulated signal varies in both frequency and amplitude. However, the intelligence being transmitted depends upon the frequency variations of this signal, while the noise and interference mainly appear in the amplitude variations. Thus the ideal FM detector should have the following characteristics:

(1) It should develop an output proportional to the variations in the frequency of the received signal.

(2) It should be insensitive to amplitude variations.

This action is achieved in several ways, but there is considerable room for further development work.

One type of FM detector is shown in figure (15.13) in block diagram form. The incoming IF signal is passed through an *amplitude limiter*,

Fig. 15.13. One type of FM detector.

or *clipper* (see chap. 17), so that the output has a constant amplitude, but varies in frequency. This is then applied to a frequency *discriminating* circuit that develops an output that is proportional to the signal frequency. This is then detected by a conventional linear diode detector. The components of this detector system are discussed in later sections.

15.8. Amplitude Limiters[1]

The simplest type of amplitude limiting is obtained by ordinary gridleak bias action. A sharp cutoff tube, preferably a pentode, is connected with zero cathode bias and a grid leak bias circuit. The tube is operated at low plate and screen potentials so that the linear operating region between cutoff and bottoming is only 3 or 4 volts in extent, as shown in figure (15.14). With a good bottoming characteristic, the gridleak biasing is scarcely required. As long as the grid signal has sufficient amplitude to cause the operating point to traverse the entire linear region, a constant output will be obtained.

[1] For a more complete discussion of the application of limiters in FM receivers, see Milton B. Sleeper, *Standard FM Handbook*, 1st ed., FM Co., Great Barrington, Mass., 1946.

A more complete discussion of clipping circuits is presented in chapter 17 and many more circuits are discussed.

The input-output characteristic of a single limiter stage appears approximately as shown in figure (15.15). Improved operation results if two stages are cascaded; the effect is shown in figure (15.15).

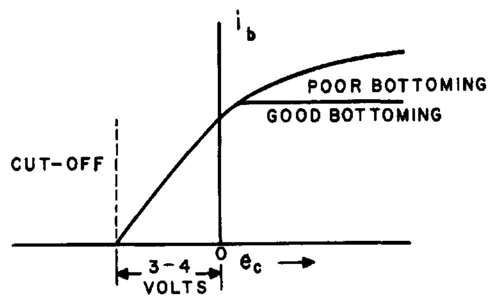


Fig. 15.14. Operation of a simple limiter.

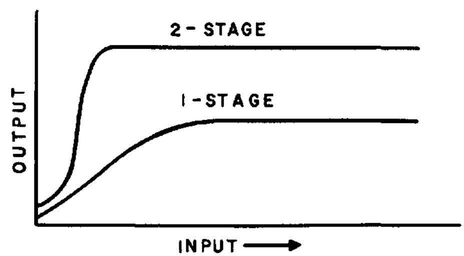


Fig. 15.15. Input-output characteristics of amplitude limiters.

15.9. Discriminators

The purpose of the *discriminator* in the detector of an FM receiver is to convert constant amplitude, variable frequency signals into an output whose amplitude is proportional to the frequency deviation. A tuned circuit is the simplest way of doing this, as shown for the case of a simple resonant circuit in figure (15.16). By setting the center frequency of the input signal off resonance, as shown, the output voltage will vary with the frequency of the applied signal. The circuit is of little practical importance because of the obvious nonlinear characteristic.

A very linear characteristic can be obtained with the *Foster-Seeley discriminator* shown in figure (15.17). The circuit is essentially a double tuned amplifier with both the primary and secondary tuned to the same

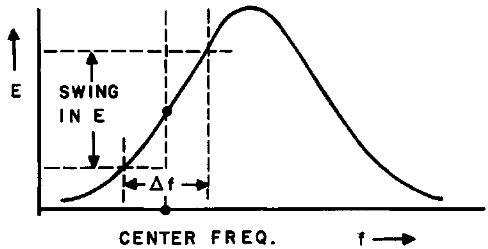


Fig. 15.16. Simple resonant circuit as a discriminator.

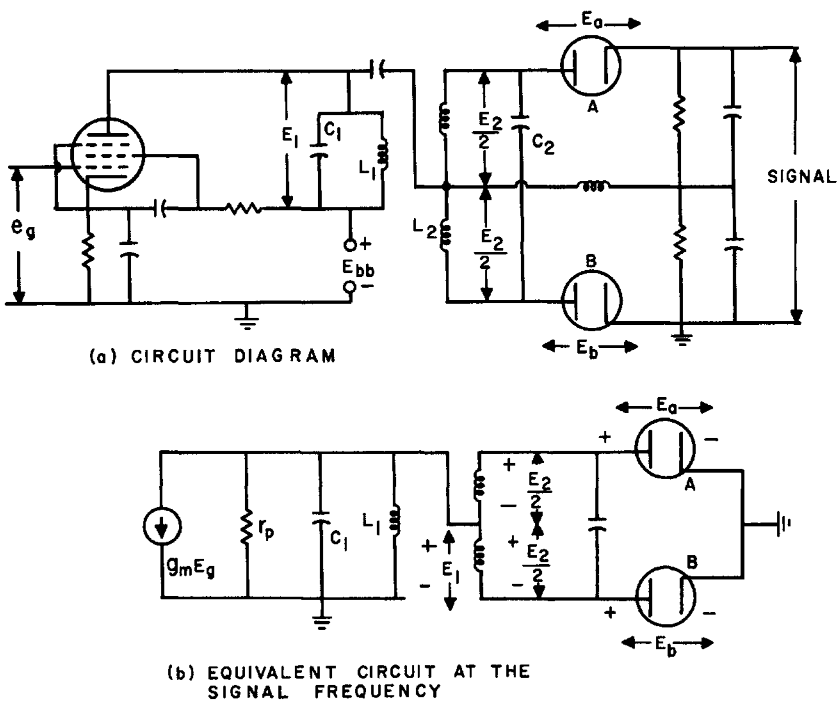


Fig. 15.17. Foster-Seeley discriminator.

frequency and having the same Q. It is assumed that the deviation in frequency off resonance is small so that

$$\delta = \frac{\omega}{\omega_{ar}} - 1$$

is much less than unity. If the coefficient of coupling k and Q are adjusted so that k^2Q is negligibly small, it is a relatively straightforward process to prove[2] the following:

(1) At resonance, E_2 leads E_1 by 90 degrees.

(2) At frequencies below resonance, E_2 leads E_1 by less than 90 degrees.

(3) At frequencies above resonance, E_2 leads E_1 by more than 90 degrees.

This will be important in a moment in explaining the operation of the circuit.

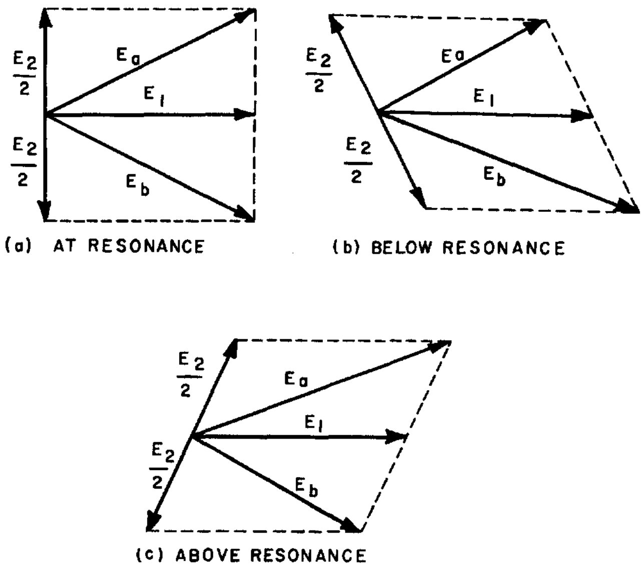


Fig. 15.18. Determination of diode voltages at various frequencies for the Foster-Seeley circuit.

The RF choke in the circuit of figure (15.17a) is a virtual open circuit at the signal frequency, while C_a, C_b, and C_c are virtual short circuits. Therefore the signal frequency equivalent circuit of figure (15.17b) is obtained. From this diagram it is obvious that the voltage drops across the two diodes are

$$E_a = E_1 + \frac{E_2}{2}; \qquad E_b = E_1 - \frac{E_2}{2}$$

[2] See, for example, J. D. Ryder, *Electronic Fundamentals and Applications*, Prentice-Hall, Inc., New York, 1950, pp. 538–546.

Thus it is an easy matter to make the vector diagrams to show the way the voltages E_1 and E_2 combine to provide the diode voltage at resonance and above and below resonance. This is shown in figure (15.18). The diode voltages are equal at resonance, but they are different at all other times.

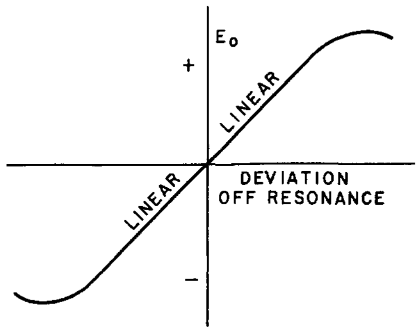


Fig. 15.19. Characteristic of a Foster-Seeley discriminator.

From the circuit diagram in figure (15.17a) it is clear that the total output voltage is the *difference* between the detector outputs. That is

$$E_0 = \eta_d(E_a - E_b)$$

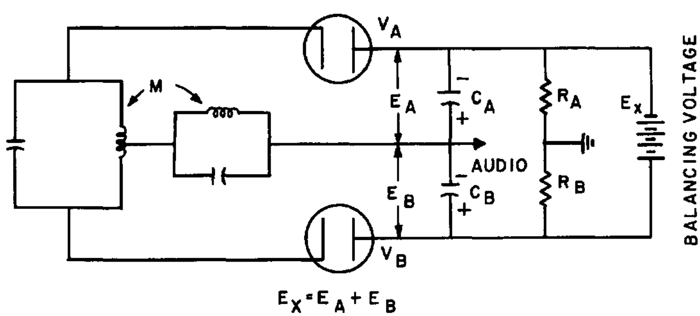


Fig. 15.20. Elementary ratio detector.

Hence at resonance the output is zero. At frequencies above resonance E_a is greater than E_b and the output is positive. At frequencies below resonance E_a is less than E_b and the output is negative. Over a confined frequency range an exactly linear characteristic is obtained as shown in figure (15.19).

Another circuit that closely resembles the one just discussed is the ratio detector shown in figure (15.20). The main advantage of this circuit is that limiters are not required. The same general type of center-tuned discriminator is used. The secondary is center tapped to provide equal voltages of opposite polarity to the diodes; the primary voltage is applied to the diodes in parallel. As a result, as the frequency of the incoming signal varies, the phase angles of the primary and secondary voltages vary as before. This circuit is different mainly in the location of the signal output and the presence of the stabilizing voltage.

Because the stabilizing voltage is connected across the two capacitors, the rectified output voltages across these capacitors must be equal to the stabilizing voltage at all times. Hence, the potential at the signal take-off point can vary only when the ratio of the rectified output voltages changes. There is a change in this ratio only when the frequency of the incoming signal changes. Hence the signal output will vary only with changes in the frequency of the incoming signal. Amplitude variations of the FM signal have no effect because this does not change of the ratio of the rectified voltages and no signal output is developed at the take-off point.

15.10. Principles of Mixers

One common application of the principle of rectification is found in the *mixer* or *converter* stage of every superheterodyne receiver. The

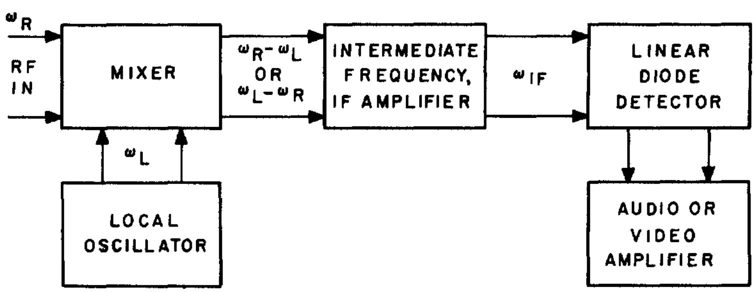


Fig. 15.21. Block diagram of a superheterodyne radio receiver.

block diagram of such a receiver is shown in figure (15.21). The incoming radio frequency signal from the antenna is combined with a locally supplied signal from the local oscillator. The total signal is then applied to a rectifier element called the *mixer*. The mixer output contains many frequency terms including the *difference* frequency as shown in figure (15.21). Depending upon the relative frequencies involved,

$$\omega_{if} = \omega_{rf} - \omega_{lo} \quad \text{or} \quad \omega_{if} = \omega_{lo} - \omega_{rf}$$

In effect, the original incoming signal is translated down the frequency scale so that such circuits perform frequency translation, or conversion.

Suppose that the incoming radio frequency signal is designated as $e_{rf} = E_R \sin \omega_R t$, and the local oscillator signal is $e_{lo} = E_L \sin \omega_L t$. If both signals are impressed in series with a *linear* circuit element, the total applied voltage is

$$e_t = e_{rf} + e_{lo} = E_R \sin \omega_R t + E_L \sin \omega_L t$$

According to the principles of linear circuit theory, these two voltages can be represented as rotating vectors or phasors. Also, if the E_L phasor is taken as the reference so that it appears stationary on the diagram, then E_R is connected to the tip of E_L and rotates with an angular velocity $(\omega_R - \omega_L)$ either clockwise or counterclockwise depending upon the relationship between ω_R and ω_L.

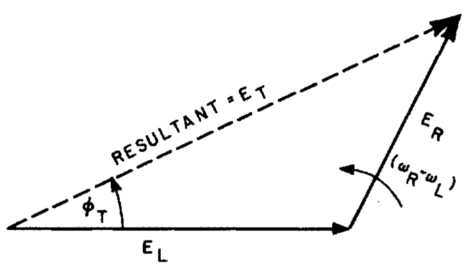


Fig. 15.22. Addition of two sinusoidal voltages of different frequency.

The total applied voltage is the sum of the two phasors E_R and E_L, and it oscillates in time with a frequency ω_L. The magnitude of the total voltage E_T can be computed from figure (15.22) by the law of cosines. That is,

$$E_T = \sqrt{E_L^2 + E_R^2 + 2E_R E_L \cos(\omega_R - \omega_L)t}$$

In nearly all practical cases the local oscillator voltage is many times larger than the received radio signal, so that E_R is much less than E_L. Therefore E_R^2 is negligible compared with E_L^2. Thus

$$E_T \doteq E_L \sqrt{1 + 2\frac{E_R}{E_L}\cos(\omega_R - \omega_L)t}$$

Because the second term under the radical is always much less than unity, the approximate form can be written

$$E_T \doteq \sqrt{1+x} \doteq 1 + \frac{x}{2}$$

Hence $$E_T \doteq E_L\left[1 + \frac{E_R}{E_L}\cos(\omega_R - \omega_L)t\right]$$

As long as E_R is much less than E_L, and this is generally true, the envelope of the new carrier at the difference frequency is a reproduction of the envelope of the carrier of the incoming RF signal.

It should also be clear from figure (15.22) that the phase angle ϕ_T of

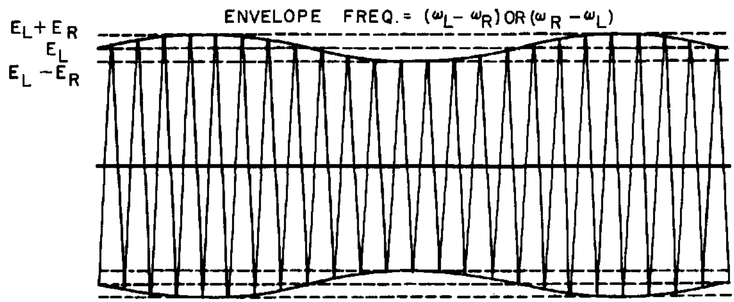


Fig. 15.23. Sum of two sinusoidal voltages of different frequency when the magnitude of one is small compared to the other.

the resultant also varies with time, and this would have the effect of making the frequency of the resultant vary somewhat. The effect is negligible under the assumed inequality.

Clearly then, the sum of the applied voltages appears as shown in figure (15.23) for the case assumed. It is evident that the *envelope* of the total voltage is a replica of the input and has a frequency equal to the *difference* frequency of the two signals.

Now suppose that this total voltage is applied to any rectifier. It could be a vacuum tube or varistor or a multielectrode tube biased to cutoff. All that is required is a sharp break between a virtually linear region of conduction and a nonconducting region. Thus, assuming a diode, operation can be indicated graphically as shown in figure (15.24). Clearly, the circuit functions as a half wave rectifier when the amplitude of the voltage of the rectified voltage varies sinusoidally at the difference frequency.

The Fourier series of the output from a half wave rectifier was given in section (14.4) as

$$i = \frac{I_m}{\pi}\left(1 + \frac{1}{2}\sin\omega_L t - \frac{2}{3}\cos 3\omega_L t + \ldots\right)$$

However, in this case, I_m, which is the amplitude of the wave, varies with time as follows:

$$I_m \doteq I_L\left[1 + \frac{E_R}{E_L}\cos(\omega_R - \omega_L)t\right]$$

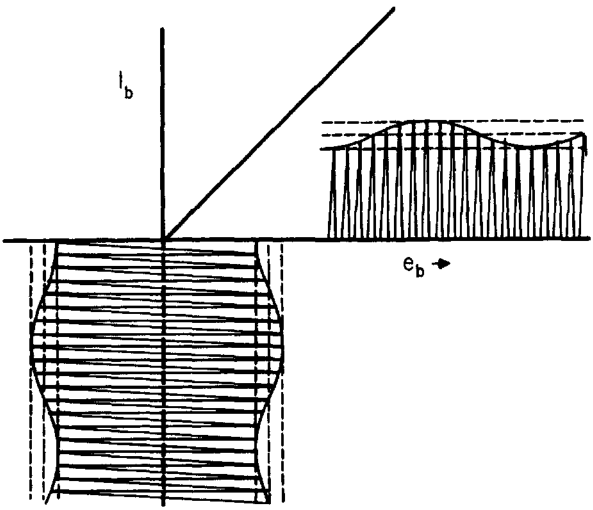


Fig. 15.24. Use of half wave rectification in mixer operation.

Substitute for I_m in the equation for i, evaluate a few terms, and the result is

$$i = \frac{I_L}{\pi}\left[1 + \frac{E_R}{E_L}\cos(\omega_R - \omega_L)t + \frac{1}{2}\sin\omega_L t + \ldots\right]$$

There are an infinite number of terms in all possible combinations, but interest centers upon the second term because it is the difference frequency component desired. If the load circuit on the diode is tuned to this frequency and the bandwidth is adjusted to exclude other terms, frequency translation is accomplished and the primary function of the mixer is fulfilled.

15.11. Some Actual Mixer Circuits

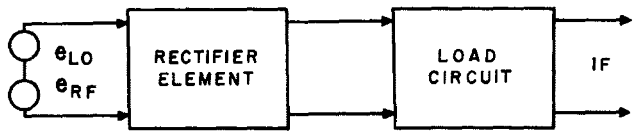
All mixer circuits involve three essential components as follows:

(1) A method of combining the two input signals.

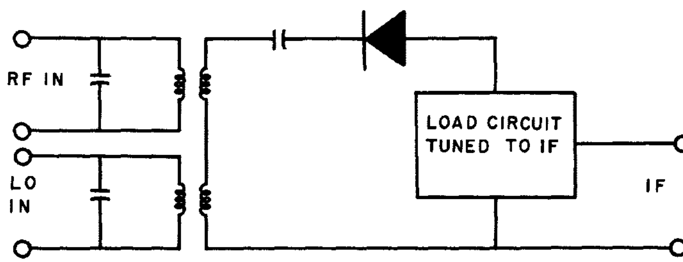
(2) A circuit element with a sharp break in its current-voltage characteristic for rectification.

(3) A tuned circuit to select the difference frequency output.

Mixer circuits are often classified according to the number of inputs. That is, a diode has only a single input. Hence the two input signals are



(a) BLOCK DIAGRAM SHOWING GENERAL COMPONENTS.



(b) USE OF A DIODE AS A RECTIFIER ELEMENT. IT COULD BE REPLACED BY A TRIODE OR A PENTODE BIASED TO CUT-OFF. CAN ALSO USE CAPACITIVE INPUT COUPLING RATHER THAN INDUCTIVE.

Fig. 15.25. Single input mixers.

applied in series with one lead as shown in figure (15.25). This is typical at ultrahigh frequencies.

General practice below the ultrahigh frequency region involves the use of special tubes developed especially for mixer service. They are called *mixers* and *converters* and identified by the names and symbols shown in figure (15.26). As indicated in the figure, the tubes are of two general types:

(1) Mixers—local oscillator requires a separate tube.

(2) Converters—local oscillator tube is enclosed in the same envelope with the mixer tube. There are three subtypes in use: pentagrid converter, triode-hexode converter, and the octode converter.

In all cases it is evident that the addition of the signal frequency and local oscillator voltages is accomplished inside the tube because each voltage acts to control the mixer plate current. In all tubes a double screen grid is used for two reasons:

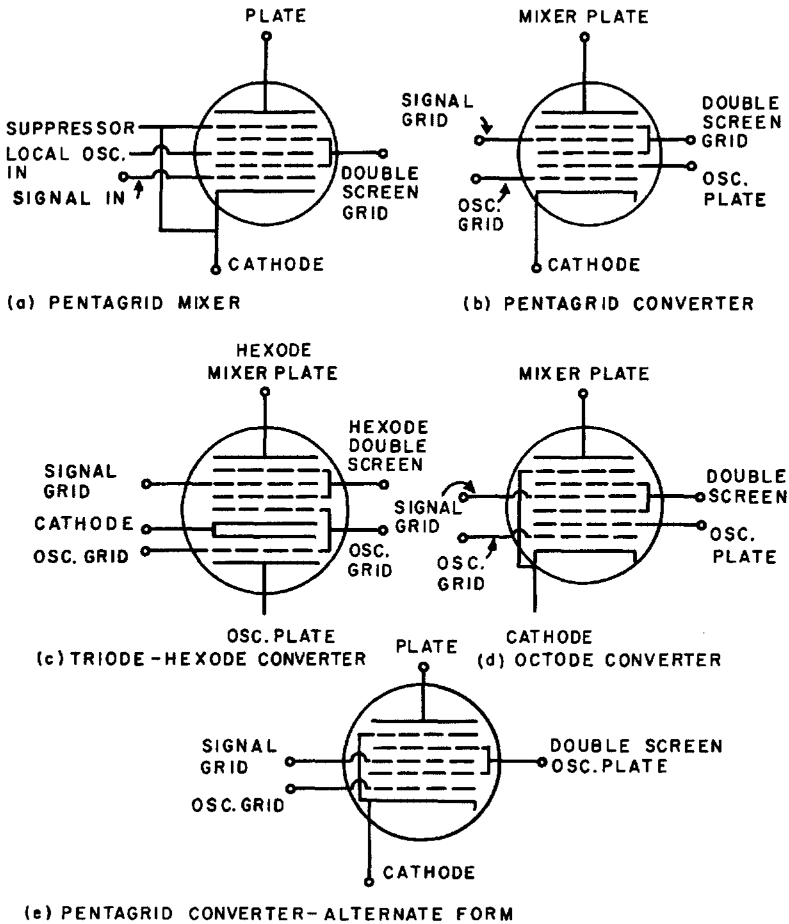


Fig. 15.26. Mixers and converters.

(1) It provides the electronic acceleration in the mixer section of the tube.

(2) It provides electrostatic shielding between the signal grid and the local oscillator input or section.

In general, the signal grid of a mixer or converter tube is made to have a remote cutoff characteristic so that the amplification of the stage can be controlled through AGC bias voltage variations. The oscillator grid normally has a sharp cutoff characteristic.

All these tubes operate on the same general principle. However, their variations make them particularly suited for use in specific applications.

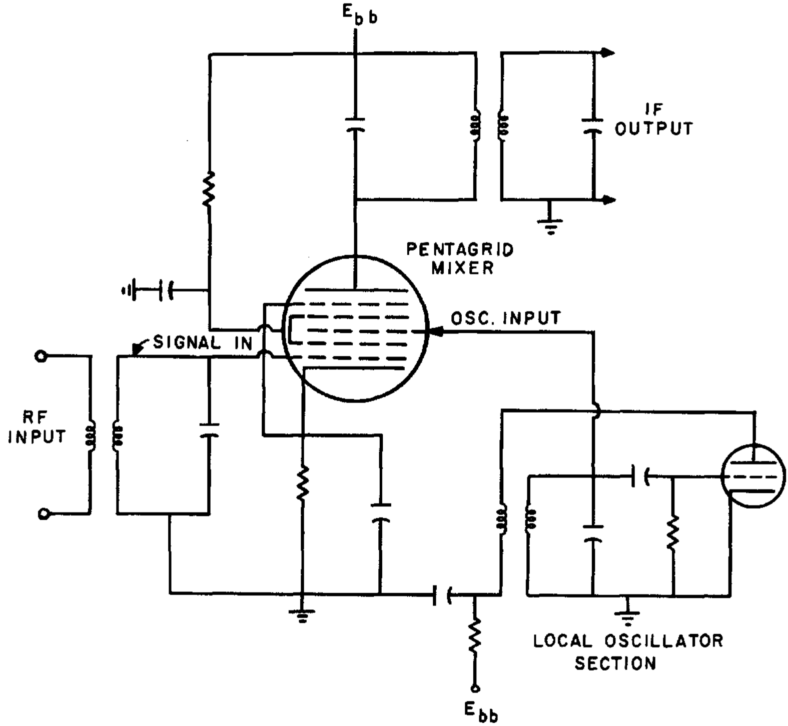


Fig. 15.27. Typical mixer circuit.

Thus pentagrid mixers are usually used for short-wave receivers; the triode-hexode converter is used in all-wave receivers because it is designed to be noncritical with respect to changes in signal bias or oscillator plate voltage. The pentagrid converter operates best at low frequencies because interactions between signal and oscillator grids increases with frequency. However, they are particularly suited for use with AGC because they are designed to minimize oscillator detuning resulting from AGC bias voltage variations on the signal grid.

A typical circuit diagram is given in figure (15.27). Although this illustrates the use of a pentagrid mixer with a separate local oscillator, essentially the same circuit applies to triode-hexode and pentagrid converter circuits.

15.12. Equivalent Circuits of Mixer Stages

The effectiveness of a mixer or converter tube in performing its function is characterized by a constant called the *conversion transconductance* g_c where

$$g_c = \frac{\partial i_{if}}{\partial e_{rf}}$$

It is analogous to the mutual transconductance of the tube except that it relates the output current of the tube at the difference frequency to the

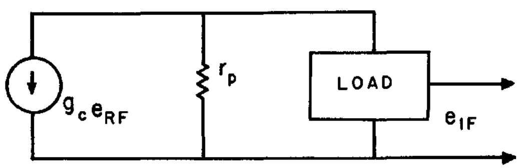


Fig. 15.28. Equivalent circuit of a mixer or converter stage under rated operating conditions.

input voltage at the radio frequency. The conversion transconductance will always be less than the mutual transconductance because the tube operates as a rectifier and conducts only for short periods. The ratio of g_c to g_m is equal to or less than $1/\pi$. Conversely, the effective plate resistance will always be larger for the same reason.

Because of the definition of the conversion transconductance, it is clear that a mixer stage can be represented by the equivalent circuit shown in figure (15.28). Because the form of this circuit is precisely the same as the form of the equivalent circuit of voltage amplifiers developed in Part II, the equation for the *conversion gain* can be written directly as

$$A_c = -g_c Z_m = \frac{E_{if}}{E_{rf}}$$

where Z_m = mutual impedance of the passive network in the equivalent plate circuit evaluated at the difference frequency. Hence, all the material developed in chapters 3 through 6 is directly applicable here and further discussion is not needed.

Fortunately, values of conversion transconductance can be obtained from the tube manuals. Representative values range from 250 μmhos to 950 μmhos with plate resistances of the order of 350,000 to 1,000,000 ohms.

PROBLEMS

15.1. An unmodulated 455 kcps carrier of 10 v amplitude is applied to a simple linear diode detector. The effective diode resistance is 1000 ohms and the load circuit has $R_L = 50{,}000$ ohms and $C = 0.01\ \mu$f. Compute the rectification efficiency, direct output voltage, and input impedance.

15.2. A 30 mcps, 10 v carrier is 50% modulated by the highest modulating frequency of 1 mcps. The diode resistance is 200 ohms; $R_L = 10{,}000$ ohms. Compute C_{max}, the direct output voltage, and the input resistance.

15.3. Compute the ripple voltage in the output of the detector of problem (15.2).

15.4. A linear diode detector is to be designed to operate with a 30 mcps carrier, $m_a = 0.45, f_m(\max) = 1$ mcps and an input resistance of not less than 4000 ohms. A germanium diode is used for which $r_p' = 100$ ohms. The d-c component is to be blocked. Design the detector.

15.5. Design a suitable trap for the ripple component in the detector of problem (15.4). Refer to any circuit theory text book for further guidance.

15.6. A mixer stage is to be designed using a 6SA7 tube. The necessary data for the tube are given in the tube manual. The incoming RF signal is 10 μv at 1 mcps and a 455 kc IF is desired. The bandwidth of the stage is to be 10 kcps. Design the stage assuming a total interstage capacitance of 25 $\mu\mu$f. What is the conversion gain?

15.7. A full wave detector is to be designed for use in an a-c vacuum tube voltmeter. The detector converts the alternating current to direct current, which is then indicated on a meter. The detector is to operate over a frequency range of 20 c to 20 kc and the scale calibration should be accurate to $\pm 5\%$ over this range. The diodes used have $r_p' = 200$ ohms. It is desirable to have the fastest possible meter recovery time following a measurement. Explore some of the various design possibilities in detail.

15.8. A parallel resonant circuit of $Q = 40$ and band center at 10 mcps is used as a discriminator in an FM detector. The incoming signal is centered on the lower 50% frequency of the characteristic of the tuned circuit. Find the change in output voltage if the input signal has a constant amplitude of 70 μamp with a frequency deviation of ± 75 kcps.

Chapter 16

MAGNETIC AMPLIFIERS

Magnetic amplifiers are rapidly becoming full-fledged working partners of vacuum tube and transistor amplifiers. Their extreme ruggedness, reliability, and nearly indefinite life make them attractive to equipment manufacturers.

The presentation here is far from complete because the subject is complicated and extensive, with abundant coverage in the periodical literature. For example, one source[1] lists nearly one thousand papers and patents. Complete coverage of the subject in the limited space available here is impossible. Only the essential fundamental characteristics of the circuits are treated here in the hope that the discussion will provide adequate preparation for further exploratory reading in the periodical literature.

Only magnetic amplifiers operating as amplifiers are presented in this chapter. Coincidence circuits, computers, and trigger circuits are incorporated in chapters 17 and 18, along with similar vacuum tube and transistor circuits.

A different and informative approach is followed in a paper by Manley.[2] The analysis presented here is standard.

16.1. Operating Principles, Zero Control Voltage

The circuit diagram of an elementary *series connected* magnetic amplifier is shown in figure (16.1a). The circuit consists of two identical *gate windings*, with ferromagnetic cores, connected in series with an alternating voltage source e_{ac} and a load resistance R_L. The hysteresis loss in the gate winding cores is assumed to be zero, so that the idealized form of the hysteresis loop appears as shown in figure (16.1b). A third coil, called the *control winding*, is wound about the two cores so that the

[1] T. G. Miles, "Bibliography of Magnetic Amplifier Devices and Saturable Reactor Art," *Trans. AIEE*, vol. 70, 1951, pp. 104–123.

[2] J. M. Manley, "Some General Properties of Magnetic Amplifiers," *Proc. IRE*, vol. 39, March, 1951, pp. 242–251.

flux linking the control coil is the difference between the fluxes produced by the two gate windings. That is $\phi_c = \phi_B - \phi_A$.

The inductance of a coil is defined as the flux linkages per ampere. That is, for the two gate windings,

$$L_A = \frac{d\lambda_A}{dI_A}; \qquad L_B = \frac{d\lambda_B}{dI_B}$$

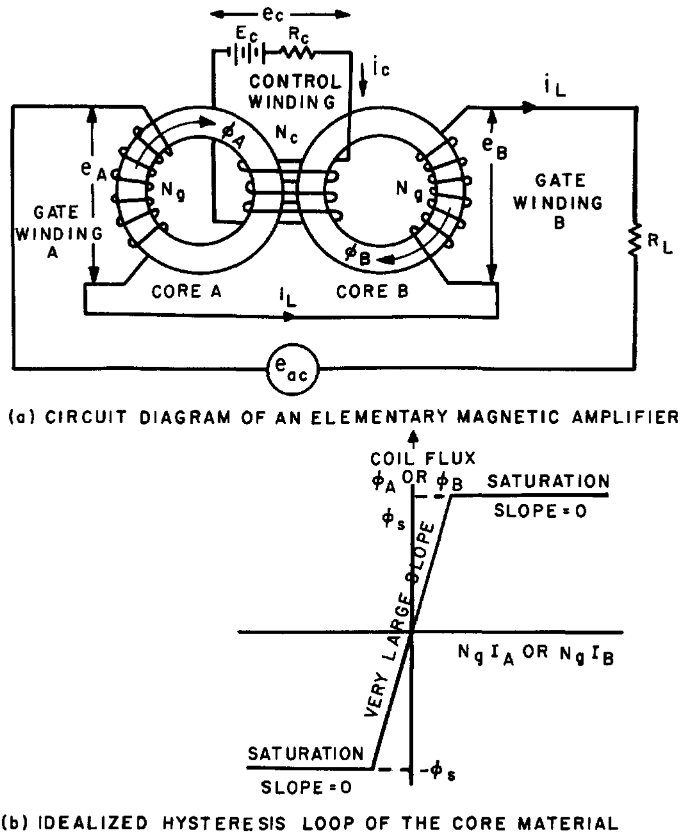


Fig. 16.1. Elementary magnetic amplifier.

where $\lambda_A = N_g\phi_A$ = flux linkages of coil A; $\lambda_B = N_g\phi_B$ = flux linkages of coil B; I_A = current through coil A; I_B = current through coil B. Therefore it is possible to write

$$L_A = N_g \frac{d\phi_A}{dI_A}; \qquad L_B = N_g \frac{d\phi_B}{dI_B}$$

The hysteresis loop given in figure (16.1b) shows the relationship between the flux in either gate winding core to the applied magnetomotive force (mmf). The slope of the hysteresis loop at any point is

$$\text{slope} = \frac{1}{N_g} \cdot \frac{d\phi_A}{dI_A} = \frac{1}{N_g} \cdot \frac{d\phi_B}{dI_B}$$

Therefore
$$\frac{d\phi_A}{dI_A} = \frac{d\phi_B}{dI_B} = N_g \text{ (slope)}$$

Hence the coil inductances are $L_A = L_B = N_g^2$ (slope of hysteresis loop).

An examination of the hysteresis loop of the two cores given in figure (16.1b) shows that there are two distinctly different values of slope:

(1) In the saturation region, the slope $\doteq 0$.

(2) In the unsaturated region, the slope $\doteq$ infinity.

Hence the coil inductances are either zero when saturated or extremely large when unsaturated. Also, when unsaturated, the gate windings are inductively coupled through the core flux to the control winding. Hence for each core we can construct an equivalent circuit of the form shown in figure (16.2a). The complete equivalent circuit of the magnetic amplifier then appears in figure (16.2b).

The voltage induced in any coil is $e = N d\phi/dt$ volts, where $\phi =$ flux in webers; $N =$ number of turns. Alternatively, the flux produced in any core is related to the applied coil voltage by the relationship:

$$\phi = \frac{1}{N} \int e \, dt$$
$$= \frac{1}{N} \text{(area under curve of } e \text{ vs. } t)$$

Thus, for the two gate windings,

$$\phi_A = \frac{1}{N_g} \text{(area under the curve of } e_A \text{ vs. } t)$$

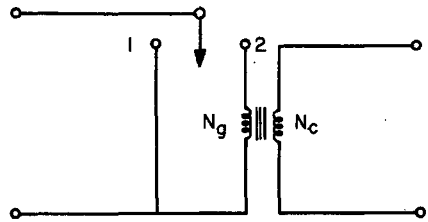
$$\phi_B = \frac{1}{N_g} \text{(area under the curve of } e_B \text{ vs. } t)$$

With zero control voltage and current, the magnitude of the alternating supply voltage is adjusted so that the flux produced in the cores almost, but not quite, reaches the saturation value ϕ_s. Because neither core is saturated, both coils appear as large (nearly infinite) inductances. These inductances have extremely high reactances at the supply frequency so that little current flows through the gate windings and load resistance

R_L. As a result, the supply voltage divides equally between the two coils, so that $e_A = e_B = \frac{1}{2}e_{ac}$ and

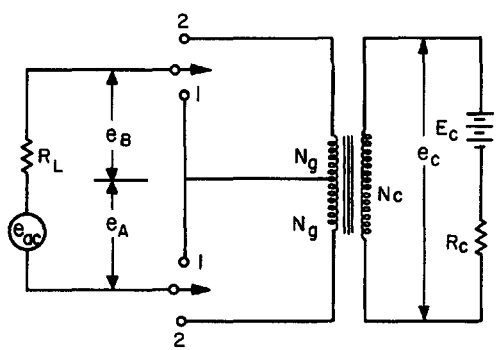
$$\phi_A = \frac{1}{2N_g} \text{(area under the } e_A \text{ vs. } t \text{ curve)}$$

$$\phi_B = \frac{1}{2N_g} \text{(area under the } e_B \text{ vs. } t \text{ curve)}$$



POSITION 1- CORE SATURATED
POSITION 2- CORE UNSATURATED

(a) SINGLE COIL



(b) COMPLETE AMPLIFIER

Fig. 16.2. Equivalent circuits for the elementary magnetic amplifier useful in explaining circuit operation.

Both cores are unsaturated, so the switches in the equivalent circuit of figure (16.2b) are in position 2. If the supply voltage is sinusoidal, the fluxes will be cosinusoidal. Both the load current and control current will be zero.

The effect of control voltage will be explained in the next section.

16.2. Operation with Control Voltage Applied

When voltage is applied to the control winding a small magnetizing current flows. This current produces an mmf N_cI_c that opposes the mmf of coil A, and aids the mmf of coil B. Thus the total mmf's acting on the cores are

$$(NI)_A = \text{total mmf on core } A = N_gI_L - N_cI_c$$

$$(NI)_B = \text{total mmf on core } B = N_gI_L + N_cI_c$$

The hysteresis loops of the two cores should now be drawn as shown in figure (16.3).

Because of the effect of the control current, the core fluxes are *biased* away from zero. The flux in core A is displaced down, while that in B

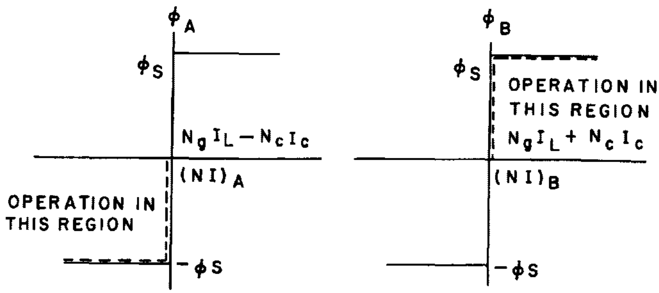


Fig. 16.3. Hysteresis loops of the two cores in the magnetic amplifier of figure (16.1a). Effect of control current appears in the abscissa.

goes up. This is illustrated in figure (16.4). As a result of this displacement of the core flux curves, core A now saturates on the negative swing of the supply voltage and core B saturates on the positive swing. This operating regime would be reversed simply by reversing the direction of the control current.

The various waveforms characteristic of series connected magnetic amplifiers can now be deduced from the equivalent circuit of figure (16.2b). Assume that core A is just coming out of saturation at $t = 0$ as shown in figure (16.4). Thus, as long as both cores are unsaturated, both switches in the equivalent circuit are in position 2. The coil inductances are large and little current flows in either the gate or control windings. This is clear from the hysteresis loops of figure (16.3) because

$$N_gI_L - N_cI_c = 0 \quad \text{for core } A$$

$$N_gI_L + N_cI_c = 0 \quad \text{for core } B$$

in the unsaturated region. This is possible only if $I_c = I_L = 0$. Therefore the supply voltage divides evenly between the two coils, so that $e_A = e_B = e_{ac}/2$.

Under these operating conditions the load circuit is a pure inductance and the flux is a negative cosine function if the supply voltage is sinusoidal.

As e_{ac} increases, the core flux increases with the area under the voltage curve. Eventually, at ωt_1, this increase is sufficient to produce saturation in core B. When this occurs, the switch on coil B in the

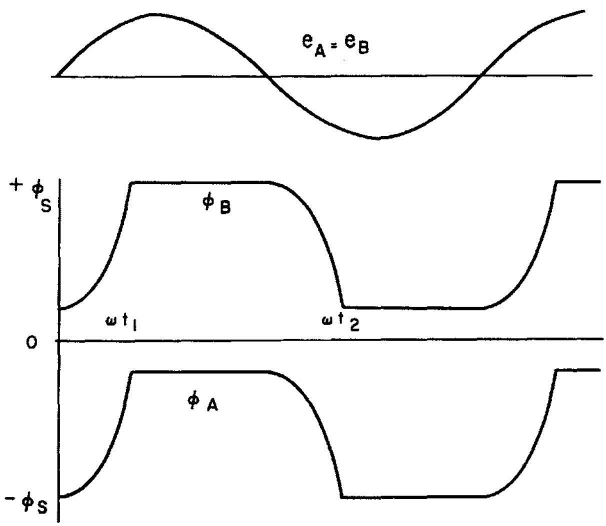


Fig. 16.4. Approximate waveforms of core flux.

equivalent circuit moves to position 1 and e_B drops to zero. Coil A remains transformer coupled to the control winding, so that $e_A = N_g e_c / N_c$. However, the voltage e_c is virtually zero, so that e_A drops almost to zero. This condition remains as long as core B is saturated. Because both coil voltages are practically zero, nearly the entire supply voltage appears across R_L and a large load current i_L suddenly flows. Core A is unsaturated, so that $N_g i_L - N_c i_c = 0$, or $i_c = N_g i_L / N_c =$ control current. Therefore a current is induced in the control winding that has exactly the same form as the load current. The load current is in phase with the supply voltage during this period because the load circuit is purely resistive.

All these statements are illustrated by the waveforms shown in figures (16.4) and (16.5).

Because the load and control currents are in phase with the supply voltage when e_{ac} drops to zero, both i_L and i_c become zero. This operating characteristic prevails in the mode when neither core is saturated, because the only time both currents can be zero is when the cores are unsaturated. Hence core B comes out of saturation, and the switch in the equivalent circuit reverts to position 2.

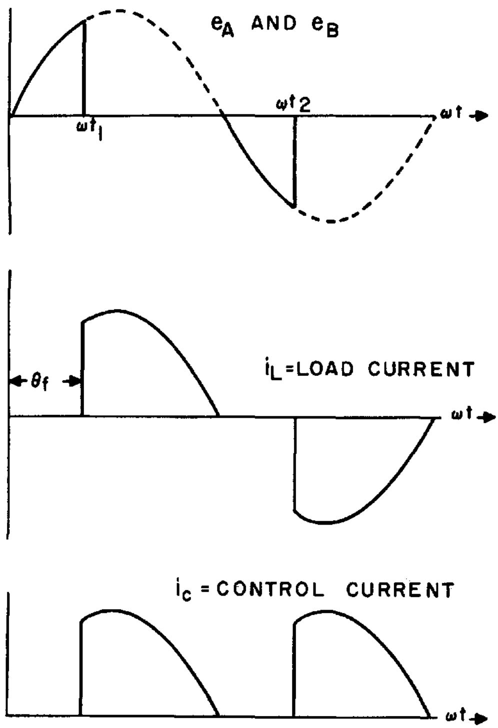


Fig. 16.5. Additional waveforms in the series connected magnetic amplifier.

The coil voltages immediately become equal to each other and to half the supply voltage. The core fluxes commence a cosinusoidal decrease because the volt-second areas of the coil voltages are now negative. Eventually, at ωt_2, core A saturates, and the entire process repeats as before. However, because core B is now unsaturated, $N_g i_L + N_c i_c = 0$, or $i_c = -N_g i_L/N_c$. The load current is reversed from its previous direction, but because of the minus sign in the preceding equation, the control current has the same direction as before. This is shown in figure (16.5).

The instant at which conduction starts in the load circuit is called the *firing angle* θ_f, as shown in figure (16.5).

16.3. Some Amplifier Characteristics

It was shown in the preceding section that the control current is related

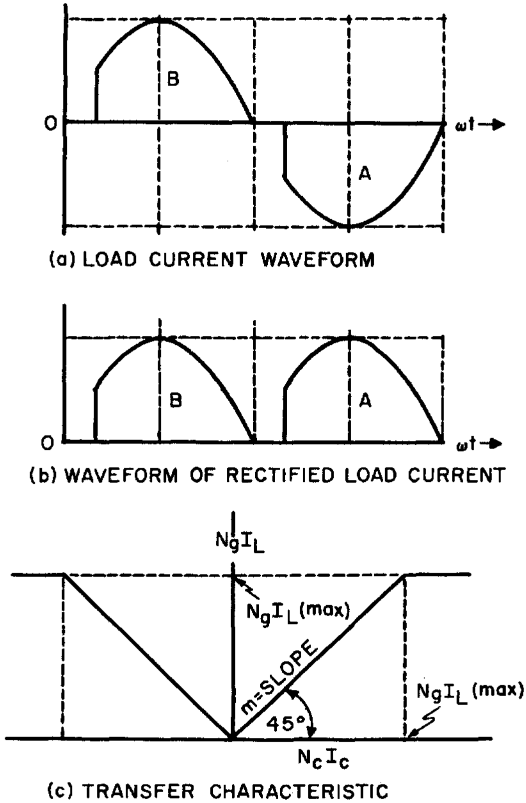


Fig. 16.6. Relationships in the elementary magnetic amplifier.

to the load current by the equation

$$i_c = \pm \frac{N_g}{N_c} i_L$$

when the cores have the idealized hysteresis loops assumed.

Under normal conditions the load current alternates as shown in figure (16.6a), while the control current does not. This accounts for the

$\pm$ sign on the preceding equation. However, if i_L is rectified as shown in figure (16.6b), the relationship between control current and *rectified* load current is $i_c = N_g i_L / N_c$. Although this equation was derived for instantaneous values of i_c and i_L, it is also true for average values because both currents now have the same waveshape. Thus, letting I_c and I_L denote the average, or d-c values of i_c and i_L,

$$I_c = \frac{N_g}{N_c} I_L \quad \text{or} \quad N_c I_c = N_g I_L$$

Obviously this relationship is true only as long as the cores do not saturate at the same time. This means that it is valid for firing angles from 180 to 0 degrees.

When the firing angle is reduced to zero, both cores are saturated. The supply voltage is always applied across the load resistance, so the load current has its maximum value. Further increases in control current will have no effect upon the load current.

With these facts in mind, the transfer characteristic of the amplifier shown in figure (16.6c) is easily drawn. The I_L(max) term designates the average value of the load current when the firing angle is 0 degrees.

16.4. The Transfer Function

The transfer characteristic of the series connected magnetic amplifier is given in figure (16.6c). The slope of the characteristic in the linear

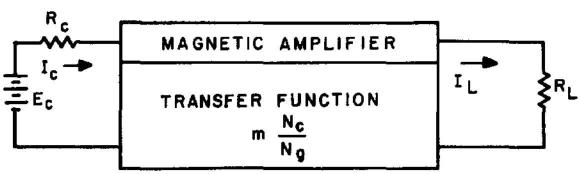


Fig. 16.7. Block diagram equivalent representation.

region is identified by the symbol m. However, in this particular amplifier $m = 1$. In other cases to be discussed later it greatly exceeds 1. So, to accommodate either possibility, let $m =$ slope of the transfer characteristic in the linear region $= N_g I_L / N_c I_c$.

The amplifier input is the average value of the control current I_c, and the output is the average value of the rectified load current I_L. Hence the current transfer function of the amplifier is $I_L / I_c = m N_c / N_g$, and the block diagram equivalent representation appears as shown in figure (16.7). This is a useful representation when the magnetic amplifier is a part of some larger system.

16.5. Bias and Signal Windings

The elementary series connected magnetic amplifier discussed so far is essentially a d-c amplifier with a d-c input and output. It is immensely

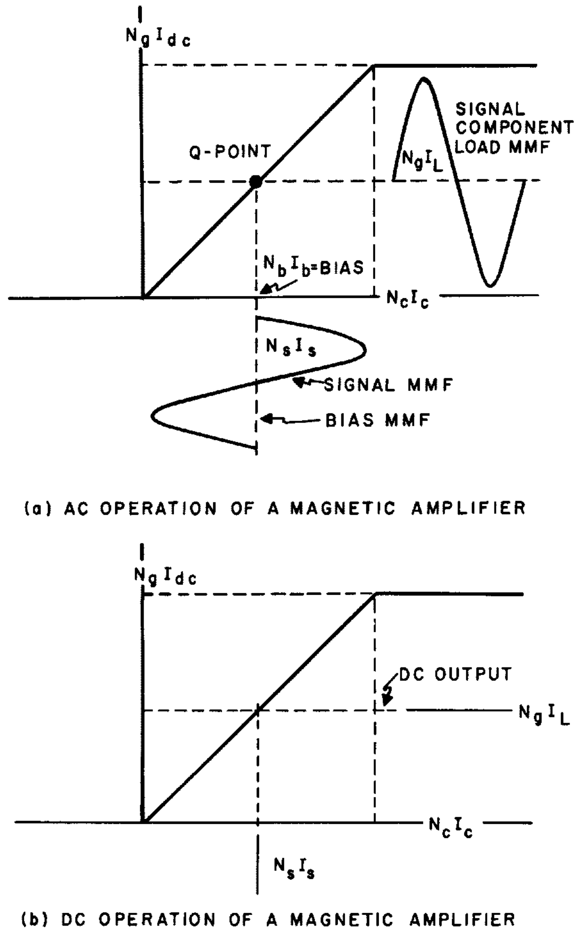


Fig. 16.8. Magnetic amplifier operation.

useful in performing this service, but it also finds extensive use in the amplification of variational signals. In such cases the signal input to the amplifier may not have a d-c component, or the d-c component may not be the value required for operation over the linear region of the transfer characteristic. Therefore it often becomes necessary to *bias* the magnetic amplifier to the desired point under no-signal conditions and apply the signal through a separate input. As a result, the control winding is generally divided into two separate windings, a *signal* winding and a *bias* winding. A subscript s will designate factors relating to the signal winding, and a subscript b will signify quantities associated with the bias winding.

Of particular interest is the fact that the amplifier will have a d-c output in the absence of a signal input because of the action of the bias winding. This is illustrated in figure (16.8). This disadvantage is not present in some advanced types of magnetic amplifier circuits.

16.6. The Use of Rectifiers

Nearly all magnetic amplifiers use one or more rectifiers for various purposes. Even the elementary series connected magnetic amplifier

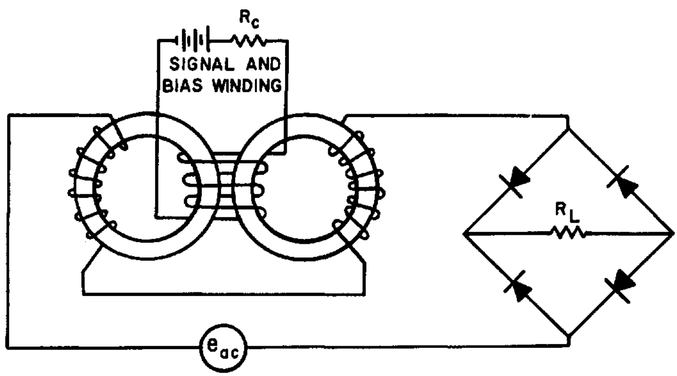


Fig. 16.9. Use of a bridge rectifier to rectify the load current.

discussed earlier required rectifiers to convert the load current into a unidirectional current.

Semiconductor diodes are nearly always used for rectifiers in preference to vacuum and gas diodes. They have low forward resistance, low voltage drop, long life, do not require filament heating power, and are relatively insensitive to shock and vibration. Thus, where rectifiers are shown on the circuit diagrams, it is understood that semiconductor diodes are involved.

The use of these diodes to rectify the load current is shown by the circuit diagram in figure (16.9). The load resistance can usually be placed on either the a-c or d-c side of the rectifier circuit. Figure (16.9) is a complete circuit diagram of a practical series connected magnetic amplifier having the transfer characteristic shown in figure (16.6c).

16.7. Gain and Time Constant, Elementary Magnetic Amplifier

Assume a magnetic amplifier of the type shown in figure (16.9), but having separate bias and signal windings. If a-c amplification is desired, separate inputs are required for both the signal and bias windings. If d-c amplification is desired, the bias winding is open circuited and the signal winding provides the necessary control magnetomotive force. Regardless of the type of amplification required, operation is restricted to the linear part of the transfer characteristic.

Operating conditions in the amplifier for both a-c and d-c conditions were illustrated in figure (16.8). In both cases you will observe that the amplitude of the signal component of load current is designated I_L and the amplitude of the signal input is shown as I_s.

The *power gain* of the amplifier is defined as follows:

$$\begin{aligned} W &= \text{power gain} \\ &= \frac{\text{useful load power}}{\text{power input to signal winding}} \\ &= \frac{P_L}{P_s} \end{aligned}$$

Hence, in the a-c case,

$$P_L = \tfrac{1}{2} I_L^2 R_L; \qquad P_s = \tfrac{1}{2} I_s^2 R_s$$

so that the power gain can be expressed as

$$W = \left(\frac{I_L}{I_s}\right)^2 \frac{R_L}{R_s} \tag{16.1}$$

It is a simple matter to show that exactly the same result is obtained in the d-c case,

However, it was previously shown that the transfer function is

$$\frac{I_L}{I_s} = m \frac{N_s}{N_g}$$

so that the power gain can be written

$$W = m^2 \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{R_s} \tag{16.2}$$

However, in the elementary series connected magnetic amplifier, m, which is the slope of the linear region of the transfer characteristic, is unity, so that the power gain is

$$W = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{R_s} \tag{16.3}$$

To calculate the time constant of the signal winding, the coil inductance must be known. Because this depends upon the flux linking the coil, it is desirable to make a plot of the flux linking the signal winding as a function of the magnetomotive force of the signal winding. The

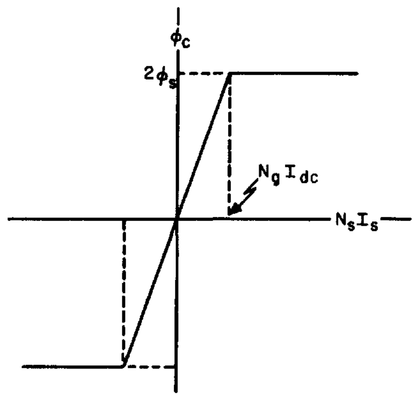


Fig. 16.10. Total average flux linking the signal, or control, winding of a d-c magnetic amplifier.

result is shown in figure (16.10). The I_{dc} appearing on this figure is numerically equal to the I_L(max) defined earlier.

The slope of the curve in figure (16.10) is easily evaluated because both cores are saturated when $N_s I_s = N_g I_{dc}$ and the total flux linking the signal winding is $2\phi_s$ where ϕ_s designates the saturation value of the core flux.

The slope of the characteristic is clearly

$$\text{slope} = \frac{2\phi_s}{N_g I_{dc}}$$

The flux linking the signal winding for any given value of signal winding mmf is therefore

$$\phi_c = N_s I_s \frac{2\phi_s}{N_g I_{dc}}$$

Then the coil inductance is the flux linkages per ampere, or

$$L_0 = \frac{N_s\phi_c}{I_s} = \frac{2N_s^2\phi_s}{N_g I_{dc}} \tag{16.4}$$

where I_{dc} is the maximum average value of rectified load current. The corresponding load voltage is $E_{dc} = I_{dc}R_L$, so that the coil inductance can be expressed as

$$L_0 = \frac{2N_s^2\phi_s R_L}{N_g E_{dc}} \tag{16.5}$$

In normal operation the supply voltage is adjusted so that the flux varies from $+\phi_s$ to $-\phi_s$ in one half cycle in the absence of any control current. Also, the signal voltage divides equally between the two gate coils. For any gate coil voltage, the change in flux through the coil is

$$\Delta\phi_B = \frac{1}{N_g}\int e_B\,dt = \frac{1}{N_g}\,(\text{area under } e_B \text{ vs. } t)$$

$$= \frac{1}{2N_g}\,(\text{area under } e_{ac} \text{ vs. } t)$$

Over one half cycle, ϕ_B varies from $+\phi_s$ to $-\phi_s$, so that $\Delta\phi_B = 2\phi_s$. The area under the e_{ac} vs. t curve over the same interval is obviously $E_{dc}T/2$, where E_{dc} = average value of e_{ac} rectified; $T = 1/f$ = period e_{ac}. Therefore

$$2\phi_s = \frac{1}{4N_g}\cdot\frac{E_{dc}}{f}$$

Hence the maximum value of E_{dc} is

$$E_{dc} = 8N_g f\phi_s$$

Substitute this into equation (16.5) and the value for the signal winding inductance becomes

$$L_0 = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4f} \tag{16.6}$$

where f = supply frequency in cycles per second.

This inductance is in series with the coil resistance, so that the circuit time constant is

$$t_s = \frac{L_0}{R_s} = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4R_s f} \tag{16.7}$$

This is the time constant expressed in seconds. In magnetic amplifier practice it is more common to express this factor in terms of the number of cycles of the supply voltage that elapse during one time constant. That is,

$T_s =$ number of supply voltage cycles in one time constant $= ft_s$

$$T_s = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4R_s} \quad \text{cycles} \tag{16.8}$$

The performance of magnetic amplifiers is characterized by a figure of merit that is the ratio of the power gain of the amplifier to the time constant expressed in cycles. That is, denoting the figure of merit by F_m, then $F_m = W/T_s$. For the elementary series connected magnetic amplifier under discussion, the figure of merit is equal to 4.

This demonstrates a fact of basic importance. Because of its low figure of merit, a high gain amplifier will have a slow response, and a fast response can be achieved only at the expense of the gain. This is also true in vacuum tube amplifiers, but the figures of merit, though defined differently, are much larger.

16.8. Response Time

The presence of inductance in the control or signal winding of the magnetic amplifier requires a slight revision in the equivalent representation previously given in figure (16.7). The arrangement shown in figure (16.11) is more exact.

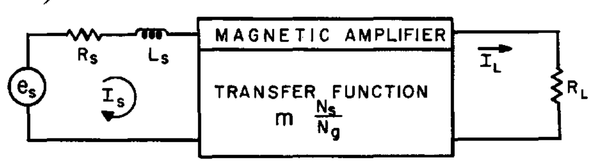


Fig. 16.11. Modified equivalent representation of the elementary magnetic amplifier.

Because of the presence of the inductance and resistance in the signal winding, a finite time must elapse for a change in magnitude of I_s to occur. Because the time constant of the series connected magnetic amplifier, expressed in cycles of the supply frequency, is

$$T_s = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4R_s}$$

then if E_s changes suddenly from one value to another, I_s will change exponentially. At the end of one time constant it will undergo 63% of the total change. Two or three time constants must elapse before the final value is attained for all practical cases. This, if T_s is of the order of 2 or 3 cycles, then 4 or 6 or more cycles must pass before a change in the signal winding exerts its full effect upon the load current. This is a basic limitation on the performance of the elementary series connected magnetic amplifier and was the primary determinant in the development of the special magnetic amplifier types to be discussed later.

16.9. Series Connected Magnetic Amplifier, Inductive Load

Although the effect of control circuit inductance has been considered in time constant and figure of merit calculations, this inductance and that in the load circuit were completely neglected in the determination of

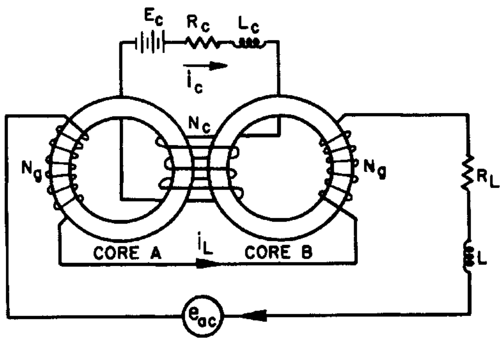


Fig. 16.12. Elementary series connected magnetic amplifier with inductances in the control and load circuits.

circuit waveforms. A detailed solution has been obtained by Wilson.[3] Because the derivations are rather lengthy, only the method of analysis and approach are illustrated here.

The circuit to be analyzed is shown in figure (16.12). To draw the circuit in this form it was necessary to make the following assumptions in line with Wilson's procedure:

(1) The series resistances represent the total circuit resistance including contributions from the coil windings, internal resistances of the generators, and the external or connected resistance.

(2) Leakage between the load and control circuits is negligible.

(3) Interwinding and distributed capacitances can be omitted.

The analysis is greatly simplified if it is assumed that the gate winding

[3] Thomas G. Wilson, "Series Connected Magnetic Amplifier With Inductive Loading," *Communications and Electronics*, vol. 71, January, 1952, pp. 101–110.

cores and coils are identical and have the idealized hysteresis loops shown in figure (16.13). This last assumption is a critical one. However, Wilson shows that actual cores of some types closely approximate this ideal, and where this condition exists, the analysis yields results that compare favorably with experimental determinations.

The equations for the various segments of the idealized hysteresis loops given in figure (16.13) are easily written as follows:

(1) In the unsaturated region

(a) for core A; $N_g i_L - N_c i_c = 0$ (16.9)

(b) for core B; $N_g i_L + N_c i_c = 0$ (16.10)

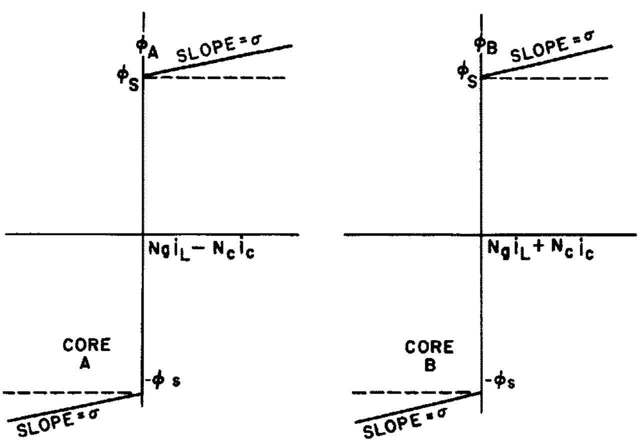


Fig. 16.13. Assumed form of the hysteresis loops of the two cores in the circuit of figure (16.12).

(2) In the saturated region

(a) Core A saturates in the negative flux region, so

$$\phi_A = -\phi_s + \sigma(N_g i_L - N_c i_c) \tag{16.11}$$

(b) Core B saturates in the positive flux region, so

$$\phi_B = +\phi_s + \sigma(N_g i_L + N_c i_c) \tag{16.12}$$

These equations are used later in the derivation.

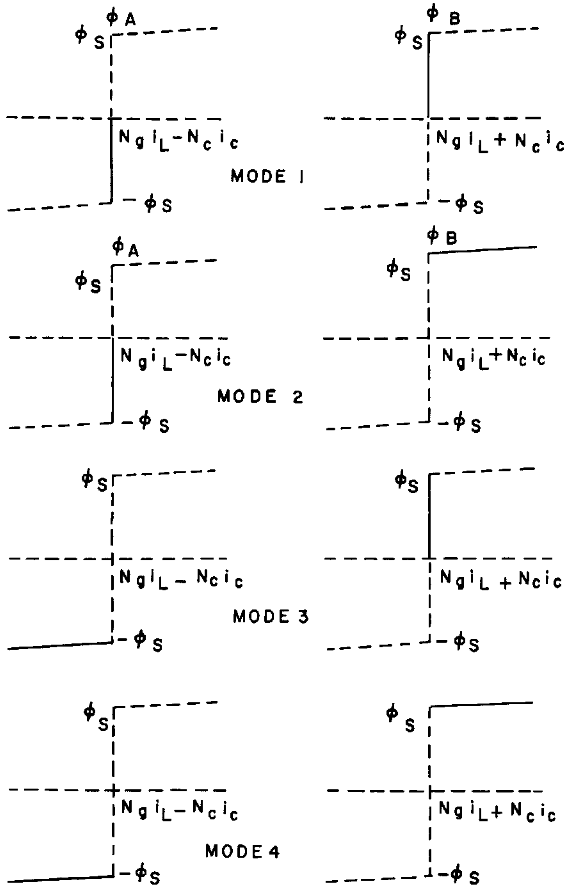


Fig. 16.14. Modes of operation. Solid line indicates the actual region of operation on the hysteresis loop in the various modes.

With the hysteresis loops assumed it is obvious that there are four possible operating possibilities as shown in figure (16.14) and tabulated:

Mode	*Core A*	*Core B*
1	unsaturated	unsaturated
2	unsaturated	saturated
3	saturated	unsaturated
4	saturated	saturated

In normal operation, mode 4 is not usually obtained and the system operates cyclically from mode-to-mode in a 2–1–3–1 sequence.

Regardless of the mode of operation, the Kirchhoff loop equations for the load and control circuits are

$$N_g \frac{d\phi_A}{dt} + N_g \frac{d\phi_B}{dt} + L \frac{di_L}{dt} + R_L i_L = E_{ac} \sin \omega t \tag{16.13}$$

$$N_c \frac{d\phi_B}{dt} - N_c \frac{d\phi_A}{dt} + L_c \frac{di_c}{dt} + R_c i_c = E_c \tag{16.14}$$

These two relationships must be satisfied at all times. At different times operation will be in different modes. Thus a solution of these equations must be obtained for each mode, and the initial conditions must be matched so that transition from one mode to the next is continuous. Actually, the process is rather simple in principle although the actual evaluation is quite complicated.

To illustrate the technique, assume that ϕ_A, ϕ_B, i_L, and i_c are to be computed during mode 2. In this mode core B is saturated, so that from equation (16.9) for the unsaturated core A, $N_g i_L - N_c i_c = 0$, or, in an alternate form

$$i_c = \frac{N_g}{N_c} i_L \tag{16.15}$$

or

$$N_g N_c \frac{di_c}{dt} = N_g^2 \frac{di_L}{dt} \tag{16.16}$$

or

$$N_g N_c \frac{di_L}{dt} = N_c^2 \frac{di_c}{dt} \tag{16.17}$$

Core B is saturated in this mode so substitute equation (16.12) for ϕ_B into the two voltage loop equations (16.13) and (16.14). After some simplification and the use of the three preceding relationships, the following results are obtained:

(1) For the load circuit

$$N_g \frac{d\phi_A}{dt} + (2\sigma N_g^2 + L) \frac{di_L}{dt} + R_L i_L = E_{ac} \sin \omega t \tag{16.18}$$

(2) For the control circuit

$$-N_c \frac{d\phi_A}{dt} + (2\sigma N_c^2 + L_c) \frac{N_g}{N_c} \cdot \frac{di_L}{dt} + R_c \frac{N_g}{N_c} i_L = E_c \tag{16.19}$$

We now have two equations and two unknowns. By using the Laplace transform method and substituting appropriate initial conditions, the necessary solutions can be found.

The same process can be followed for any of the four modes. Space limitations preclude the inclusion of the complete derivation here, but this can be found in the reference to Wilson's work. His experimental results closely approximate his theoretical calculations as long as the core material hysteresis loop approaches the assumed form.

The net effect of these circuit inductances is to remove the abrupt discontinuities in the waveforms of current and flux. The resulting rounding often gives them an appearance that is considerably different from the form predicted by an elementary analysis that does not include the effects of inductance.

16.10. The Use of Feedback

For the general feedback amplifier discussed in chapter 7 it was shown that current-controlled negative feedback could be used to stabilize the output current from an amplifier. That is, the load current is made nearly constant and independent of variations in load resistance. Reduced amplification is the ultimate penalty paid for the increased gain stability. Also, in some special cases, such as vacuum tube amplifiers with cathode degeneration, the current feedback also degrades the high frequency figure of merit.

In some respects the elementary series connected magnetic amplifier has the characteristics of a system involving a large amount of negative feedback. The load current is $I_L = I_c N_c / N_g$, and is clearly independent of the load resistance over the linear part of the transfer characteristic. This suggests the presence of current-controlled negative feedback. Hopefully, if the constant current characteristic of the amplifier can be neutralized, the amplifier figure of merit might be increased. It will be shown that precisely this occurs.

One method of achieving the desired neutralization involves the use of current-controlled *positive* feedback. A magnetomotive force proportional to the load current is added to the magnetomotive forces of the control windings. Thus, considering the control winding to include the signal and bias windings, one more coil, called the *feedback winding*, must be added to the system. The circuit connections are shown in figure (16.15).

In this case the total *control* magnetomotive force is the sum of the mmf's produced by the signal, bias, and feedback windings. That is,

$$(NI)_c = (N_s I_s + N_b I_b) + (N_f I_L) \tag{16.20}$$

or

$$(NI)_c = N_c I_c + N_f I_L \tag{16.21}$$

The transfer characteristic of the elementary series connected magnetic amplifier without feedback is shown in figure (16.16a). In terms of the quantities just defined, this is a plot of the gate winding mmf N_gI_L against the total control mmf $(NI)_c$.

The characteristic of the feedback coil is shown in figure (16.16b) as a plot of the gate winding mmf N_gI_L as a function of the feedback winding mmf N_fI_L. The slope of the characteristic is N_g/N_f.

The transfer characteristic of the feedback magnetic amplifier is a plot of gate winding mmf N_gI_L against the signal and bias winding

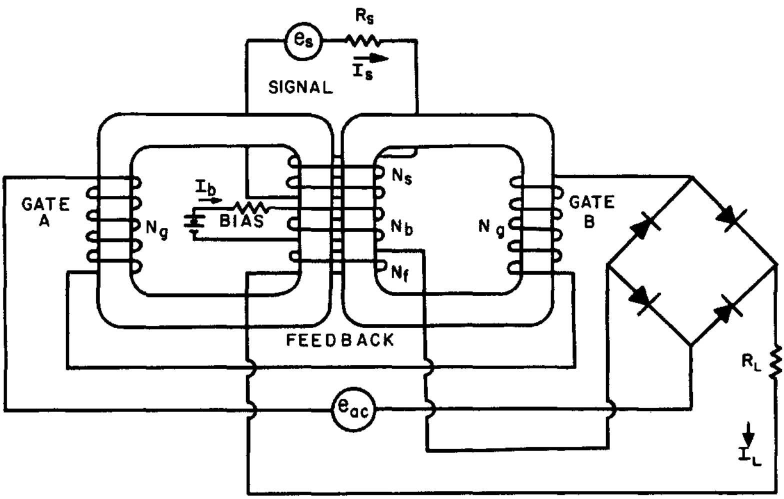


Fig. 16.15. Magnetic amplifier with positive current controlled feedback.

mmf's, $N_sI_s + N_bI_b = N_cI_c$. The result, which is shown in figure (16.16c), can be obtained by combining figures (16.16a) and (16.16b) as follows:

(1) Assume a series of values for N_gI_L.

(2) For each such value assumed, read off the corresponding values of $(NI)_c$ and N_fI_L from figures (16.16a) and (16.16b).

(3) Compute $N_cI_c = (NI)_c - N_fI_L$.

(4) Plot the values of N_cI_c so obtained against the values of N_gI_L assumed. The result is figure (16.16c).

The transfer function of the feedback loop is

$$\beta = \frac{N_fI_L}{N_gI_L} = \frac{N_f}{N_g}$$

If 100% feedback is used, $N_f = N_g$ and $N_cI_c = (NI)_c - N_fI_L = 0$.

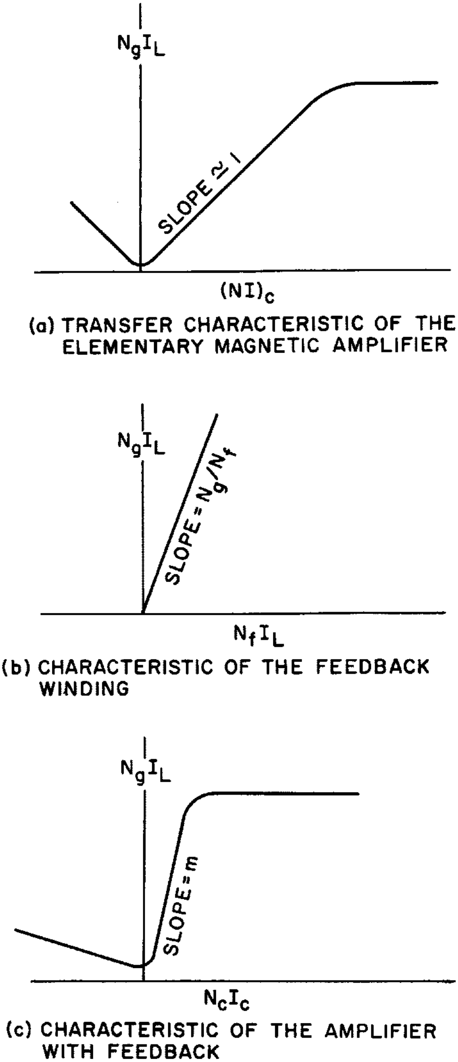


Fig. 16.16. Construction of the transfer characteristic of a magnetic amplifier with positive feedback.

Thus the transfer characteristic of the amplifier with feedback would rise sharply on the right side of the axis, and the slope m would be a large number. In such a case, the value of m is primarily determined by the deviations in the system from the ideal conditions assumed. In the absence of any such limiting factors, m is infinite. However, under actual conditions the operation of the amplifier is governed by the imperfections in the system and an exact analysis is extremely difficult. Even an approximate analysis is lengthy and complicated.

If the feedback exceeds 100%, the transfer characteristic may bend back over itself and the circuit may trigger.

Note in figure (16.16c) that reversing the control current gives a degenerative effect and degrades the performance of the amplifier.

16.11. Figure of Merit with Feedback[4]

The increase in the slope of the transfer characteristic produced by current-controlled positive feedback improves the figure of merit of the

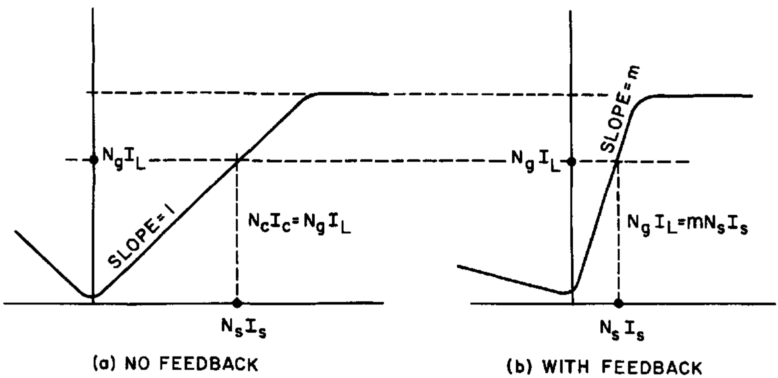


Fig. 16.17. Effect of positive feedback on the transfer characteristic.

amplifier. However, this improvement is not obtained without penalty. As in any positive feedback system the nonlinearities are accentuated and harmonic distortion of relatively high order results.

In section (16.7) it was shown that the power gain of a series connected magnetic amplifier without feedback is

$$W = m^2 \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{R_s}$$

[4] A. O. Black, "Effect of Core Material on Magnetic Amplifier Design," *Proc. Natl. Electronics Conf.*, vol. 4, November, 1948, pp. 427–435.

Because $m = 1$ in an elementary amplifier and is always greater than 1 in feedback circuits, the power gain is increased by the square of the slope of the transfer function characteristic.

In the absence of feedback, the inductances of the signal winding was shown to be

$$L_s = \frac{2N_s^2\phi_s}{N_g I_L} = \frac{2N_s\phi_s}{I_s} \cdot \frac{N_s I_s}{N_g I_L} \tag{16.22}$$

This is the general relationship independent of the presence or absence of feedback. However, in the absence of feedback, $N_s I_s = N_g I_L$ and the signal winding inductance is

$$L_s = L_0 = \frac{2N_s\phi_s}{I_s} \tag{16.23}$$

When feedback is applied, the same gate winding mmf $N_g I_L$ is obtained as in the no-feedback case if the $N_s I_s$ used in the no-feedback case is replaced by $mN_s I_s$. This is illustrated in figure (16.17). If this substitution is made in equation (16.22) for the signal winding inductance, the result is

$$L_s = \frac{2N_s\phi_s}{I_s} \cdot \frac{mN_s I_s}{N_g I_L} = m\,\frac{2N_s\phi_s}{I_s} \cdot \frac{N_s I_s}{N_g I_L}$$

or

$$L_s = mL_0 \tag{16.24}$$

Thus the use of feedback increases the signal winding inductance in direct proportion to the increase in the slope of the transfer characteristic.

It was shown in equation (16.6) that the signal winding inductance in the absence of feedback was

$$L_0 = \left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4f}$$

so that the inductance in the presence of feedback is

$$L_s = m\left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4f}$$

Therefore the signal winding time constant becomes

$$t_s = \frac{L_s}{R_s} = m\left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4fR_s} \quad \text{sec} \tag{16.25}$$

$$T_s = ft_s = m\left(\frac{N_s}{N_g}\right)^2 \frac{R_L}{4R_s} \quad \text{cycles} \tag{16.26}$$

Thus the amplifier figure of merit becomes

$$F_m = \frac{W}{T_s} = 4m \tag{16.27}$$

The figure of merit is linearly dependent upon the slope of the transfer characteristic. With good core materials and without excessive nonlinear distortion, values of m in excess of 100 are possible. The operational advantage of positive feedback is obvious.

16.12. Self-Balancing Magnetic Amplifiers[5]

Ordinarily, at least in the amplifiers discussed here, when feedback is used it is provided through extra windings on the core and does not connect directly back into the control winding. As a result, the effective impedance of the control circuit is affected only by the increase in inductance. Geyger[5] has developed a whole series of amplifiers in which true feedback directly into the control winding is used. The resulting circuits have unique characteristics and will be briefly discussed here.

An elementary form of Geyger's magnetic amplifier is the *self-balancing potentiometer type* shown in figure (16.18). In essence it consists of a series connected magnetic amplifier with separate signal, bias, and feedback windings. However, there are some unique features:

(1) Each load winding has its own rectifier circuit.

(2) The load resistance circuit is conductively coupled directly into the signal winding. This is the main distinguishing feature of the circuit.

It is assumed that the resistor R_s includes the internal resistance of the signal source e_s. If the signal current I_s is much less than the load current I_L the following equation can be written for the signal circuit:

$$E_s - E_k = I_s(R_s + R_k) \tag{16.28}$$

Consequently the signal winding current is

$$I_s = \frac{E_s - E_k}{R_s + R_k} \tag{16.29}$$

and the mmf of the signal winding is

$$(NI)_s = N_s I_s = N_s \frac{E_s - E_k}{R_s + R_k} \tag{16.30}$$

Because the feedback voltage E_k subtracts from the signal voltage E_s, negative feedback or degeneration is introduced.

[5] W. A. Geyger, "Magnetic Amplifiers of the Self-Balancing Potentiometer Type," *Communications and Electronics*, January, 1953, pp. 383–394.

The current flowing in the feedback winding is that fraction of the load current controlled by the current divider consisting of R_p and R_f in figure (16.18). That is,

$$I_f = I_L \frac{R_p}{R_p + R_f}$$

so that the mmf produced by the feedback winding is

$$(NI)_f = N_f I_f = N_f I_L \frac{R_p}{R_p + R_f} \tag{16.31}$$

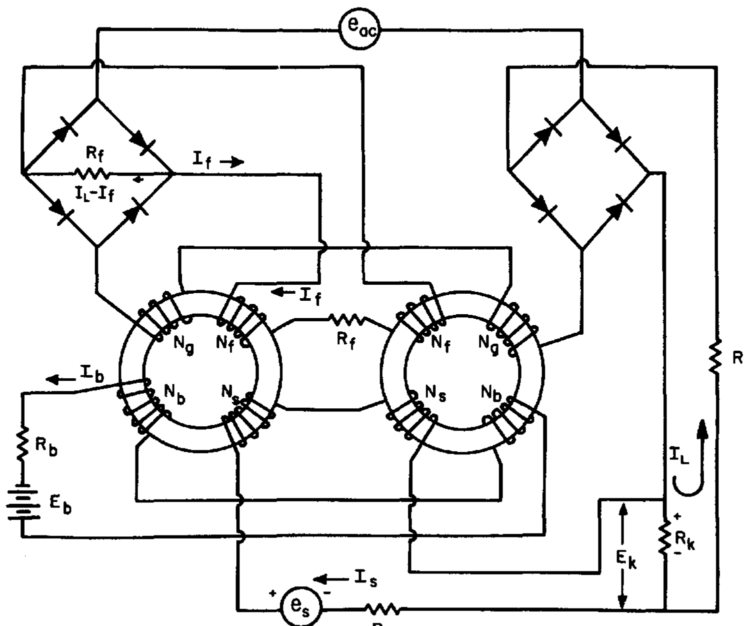


Fig. 16.18. Self-balancing magnetic amplifier of the potentiometer type.

For the case of an ideal rectangular hysteresis loop, the total mmf of the load winding is equal to the total applied mmf of all other windings. Neglecting the d-c components introduced by the bias winding,

$$(NI)_g = (NI)_f + (NI)_s$$

and therefore

$$N_g I_L = N_f I_L \frac{R_p}{R_p + R_f} + N_s \frac{E_s - E_k}{R_s + R_k}$$

However, from figure (16.18), under the assumed relationship between the currents, it is clear that $E_k = I_L R_k$. Hence the preceding equation for the mmf's becomes

$$N_g I_L = N_f I_L \frac{R_p}{R_p + R_f} + N_s \frac{E_s - I_L R_k}{R_s + R_k}$$

Solve this equation for the ratio E_s/I_L. The result is

$$\frac{E_s}{I_L} = R_k + \frac{R_s + R_k}{N_c} \left(N_g - N_f \frac{R_p}{R_p + R_f}\right) \tag{16.32}$$

If 100% feedback is assumed in the feedback winding,

$$N_g I_L = N_f I_f = N_f I_L \frac{R_p}{R_p + R_f}$$

so that

$$N_g = N_f \left(\frac{R_p}{R_p + R_f}\right)$$

As a result, the ratio given in equation (16.32) reduces to

$$\frac{E_s}{I_L} = R_k = \left(\frac{1}{\text{transconductance of circuit}}\right) \tag{16.33}$$

When this condition exists, $E_s = I_L R_k = E_k$, and the signal current is $I_s = (E_s - I_L R_k)/(R_s + R_k) = 0$.

In actual practice the signal current will not be exactly zero mainly because the assumption of 100% positive feedback is not quite met. However, this problem, which is primarily caused by rectifier imperfections, can be largely eliminated by adjustment of R_p.

The unique feature of the circuit is the fact that the signal current required is theoretically zero so that the input impedance is theoretically infinite. In practice, $R_{in} = E_s/I_s$ is usually 100 to 1000 times larger than $(R_s + R_k)$.

Equation (16.33) for the transconductance of the circuit shows that operation is independent of R_L, E_{ac}, and the supply frequency. This is generally true over a particular range of values.

The small signal winding current required produces a nearly linear input-output characteristic.

The circuit has many modifications, variations, and applications that are discussed in Geyger's article.

16.13. Self-Saturating Amplifier Circuits[6]

From the preceding discussion you can see that real operational advantages can be achieved by using large amounts of positive, current-controlled feedback. It is possible to achieve the effect of 100%

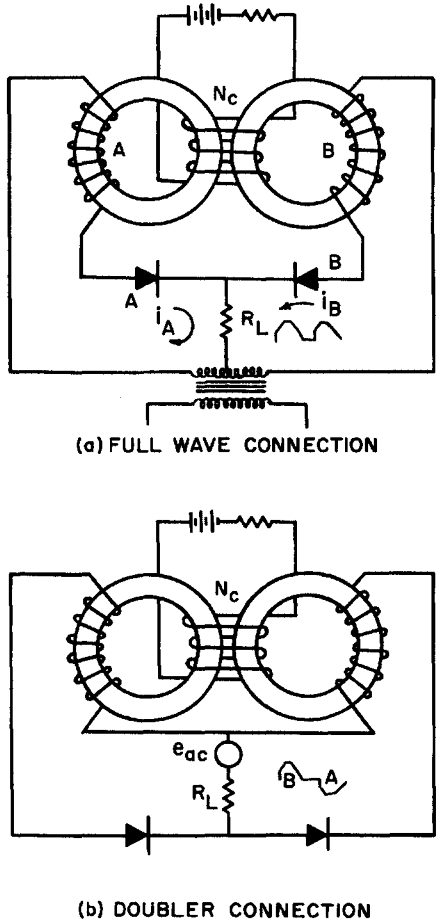


Fig. 16.19. Some self-saturating magnetic amplifier circuits.

feedback by using *self-saturating* amplifiers, and the circuit is somewhat simpler in operation. Several schemes are possible, but all involve

[6] Sidney B. Cohen, "Analysis and Design of Self-Saturable Magnetic Amplifiers," *Proc. IRE*, vol. 39, September, 1951, pp. 1009–1020.

removing the transformer coupling between the gate winding and control winding when the cores are not saturated. This reduces the control current required to a small figure that is zero in the ideal case. The effect is accomplished through the use of semiconductor diodes operating as synchronous switches.

For example, consider the *full wave* self-saturated amplifier shown in figure (16.19a). Note the placement and *polarity sense* of the diode switches. When core A saturates, diode A conducts, while diode B does not. This is apparent from a consideration of the voltage relationships involved. Because diode B is an open circuit, no current flows through coil B and there is no transformer coupling to the control winding. Thus the control current is essentially zero. The converse situation occurs when B saturates. Thus the only control current required is the magnetizing current, and this is zero in the ideal case. This is directly analogous to the feedback amplifier with 100% feedback. Of course, the control current is never zero, but it is small enough so that the slope of the transfer characteristic is large.

Another self-saturating magnetic amplifier is the *doubler* circuit shown in figure (16.19b). Its operation and characteristics can be explained in the same general way as the full-wave circuit because the same general principle is involved.

A detailed discussion and analysis of these circuits has been made elsewhere by Finzi, Pittman, and Durand[7] to show the separate influences of design parameters that affect the self-saturating magnetic amplifier.

16.14. Single Core Magnetic Amplifiers[8]

All the magnetic amplifiers discussed so far used two ferromagnetic cores. It has been found that single core circuits can be used and faster response times result. A typical circuit is shown in figure (16.20).

The operation of this circuit can be explained if both the signal and supply voltages are assumed to be sinusoidal and of the same frequency and phase. It is also assumed that the signal is rectified so that it appears as shown in figure (16.21).

When e_{ac} is positive with respect to ground, diode 2 conducts. The

[7] L. A. Finzi, G. F. Pittman Jr., H. L. Durand, "The Effective Feedback Ratio of Magnetic Amplifiers," *Communications and Electronics*, April, 1952, pp. 157–164.

[8] R. A. Ramey, "On the Mechanics of Magnetic Amplifier Operation," *Trans. AIEE*, vol. 70, Part II, pp. 1214–1223; also "The Single Core Magnetic Amplifier as a Computer Element," *Communications and Electronics*, January, 1953, pp. 442–446.

core magnetization is assumed to be at some point ϕ_0 as shown in figure (16.22). Thus only a negligibly small magnetizing current flows until saturation occurs. When this happens, the full supply voltage is applied across R_L and a large current flows. The core is saturated to

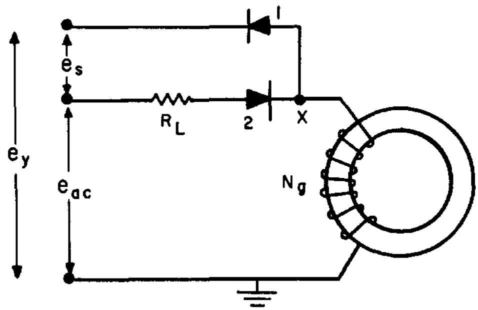


Fig. 16.20. A single core magnetic amplifier.

point c as shown in figure (16.22). As the supply voltage drops to zero, the coil current and mmf also drop to zero, so that at time t_1, the core flux has a value as at point d in figure (16.22). During this entire period

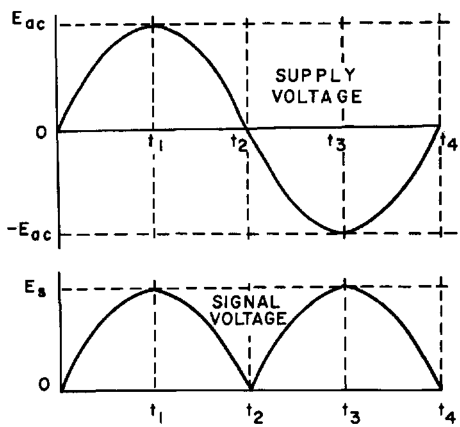


Fig. 16.21. Waveforms of supply and signal voltages.

the voltage across diode 1 has a polarity such that diode conduction is impossible.

On the negative swing of e_{ac}, diode 2 does not conduct and point x in the circuit diagram is at ground potential. The potential at point y, with respect to ground, is

$$e_y = e_s + e_{ac} = \text{drop across diode 1}$$

However, during this half cycle,

$$e_y = E_s \sin \omega t - E_{ac} \sin \omega t$$
$$= (E_s - E_{ac}) \sin \omega t \qquad (16.34)$$

Because E_s is less than E_{ac}, the voltage e_y is negative with respect to ground, and diode 1 conducts. The resulting current through the coil is in the reverse direction from the previous case. Thus the core flux is reduced from the saturation point to the value ϕ_0 at time t_2. This is called the *reset* process because the core flux is reset to its original value. The system is then ready to repeat the process.

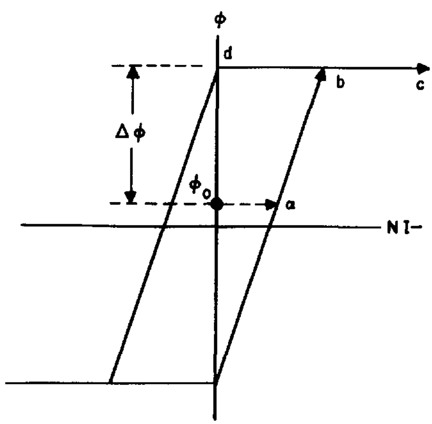


Fig. 16.22. Hysteresis loop of the core in the circuit of figure (16.20).

The essential point is that the core flux is brought down from saturation by an amount

$$\Delta\phi = \frac{1}{N_g} \text{(area under the } e_y \text{ vs. } t \text{ curve)}$$

Therefore it is clear that the more nearly E_s approaches the size of E_{ac}, the smaller the value of $\Delta\phi$ and the smaller the subsequent flux change before saturation and firing on the next half cycle. Hence decreasing the signal voltage increases the firing angle as in other magnetic amplifiers. However, if the change is made prior to time t_1, when diode 1 is nonconducting, the change is effective in controlling the firing angle immediately after t_2. There is no exponential rise in output because the control circuit is noninductive when the change is made in the control voltage. The output reaches its *full* value a *discrete* instant of time after a change in signal voltage. This time is evidently equal to about one half period of the supply frequency.

This response is considerably faster than that obtainable from two-core magnetic amplifiers. The relative speed of the single core circuit may be 6 or more times higher than that of a two core amplifier.

Another type of single core amplifier is shown in figure (16.23).

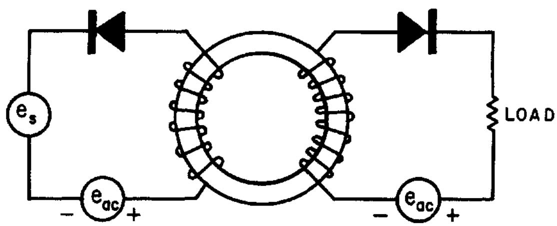


Fig. 16.23. Single core magnetic amplifier double winding type.

16.15. Cascading

It has been shown (see Reference 6) that important improvements in amplifier figure of merit can be achieved by cascading stages. This improvement results because, when three or more stages are involved, the total input time constant is practically equal to the sum of the input time constants of the individual stages. Thus $T_n = nT_s$ if all stages are the same.

The power gains multiply, so that the over-all power gain of an n stage cascade is $W_n = W_s^n$. Therefore, the over-all figure of merit of the cascade is

$$F_m \text{ (over-all)} = \frac{W_n}{T_n} = \frac{W_s^n}{nT_s}$$

The over-all gain increases more rapidly than the over-all time constant, so that the figure of merit is increased. The cascading cannot be continued indefinitely, however, because a point of diminishing returns eventually sets in just as in vacuum tube amplifier cascades.

Chapter 17

WAVE SHAPING AND COMPUTING CIRCUITS

Wave shaping circuits are extensively used throughout the electronics industry, in electronic computers, radar systems, television, nuclear instrumentation, pulse communication, measuring instruments, and so on indefinitely.

These circuits *shape* waves in various ways to achieve special waveforms from relatively standard input signals. Typical outputs might assume the form of sawtooths, pulses, exponential spikes, trapezoids, parabolas, pulsed oscillations, square waves, and so on. However, similar circuits are also used in other ways to perform mathematical operations, time selection, counting, and to indicate the coincidence of two or more events.

The subject is difficult to present because it is nearly impossible to classify the circuits in a consistent system. Classification is often made according to the use of the circuit. This leads to undue repetition because the same circuit often fulfills a number of different functions.

The emphasis in this book is upon the method of *formulating* circuits. Thus, as in all preceding chapters, we will construct the equivalent circuit of each of the various circuits and then derive suitable formulas from this equivalent. A number of circuits are worked out in detail to illustrate the method to be followed in attacking circuits that are not covered here or elsewhere.

17.1. Diode Clippers

It is often desirable or necessary to remove or clip off a part of some specified waveform. Circuits providing this function are called *clippers* or *amplitude limiters.*

In any clipping circuit the operating point of the electronic component is forced to traverse such a wide range that the device is alternately conducting and nonconducting. Thus, operation in the switching mode is obtained and diodes, triodes, pentodes, varistors, and transistors can all be made to operate as clippers by suitable adjustment of their operating conditions. The simple *diode clipper* provides a convenient illustration of the method of clipping.

Two representative diode clippers are shown in figure (17.1). The

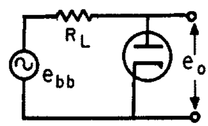
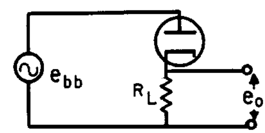
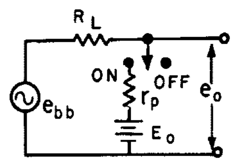
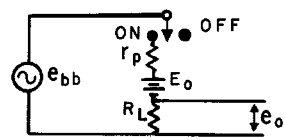
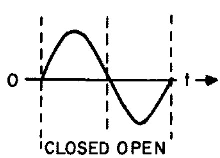
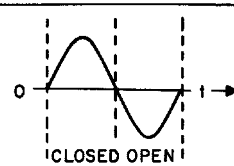
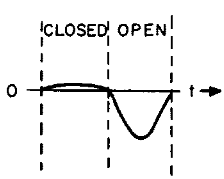
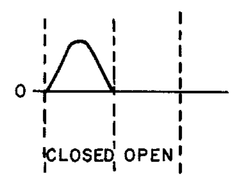
	POSITIVE CLIPPING	NEGATIVE CLIPPING
CIRCUIT DIAGRAM		
VOLTAGE SOURCE EQUIVALENT CIRCUITS		
OUTPUT VOLTAGE, SWITCH OPEN TUBE NON-CONDUCTING	$e_o = e_{bb}$	$e_o = 0$
OUTPUT VOLTAGE SWITCH CLOSED, TUBE CONDUCTING	$e_o = R_e\left(\frac{e_{bb}}{R_L} + \frac{E_o}{r_p}\right)$ where $R_e = \frac{r_p R_L}{r_p + R_L} = \frac{r_p}{1 + r_p/R_L}$	$e_o = (e_{bb} - E_o)\frac{R_e}{r_p}$ where $R_e = \frac{r_p R_L}{r_p + R_L} = \frac{r_p}{1 + r_p/R_L}$
ASSUMED SIGNAL INPUT VOLTAGE e_{bb}		
OUTPUT VOLTAGE WAVEFORM e_o		

Fig. 17.1. Diode clipper computations.

diode may be a vacuum tube or varistor. The load resistance can be connected in series with either lead of the diode. Also, the polarity sense of the diode may be reversed from that shown. Thus a number of different combinations and possibilities are possible. It should also be noted that the signal source impedance is seldom zero in practical cases and this may provide all or part of the required diode load resistance.

Regardless of the circuit connection, the diode is nonconducting and the switch in the equivalent circuit is in the open position as long as the cathode is more positive than the plate. The diode conducts and the switch closes when the plate-to-cathode voltage is positive and larger than E_0, the intercept voltage.

If the diode is replaced by its equivalent circuit as shown in figure (17.1), it is a simple matter to compute the output voltage waveforms for any assumed input. The entire solution is carried through in figure (17.1) for a given signal input.

For the cases shown in the figure, clipping commences at practically zero volts. This *clipping level* can be adjusted to nearly any desired point by biasing the diode.

The equations for the output voltage given in figure (17.1) show that the clipping action will be most effective when the ratio r_p/R_L is much less than 1. Thus large values of R_L should be used, or low r_p diodes, or both. Germanium diodes make excellent clippers because they have plate resistances from only a few ohms to a few hundred ohms, values that are much less than those in vacuum diodes. It is also desirable to use diodes with intercept voltages E_0 that are small in comparison with the peak value of e_{bb}.

The clipping action is generally more effective when the output voltage is taken across the load resistance as shown in figure (17.1). Some signal leakage into the output always occurs when the output is taken across the diode, because r_p is not zero.

17.2. Triode and Pentode Clippers

The plate current in a triode or pentode can be cutoff by the grid voltage so that these tubes can operate as clipping circuits. This type of clipping is called *plate current cutoff clipping*. Such circuits may be preferred to diodes because amplification of the clipped signal is obtained without the addition of another tube.

The basic mechanism involved in cutoff clipping is shown in figure (17.2), which indicates the necessary relationships between signal voltage, bias, and cutoff. As long as the total grid voltage $e_c = e_g - E_{cc}$ is more negative than cutoff E_{co}, the plate current is zero and the output voltage is equal to E_{bb}. When the grid voltage is less negative than the cutoff voltage the tube conducts and the output voltage drops. Actual computations are easily made from the equivalent circuit. A typical solution is carried through in the left half of figure (17.5). The clipping point is changed by variations in grid bias or cutoff voltage.

The equations developed for the cutoff clipper of figure (17.5) show that the output voltage involves a d-c component and a variational component. In some cases it is convenient to use E_{co} as a reference point and measure the positive going signal excursions from this point.

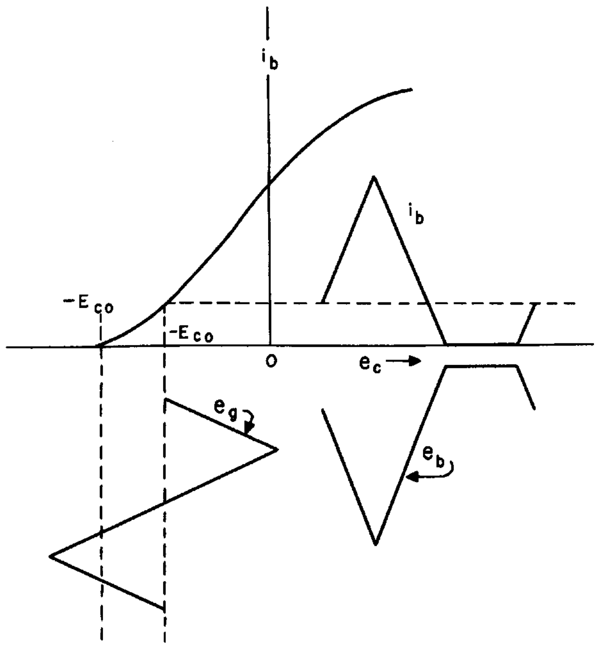


Fig. 17.2. Mechanism of plate current cutoff clipping.

Specifying this voltage as e_s, the equation for e_o in figure (17.5) for the cutoff clipper is

$$e_o = -e_s g_m R + E_{bb}$$

However, for the circuit shown,

$$R = Z_m = \text{mutual impedance of the plate circuit}$$

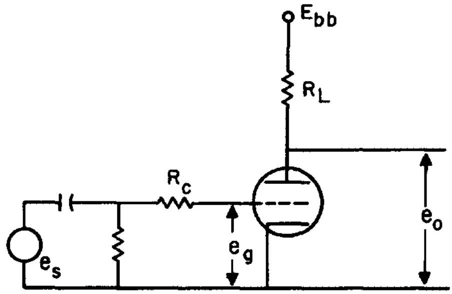
Hence

$$e_o = -e_s(g_m Z_m) + E_{bb}$$

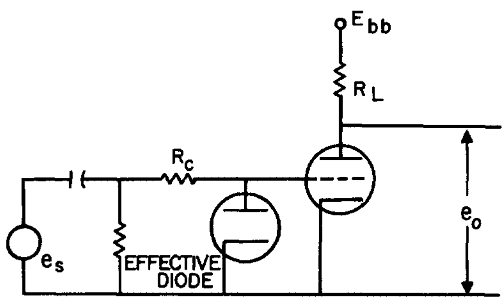
$$= E_{bb} - e_s A_r \tag{17.1}$$

where

$$A_r = g_m R.$$

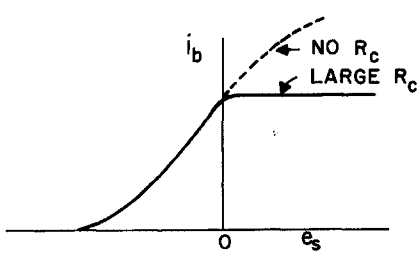


(a) GRID CLIPPING CIRCUIT



(b) EQUIVALENT REPRESENTATION OF PART (a)

17-3



(c) REPRESENTATION IN TERMS OF THE TRANSFER CHARACTERISTIC

Fig. 17.3. Grid circuit clipping.

The grid-to-cathode part of a multielectrode tube behaves in much the same manner as a diode. Thus diode type clipping is easily obtained by inserting a high resistance in series with the grid lead of the tube as shown in figure (17.3a). It is clear that the grid-to-cathode circuit is exactly the same as the diode clipper of figure (17.1a); the same considerations apply. This equivalence is illustrated by the circuit of figure (17.3b) where the effect of the grid is indicated by an equivalent diode. An alternative representation is shown in figure (17.3c). For

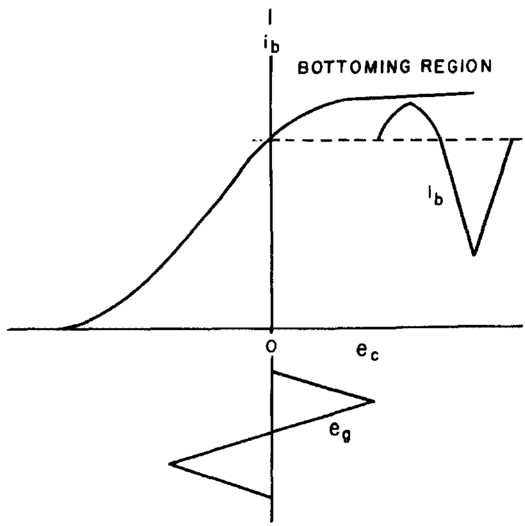


Fig. 17.4. Bottoming in a pentode.

proper operation, R_c should be much larger than the grid-to-cathode resistance r_g. The operation is called *grid circuit clipping*.

A similar type of clipping is possible with pentodes, but the grid clipping resistor is not required. The tube is operated so that it bottoms on the positive grid swing; this achieves the same effect as grid circuit clipping. The dynamic transfer characteristic of a pentode with a large load resistance shows the bottoming region clearly. The mechanism involved in pentode bottoming is shown in figure (17.4).

Suppressor grid clipping is also possible in pentodes. Special tubes, such as the 6AS6, are designed with sharp cutoff characteristics for the suppressor grid and can be successfully operated as suppressor cutoff clippers.

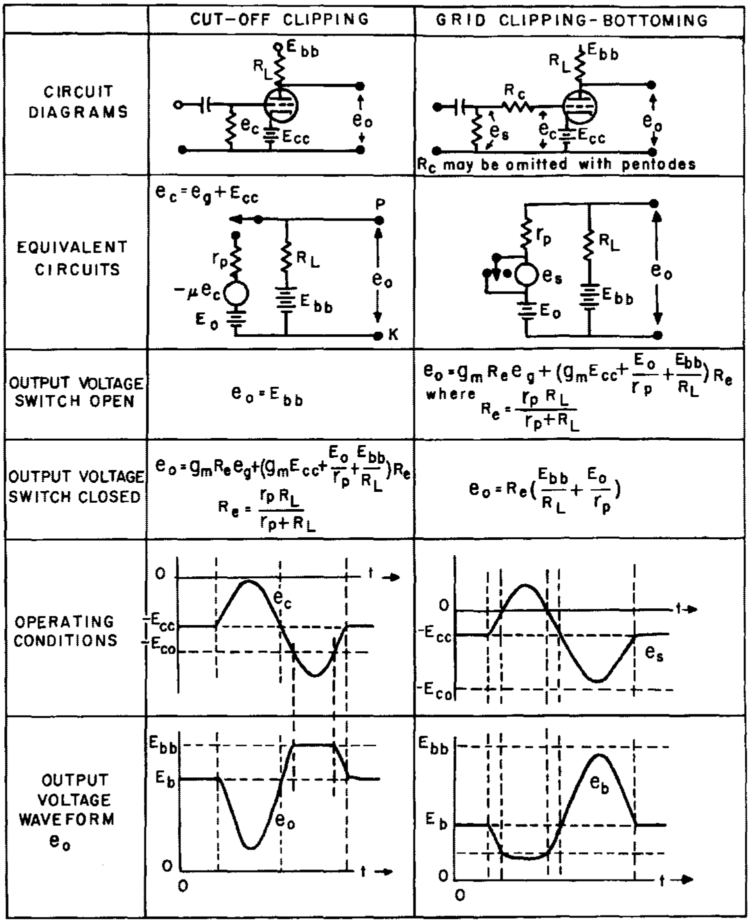
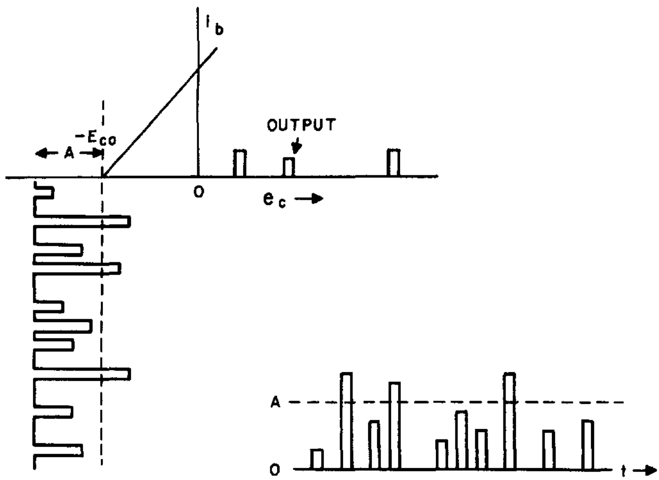


Fig. 17.5. Clipping in triodes and pentodes.

17.3. Some Applications of Clipping Circuits

The various clipping circuits find an enormous field of application in electronic systems. They are often used in simple wave shaping applications where, for example, a sine wave is to be converted to a square wave. This is done by cascading triode or pentode clippers. The output from such a system will not be a perfect square wave; there will always be some finite slope to the edges because of its sinusoidal origin. The ultimate limit to the steepness of the edges will be controlled by the same factors as those affecting rise time in pulse amplifiers. Other signal waveforms can be shaped in a similar manner. Exponentials can be clipped and converted into nearly triangular pulses. The possibilities are almost limitless.

Clippers are often used for *amplitude selection* purposes. For example, it might be desirable to transmit only those signals rising above a certain specified level as shown in figure (17.6). By simply biasing any cutoff clipper so that the critical level coincides with the point of non-conduction, the desired amplitude selection is achieved. The principle involved should be clear from figure (17.6).



(b) OPERATING CONDITIONS REQUIRED

(a) ASSUMED SIGNAL PATTERN; ONLY THOSE PULSES RISING ABOVE A ARE TO BE TRANSMITTED

Fig. 17.6. Amplitude selection by clipping.

If a sawtooth voltage is applied to the clipper input, the time of initiation of current in the tube can be made to vary as a function of time. The principle is illustrated in figure (17.7). Changes in the base-line of the sawtooth brought about by grid bias variations, cause the output to appear at a time controlled by the grid bias. *Time modulation* can be produced this way.

The cutoff characteristics of electronic devices are widely used for *time selection* purposes. For example, only those signals occurring over a particular time interval may be of interest. By turning the tube or diode off at all times except for the particular period of interest, time selection is accomplished.

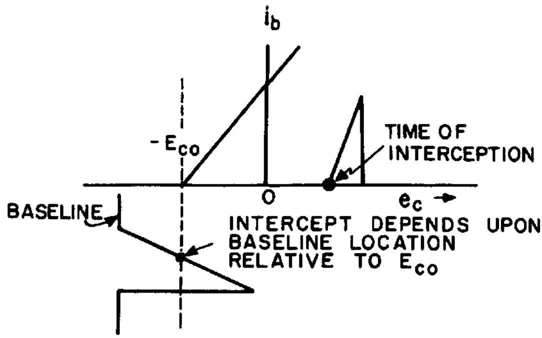


Fig. 17.7. Time modulation mechanism.

17.4. Direct Current Removal and Restoration

The signals appearing at the output electrode of a vacuum tube or transistor ordinarily contain d-c components. This is caused by the polarizing potentials that are applied. In nearly all cases it is desirable to remove the d-c component before the signal is passed on to the next

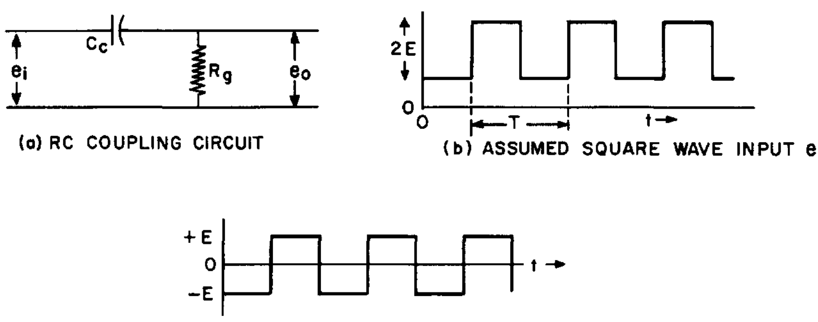


Fig. 17.8. D-c removal in an *RC* coupling circuit.

electronic circuit, because the bias relationships of the following stage might be upset if the d-c component is allowed to remain. Often *RC* circuits of the form shown in figure (17.8a) are used for this purpose, though bifilar wound coils are used in certain special cases. Because the d-c component of the input signal is blocked by the coupling capacitor in the circuit of figure (17.8a), the average voltage appearing across R_g is zero. If the time constant of the coupling circuit is large compared with the half period of the input, the output appears as shown in figure (17.8b). The d-c component of the output voltage waveform is zero because the volt-second areas above and below the reference axis are equal.

The removal of the d-c component from the signal by the coupling circuit is an important effect, particularly in the operation of clipper circuits where relative voltage levels are extremely important. An illustration of how this effect can disrupt normal circuit operation is shown in figure (17.9). The small signal pulse, superimposed on a pedestal, is to be picked off by clipper action. If the pedestal width changes, as shown in figure (17.9), the baseline of the waveform drops and the desired pulse is lost.

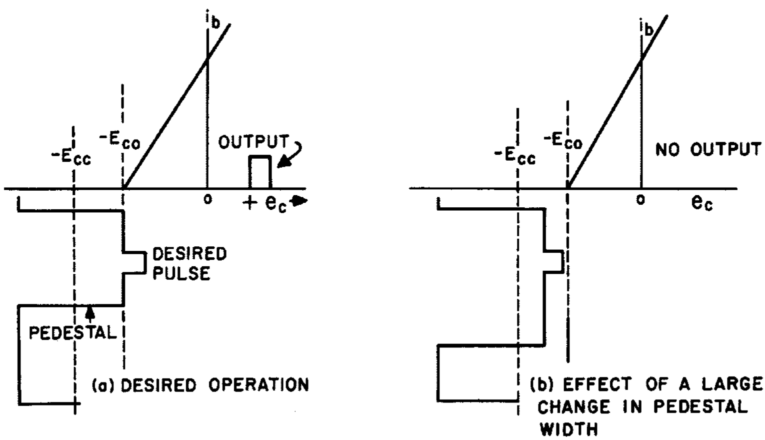


Fig. 17.9. Possible undesirable effect resulting from the removal of the d-c component of a wave by the coupling circuit.

This baseline shift caused by the removal of the d-c component is usually undesirable. Clipper stages operate best when the input signal baseline is independent of any changes in the size or shape of the waveform. Circuits that stabilize the baselines of waveforms are called *d-c restorers* or *clampers*. Simple diode clamping circuits are shown in figure (17.10).

A clamper consists of a conventional *RC* coupling circuit with the gridleak paralleled by a diode. The two parallel elements are connected to a bias voltage E_{cc} equal to the desired baseline voltage. In the absence of any signal input to the circuit of figure (17.10a), the output voltage is equal to E_{cc}. If a positive going signal is applied to the input, the circuit behaves in the normal manner because the diode does not conduct. However, a negative going signal causes the diode to conduct, and C_c rapidly charges to E_{cc} through the low diode resistance r_p. Thus the baseline is clamped at E_{cc} and only positive excursions of signal voltage are permitted. The circuit of figure (17.10b) can be analyzed in a similar manner.

It should be clear that proper clamping is possible only as long as the charging time constant r_pC_c is much less than the time constant RC_c of the coupling circuit. Hence it is necessary that R be much larger than r_p, and low r_p diodes should be used.

There are two instances in which clampers can be detrimental to circuit operation. The most obvious problem occurs in the low frequency response of amplifiers which will be affected by the nonlinear

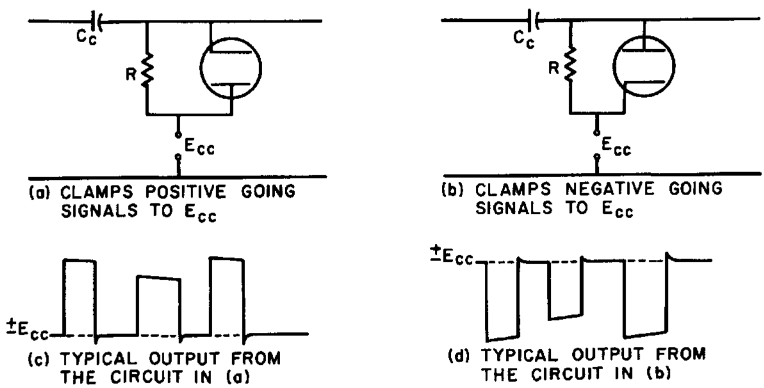


Fig. 17.10. Diode clamping circuits.

grid leak composed of R_g and the parallel diode. The other difficulty is discussed elsewhere.[1]

17.5. The Switch Tube

The simple clipping and clamping circuits discussed so far are not always sufficient to meet the varied demands for the generation of unusual waveforms or for the performance of specialized functions. Hence the switching properties of electronic components are combined with the transient response characteristics of simple *RC*, *RL*, and *RLC* circuits. The cascading of these circuits with clippers makes an endless array of waveshapes and functions available.

It is characteristic of these circuits that the electronic components

[1] S. N. Van Voorhis, *Microwave Receivers*, vol. 23, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, p. 232.

nearly always operate as simple switches. Class A operation may be used, but only occasionally.

For this application of vacuum tube circuits the grid-to-cathode signal usually approximates a square wave of large amplitude, and the tube is either cut off or highly conducting; this is shown in figure (17.11). This operating regime is usually obtained by driving the grid with an approximately square signal of large amplitude and then, by using grid circuit clipping, the grid-to-cathode voltage is effectively clamped at 0 volts. When the tube is conducting, the grid voltage e_c is virtually zero. As a result, the generator $\mu e_c = \mu(e_g - E_{cc})$ in the equivalent plate circuit is also zero, and a simplified equivalent circuit can be drawn as shown in figure (17.12). The function of the tube is reduced to that of a simple switch, alternately connecting the external circuit between two different voltage sources.

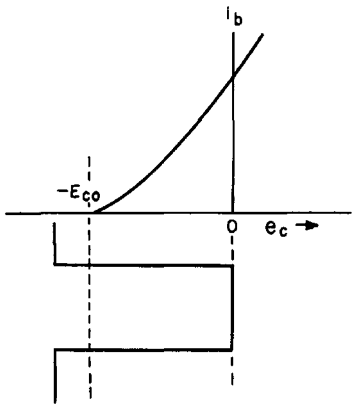
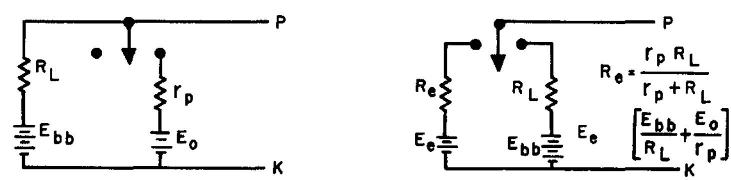


Fig. 17.11. Switch tube operation.

This is called *switch tube operation* in this book. The particular advantage of switch tube operation over class A operation is that the response characteristics become more nearly independent of variations in tube parameters.



(a) DIRECT EQUIVALENT CIRCUIT

(b) SIMPLIFIED EQUIVALENT CIRCUIT

Fig. 17.12. Switch tube equivalent circuits.

17.6. Transient Response of *RC* Circuits

It was shown in the preceding section that vacuum tubes can be made to operate as switches, alternately connecting an external circuit to two different direct voltage supplies with different internal impedances. The external circuit will usually assume one of three forms: a series *RC*

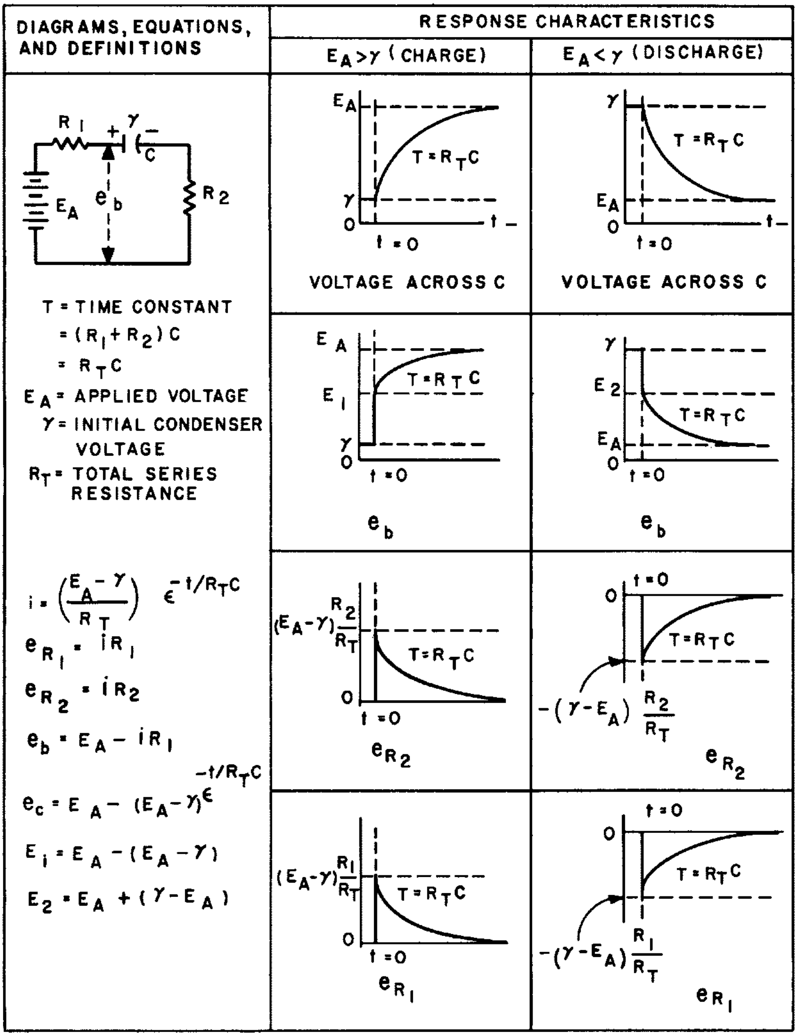


Fig. 17.13. Characteristics of *RC* circuits.

circuit, series RL circuit, or a parallel RLC circuit. The switching transients in these circuits caused by the operation of the switch tube are the response characteristics of interest.

The determination of the transient responses of these circuits is straightforward. For example, consider the series RC circuit shown in figure (17.13). The circuit equation is

$$E_A = i(R_1 + R_2) + \frac{1}{C}\int i\,dt$$

The Laplace transform of this equation is written by the method given in chapter 2 as

$$\frac{E_A}{s} = I(s)\left(R_1 + R_2 + \frac{1}{sC}\right) + \frac{\gamma}{s}$$

Solve this equation for the transform current.

$$I(s) = \frac{E_A - \gamma}{R_T} \cdot \frac{1}{s + 1/R_T C}$$

where $R_T = R_1 + R_2$; $\gamma =$ initial capacitor voltage (note polarity sense); $E_A =$ applied voltage. The inverse transform is given by pair no. 8 in the table of function-transform pairs. Hence

$$i(t) = \frac{E_A - \gamma}{R_T}\,\varepsilon^{-t/R_T C} \qquad (17.2)$$

The voltage across any one of the circuit elements can now be computed because the current is known. The results are tabulated in convenient form in figure (17.13). A similar evaluation can be made for the series RL circuit.

When simple series RC circuits are connected to switch tubes, the transients are made to repeat at intervals governed by the timing of the grid signal. Typical results for an RC circuit are shown in figure (17.14). It is apparent that the responses of switch tube RC circuits offer a number of waveform possibilities. Thus the approximately square grid signal can be converted into three different outputs:

(1) A peaked output: short time constant circuit with the output taken across R_g.

(2) Sawtooth output: long time constant circuit with the output taken across C.

(3) Trapezoidal output: long time constant circuit with the output taken across R and C in series, the voltage e_b in figure (17.14).

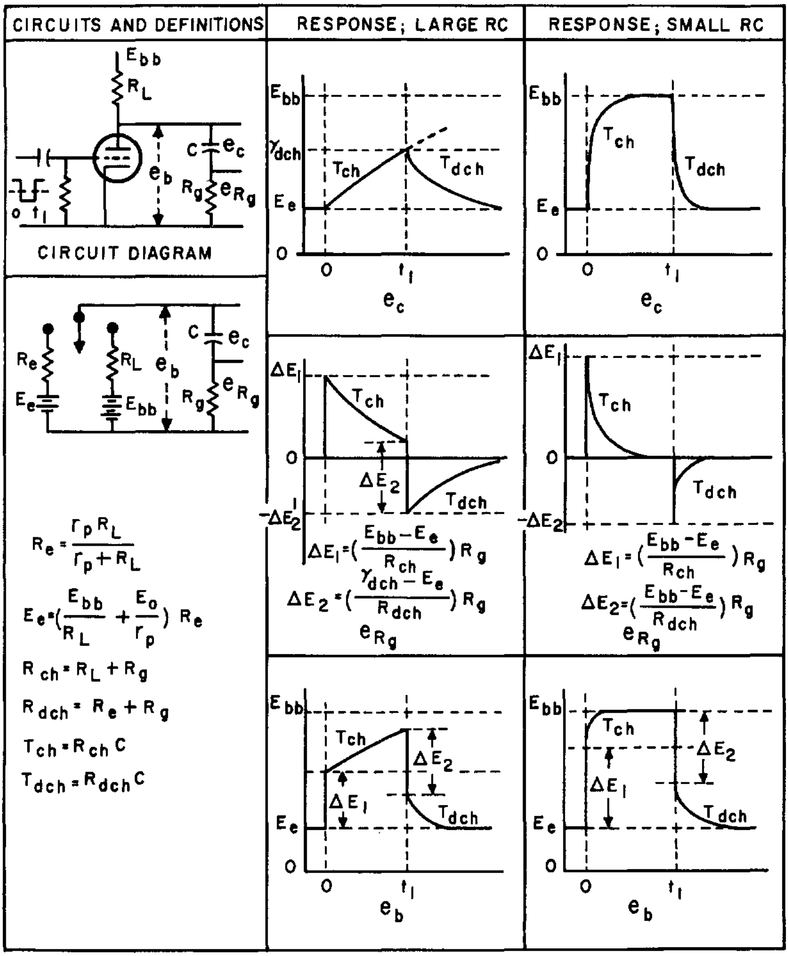


Fig. 17.14. Response of a series *RC* circuit when connected to a switch tube.

17.7. Differentiators and Integrators

From figure (17.14) which shows the response characteristics of a switch tube *RC* circuit, you can see that a square wave grid signal can be converted to a *peaked* output when the circuit time constant is small compared with the half period of the grid signal. Thus the circuit can be used as a *peaker*.

The peaking action of a short time constant *RC* circuit is closely related to mathematical differentiation, as shown in figure (17.15). If the input signal is a square wave, the true mathematical derivative consists of positive and negative pulses of zero width and infinite amplitude. In general, the shorter the circuit time constant, the more nearly the output voltage from the RC peaker circuit approaches the form of the true derivative.

The mathematical similarity between differentiation and the operation of a short time constant RC circuit can be shown very quickly in trans-

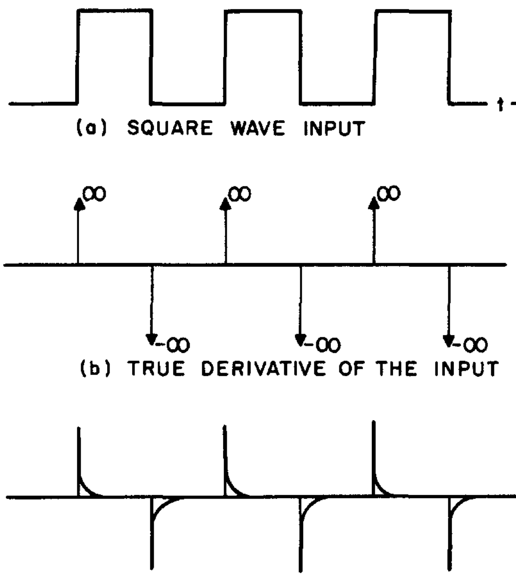


Fig. 17.15. Peaker waveform relationships.

form notation. Assume that some arbitrary signal $e_i(t)$ is applied to the RC circuit shown in figure (17.16). In terms of complex frequency, the circuit transfer function is

$$\frac{E_o(s)}{E_i(s)} = \frac{\mathcal{L}[e_o(t)]}{\mathcal{L}[e_i(t)]} = RCs\frac{1}{RCs+1} \tag{17.3}$$

Now, if $e_o(t)$ is actually the derivative of the input signal, $e_o(t) = Kde_i(t)/dt$ or, in terms of complex frequency, $E_o(s) = KsE_i(s)$. Thus, in the ideal differentiator, the circuit transfer function is

$$\frac{E_o(s)}{E_i(s)} = Ks \tag{17.4}$$

Equations (17.3) and (17.4) show that the difference between true differentiation and the properties of an RC peaker is given by the factor $1/(RCs + 1)$. Thus the RC circuit can approach the function of a true

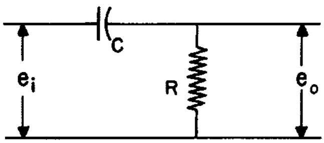


Fig. 17.16. RC circuit.

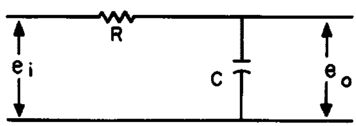


Fig. 17.17. RC circuit.

differentiator if the time constant RC is made so small that the product RCs is much less than 1 for all significant values of s. If this condition is assumed, the transform output voltage, obtained from equation

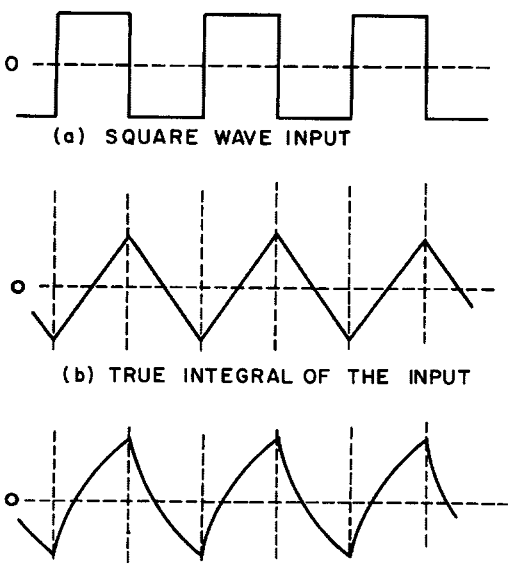


Fig. 17.18. Waveform relationships in an integrator.

(17.3), is approximately $E_o(s) \doteq RCsE_i(s)$. Thus, as a function of time,

$$e_o(t) \doteq RC\frac{de_i}{dt} \tag{17.5}$$

For short time constant peaker circuits, equation (17.5) is extremely useful in estimating the output for a given input signal. The amplitude

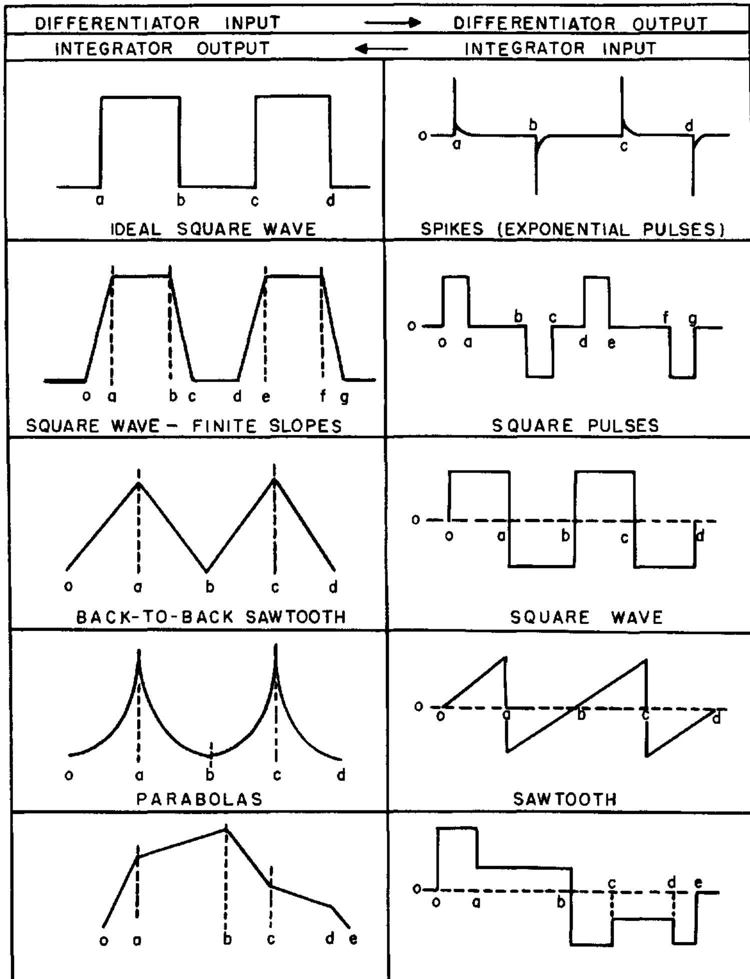


Fig. 17.19. Waveform relationships in integrators and differentiators.

of the output signal is reduced as the time constant is shortened, and this is a major disadvantage. However, even with this drawback, the differentiating characteristics of short time constant *RC* circuits are useful in generating unusual waveforms. Some typical cases are shown in figure (17.19). Of course, when input signals other than square waves are to be differentiated, class A amplifiers must be used in place of switch tubes.

When a square wave is applied to a *long time constant RC* circuit and the output is taken across the capacitor, the output voltage has the appearance of a sawtooth. The similarity between this action and mathematical integration is shown in figure (17.18). The circuit used is shown in figure (17.17).

The voltage transfer function of the *RC* circuit is

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{RC} \cdot \frac{1}{s + 1/RC} \tag{17.6}$$

When the circuit time constant is large, so that $1/RC$ is very much less than s for all significant values of s, the transfer function is approximately

$$\frac{E_o(s)}{E_i(s)} \doteq \frac{1}{RC} \frac{1}{s} \tag{17.7}$$

In an ideal integrator,

$$e_o(t) = K \int e_i(t)\, dt \tag{17.8}$$

or

$$E_o(s) = K \frac{E_i(s)}{s} \tag{17.9}$$

Thus the ideal transfer function is

$$\frac{E_o(s)}{E_i(s)} = K \frac{1}{s} \tag{17.10}$$

It is clear that an *RC* circuit can be used to approximate mathematical integration. A few possible waveform relationships are shown in figure (17.19). Integrating circuits using amplifiers require class A operation of the amplifier for all input waveforms other than square waves.

17.8. Integration and Differentiation with Feedback Amplifiers

In chapter 7, covering feedback circuits, as well as in the discussion of the Miller effect in chapter 3, it was shown that the introduction of a feedback impedance between the plate and grid of an amplifier tube caused marked changes in the input impedance of the stage. Thus if the feedback impedance Z_{fb} is connected as shown in figure (17.20a), the input impedance of the amplifier becomes $Z_{in} = Z_{fb}/(1 - A)$, where $A =$ voltage gain of the stage $= -g_m Z_m$. If the feedback impedance is either resistive or capacitive,

$$R_{in} = \frac{R_{fb}}{1 - A} \tag{17.11}$$

$$C_{in} = C_{fb}(1 - A) \tag{17.12}$$

Normally, the amplifier is designed to have a bandwidth of sufficient width so that the reference gain can be used in place of the complex voltage gain. Hence

$$R_{in} = \frac{R_{fb}}{1 + A_r} \tag{17.13}$$

$$C_{in} = C_{fb}(1 + A_r) \tag{17.14}$$

The circuit connections and equivalent representations are shown in figure (17.20).

This effect can be used to improve the integrating and differentiating action of RC circuits. Consider the two circuits shown in figure (17.21).

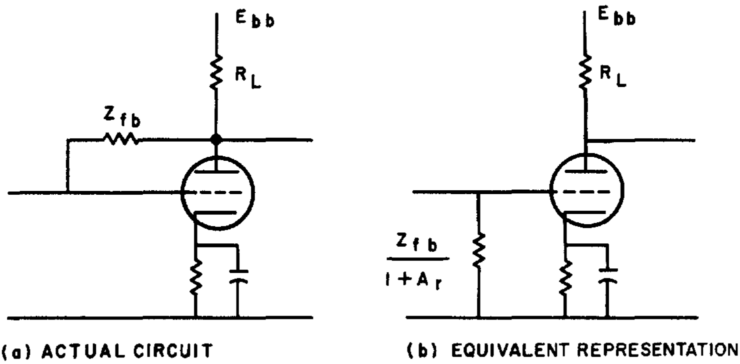


Fig. 17.20. Effect of grid-to-plate feedback on the input impedance of class A amplifiers.

With a feedback capacitor C_{fb} as in figure (17.21a), the input capacitance is increased to $C_{fb}(1 + A_r)$, causing a corresponding increase in the time constant to $RC_{fb}(1 + A_r)$. Thus large time constants can be obtained using high gain amplifiers. Also, because the output voltage is $e_o = -e_g A_r$ and because

$$e_g \doteq \frac{1}{RC_{fb}(1 + A_r)} \int e_i(t)\, dt$$

the output voltage can be expressed as

$$e_o \doteq \frac{1}{RC_{fb}} \cdot \frac{A_r}{1 + A_r} \int e_i(t)\, dt \tag{17.15}$$

It is clear that the integrated output from the amplifier will be relatively independent of increases in the time constant brought about by gain increases.

A similar analysis of the feedback differentiator of figure (17.21b) shows that the output voltage is approximately

$$e_o \doteq R_{fb}C \frac{A_r}{1 + A_r} \cdot \frac{de_i}{dt} \tag{17.16}$$

Again short time constants and excellent differentiating can be obtained without loss of output voltage simply by changing the reference gain of the amplifier.

Essentially the same effect can be achieved by simply applying the output from a conventional integrator or differentiator to an amplifier

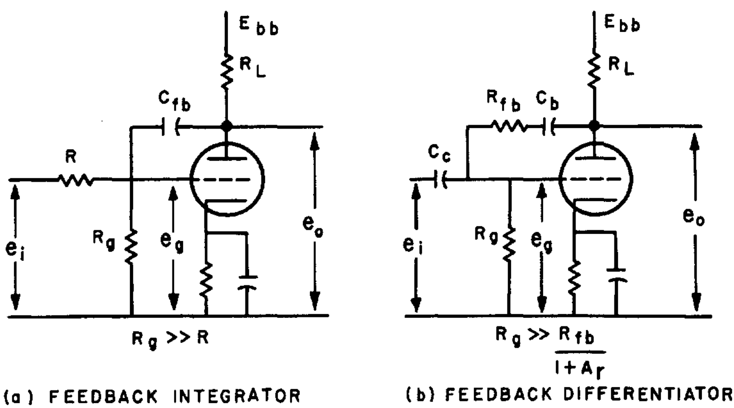


Fig. 17.21. Feedback integrator and differentiator.

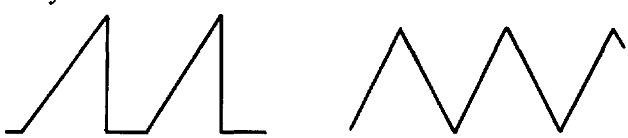
of reference gain A_r. The amplitude of the output from the system will be the same as the output from the feedback circuits if (1) the differentiator time constant is decreased by A_r; (2) the integrator time constant is increased by A_r, and if A_r is much larger than 1. Thus the end result is essentially the same as that for the feedback amplifier, and the purported advantage of feedback is open to question.

The advantage of the feedback circuit lies in the fact that the time constant change brought about by feedback is effective only as long as the amplifier tube conducts. When nonconducting, the gain is zero and the time constant of the input circuit is the unmodified value. Thus, by gating the amplifier tube, the time constant can be made to vary with time and it is in this sense that feedback circuits have a clear-cut advantage.

17.9. Sawtooth Voltage Generators

Voltage waveforms having sawtooth appearances, such as those shown in figure (17.22), find extensive application in electronic systems. Waveform (a) is widely used in laboratory oscilloscopes to provide a horizontal deflection of the cathode ray tube beam that is a linear function of time. This simple sawtooth is also used in time modulation systems and in the generation of parabolic and trapezoidal voltages.

The simplest sawtooth generator consists of a long time constant RC circuit connected to the output of a switch tube. The output voltage is taken across the capacitor. This possibility was noted in an earlier section and is clear from figure (17.14); this figure also shows the circuit diagram. When the circuit is used for this purpose, R_g is usually replaced by a short circuit.



(a) SIMPLE SAWTOOTH (b) BACK-TO-BACK SAWTOOTH

Fig. 17.22. Sawtooth voltage waveforms.

If a back-to-back sawtooth voltage output from the circuit is desired, the charging and discharging time constants of the capacitor should be equal. Also, the total voltage change ($E_A - \gamma$) should be the same in both cases. Because neither of these conditions can ever be met exactly with switch tubes, a true symmetrical back-to-back sawtooth is never developed by such a circuit. However, symmetry can be obtained by using a square wave input to a class A amplifier.

To obtain a simple sawtooth, the charging time constant, T_{ch}, must be much larger than the discharging time constant, T_{dch}. With R_g short circuited as previously noted, $T_{ch} = R_L C$, $T_{dch} = R_e C$, where $R_e = r_p R_L/(r_p + R_L)$. The specified inequality of time constants results if $R_L \gg R_e$, which simply requires that R_L be much greater than r_p. This can be obtained by using low plate resistance triodes or triode gas tubes.

A negative sawtooth output can also be obtained, but the reverse of the preceding inequality is required.

It should be recognized that this simple circuit will never provide a true sawtooth waveform because some exponential curvature in the charging curve of the capacitor voltage always exists. The amount of exponential curvature can be reduced by either of two simple methods:

(1) The charging time constant can be increased, mainly by increasing R_L.

(2) The power supply voltage E_{bb} can be raised.

The effects of these changes are shown in figure (17.23). For the same change in output voltage, the linearity improvement is apparent.

In each of the two preceding cases there are practical maximum limits that cannot be exceeded conveniently. An excessively large value of R_L reduces the output voltage magnitude to unreasonably small values, while the usual limitations exist with regard to the maximum practical power supply voltage.

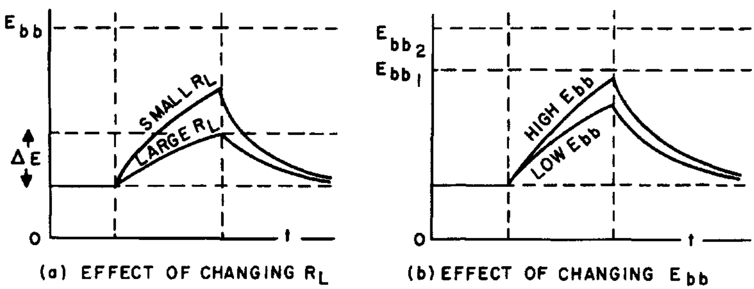


Fig. 17.23. Brute force attempts to linearize an exponential sawtooth.

An alternative method of sawtooth linearization uses constant current charging of the capacitor. If a constant current I is supplied to a capacitor, the voltage across the capacitor is

$$e_c = \frac{1}{C}\int I\,dt = \frac{It}{C}$$

and this is a linear function of time.

Constant current charging is possible with various circuits. One method is shown in figure (17.24). Here the load resistor of the switch tube is replaced by a suitably biased pentode. In the normal region of operation the pentode is essentially a constant current device. The effect obtained is equivalent to that resulting from charging C through a load resistance r_p. Thus some exponential curvature remains in the sawtooth.

In some cases it might be desirable to interchange the positions of the switch tube and the pentode to simplify the establishment of the proper polarizing potentials on the pentode. If this is done, the pentode must be gated on and off in synchronism with the switch tube.

Constant current charging can also be achieved by using positive feedback in a way that will hold the voltage across the charging resistor constant. This is the essential principle used in the *bootstrap cathode follower*, which is shown in figure (17.25). The cathode follower is designed to have a gain approaching unity. As a result, the signal

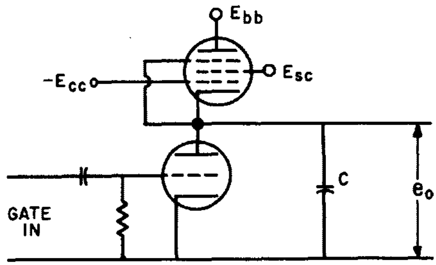


Fig. 17.24. Constant current charging of C through a pentode.

output from the cathode follower will nearly equal the voltage variation across the capacitor C.

In the quiescent condition, the switch tube is highly conducting, and the voltage at point A is nearly equal to E_{bb}, differing only by the drop

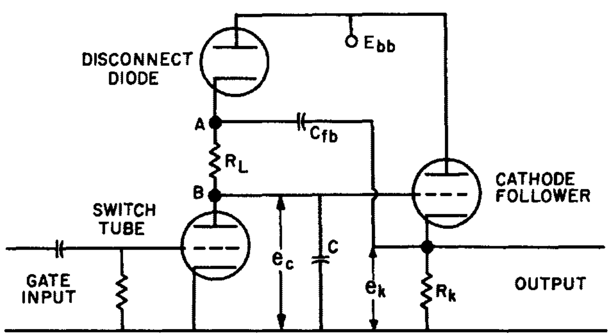
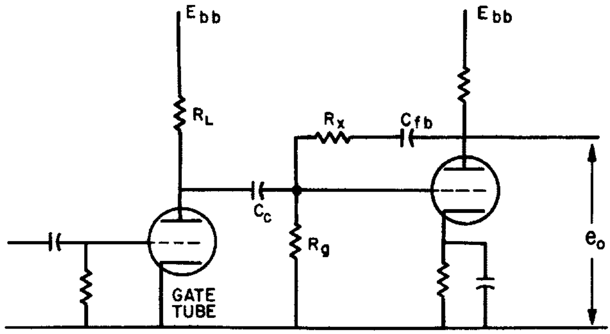


Fig. 17.25. Bootstrap cathode follower.

across the diode. The voltage at point B is equal to the drop across the tube and this is usually small. The feedback capacitor C_{fb} is charged to a voltage approximately equal to $(E_{bb} - E_{kk})$, where E_{kk} is the quiescent cathode voltage.

When the switch tube is cut off, the capacitor C commences charging through R_L, and this causes the capacitor voltage e_c to increase. The cathode follower output increases by an almost equal amount. If the feedback capacitor is so large that the voltage across it does not change noticeably during the charging period of C, the increase in cathode follower output will cause *both* sides of the load resistor to increase in voltage by nearly the same amount. Thus the potential across the resistor remains constant, and a constant current flows to charge the capacitor C. The potential at A rises above E_{bb}, and the disconnect diode becomes nonconducting. This disconnects the power supply. If the diode were omitted, the potential at point A could not rise above E_{bb}, and the voltage across R_L would not be constant.

Because the gain of the cathode follower is never quite unity and



$C_c - R_g$ PROVIDE IDEAL COUPLING

Fig. 17.26. Miller integrator for sawtooth linearization.

because the voltage across C_{fb} is not exactly constant, some curvature in the sawtooth results.

Another feedback system for sawtooth linearization uses a conventional switch tube coupled to a feedback integrator of the type shown in figure (17.20a). The feedback improves the integrating properties of the circuit. Practical circuits of this type are called *Miller integrators* and an example is shown in figure (17.26).

The resistor R_x shown in the feedback path of figure (17.26) is a practical refinement required to avoid certain transient effects. Its magnitude is usually[2] about $1/g_m$.

It is advisable in practical cases to gate the amplifier tube V_2 so that it conducts only during the charging of the capacitor. This prevents the capacitor discharge time constant from being multiplied by the amplifier gain.

[2] Chance, *et al.*, *Waveforms*, vol. 19, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, p. 280.

17.10. Flyback Time Reduction

In electrostatically deflected cathode ray tubes the position of the light spot on the face of the screen is a linear function of the voltage between the deflecting plates. In many applications the spot should move across the screen with a uniform speed so that there is a linear relation between spot position and time. After moving a specified distance across the screen, the spot should return to its point of origin. The period of uniform spot velocity is called the *sweep time.* The period required for the spot to return to its point of origin from the end of the sweep is called the *flyback time.* Ideally, the flyback time is negligible relative to the sweep time. If it is not, the effect shown in figure (17.27)

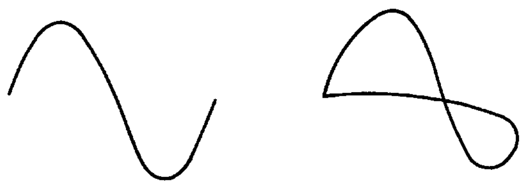


Fig. 17.27. Waveform distortion caused by too much flyback time. (a) Waveform obtained with negligible flyback. (b) Same signal when flyback time is excessive.

occurs. Evidently a sawtooth voltage is required on the deflecting plates of cathode ray tubes when used as described.

The simplest method of making modest reductions in flyback time was mentioned in section (17.9). If additional circuit complexity can be tolerated, marked reductions in flyback time can be achieved through the use of a gated Miller integrator.

A typical Miller integrator using a cathode follower and connected to an *RC* switch tube circuit is shown in figure (17.28). In the usual Miller integrator, feedback is applied through a series *RC* circuit directly from the plate to the grid of the amplifier tube. In the circuit of figure (17.28), the plate of the amplifier is connected to the cathode follower input and the output of the cathode follower is coupled back to the amplifier grid through the feedback capacitor C_{fb}. Thus the cathode follower replaces the compensating resistor R_x shown in figure (17.26).

The gain of the pentode used in the Miller integrator and the value of C_{fb} are adjusted to give the required value for the sweep capacitance.

The pentode amplifier is turned off by a gating signal on the suppressor grid during the discharge period so that the time constant is enormously reduced. Flyback time reduction is thereby achieved.

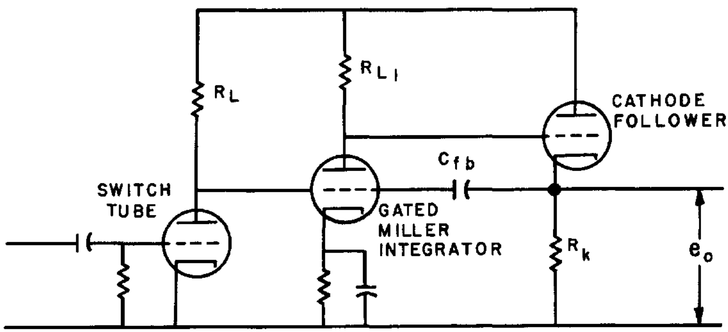


Fig. 17.28. Use of a cathode follower to reduce flyback time in a Miller integrator sweep circuit.

17.11. Trapezoidal Voltages (Sawtooth Currents)

The response characteristics of a series *RC* circuit connected to a switch tube were shown in figure (17.14). It was shown that the voltage e_b, indicated in figure (17.29) or (17.14), had a trapezoidal form. The

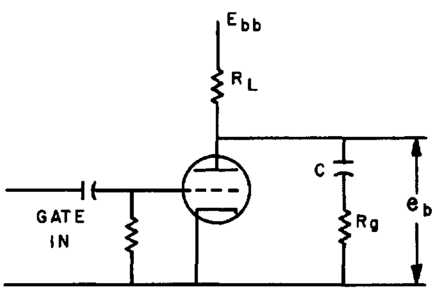


Fig. 17.29. *RC* switch tube trapezoidal voltage generator.

linearity of the increasing part of the trapezoid can be improved by the same general methods as those outlined for sawtooth improvement.

The trapezoidal voltage waveform has two principal fields of application in sawtooth current generation and in self-gated sawtooth voltage sources.

Magnetically deflected cathode ray tubes are widely used in radar and television systems. In such cases, position of the light spot on the screen is nearly a linear function of the current through the deflecting coils. Hence if a linear sweep of the electron beam across the face of the screen is desired, the current through the deflecting, or sweep, coil should be a sawtooth function of time. This is shown in figure (17.30). If the slope of the sawtooth current is K, the equation for the current during the interval T is

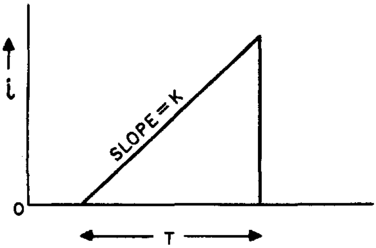


Fig. 17.30. Sawtooth current required for a linear sweep in a magnetically deflected cathode ray tube.

$$i = \pm Kt; \qquad 0 \leq t \leq T$$

The sweep coil is usually connected as the load element in a class A power amplifier as shown in figure (17.31a). The corresponding equivalent plate circuit, including the coil resistance R_c is shown in figure (17.31b). It is necessary to determine the grid voltage e_g required to make the sweep coil current i a linear function of time.

From the equivalent plate circuit of the sweep amplifier shown in figure (17.31b), the voltage loop equation is

$$\mu e_g = Ri_p + L\frac{di_p}{dt}$$

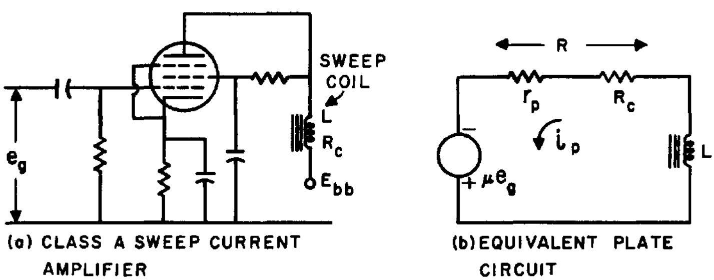


Fig. 17.31. Sweep amplifier.

However, over the interval from $t = 0$ to $t = T$, the coil current is assumed to be

$$i_p = Kt$$

Hence the voltage loop equation reduces to

$$\mu e_g = (RK)t + (LK)$$

and the voltage required on the grid of the sweep amplifier is

$$e_g = \frac{LK}{\mu} + \frac{RK}{\mu} t \tag{17.17}$$

where R, L, K, and μ are all constants. Thus the grid voltage has a d-c component LK/μ, called the *initial jump*, and a linearly increasing term of slope $= RK/\mu$. Hence a trapezoidal grid voltage of the form shown in figure (17.32) is required to produce a sawtooth current through the sweep coil. The ratio of the jump to the slope is L/R, which is the time constant of the coil circuit.

It was noted at the beginning of this section that the switch tube RC circuit generates a trapezoidal voltage waveform. Consequently, such

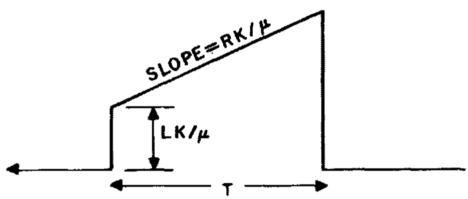


Fig. 17.32. Grid voltage required to produce a sawtooth sweep current.

a circuit can be used to supply the grid excitation for the sweep amplifier tube if the constants of the RC circuit are adjusted to provide the proper jump and slope. Reference should be made to the waveforms and formulas given in figure (17.13) and (17.14). Thus, from the waveform for e_b in figure (17.14), it is clear that the initial jump is ΔE_1. Hence, equating this jump to that determined in equation (17.17), we get

$$\text{jump} = \frac{E_{bb} - E_e}{R_{ch}} R_g = \frac{LK}{\mu_1} \tag{17.18}$$

where $$R_{ch} = R_L + R_g$$

and $$E_e = \frac{r_p R_L}{r_p + R_L} \left(\frac{E_{bb}}{R_L} + \frac{E_0}{r_p} \right)$$

It is generally valid to assume that E_{bb}/R_L is much larger than E_0/r_p, so that

$$E_e \doteq E_{bb} \frac{r_p}{r_p + R_L}$$

Hence
$$E_{bb} - E_e \doteq E_{bb}\left(1 - \frac{r_p}{r_p + R_L}\right)$$

$$\doteq E_{bb}\frac{R_L}{r_p + R_L} \tag{17.19}$$

Therefore the jump voltage is

$$\text{jump} = \frac{LK}{\mu_1} = E_{bb}\frac{R_L}{r_p + R_L}\cdot\frac{R_g}{R_L + R_g} \tag{17.20}$$

where $\mu_1 = \mu$ of the sweep amplifier tube; $r_p = r_p$ of the switch tube. This is a basic design equation.

The equation for the slope can be obtained in a similar manner by consolidating the data in figures (17.13) and (17.14). Thus, the trapezoidal voltage is clearly e_b, where

$$e_b = E_{bb} - iR_L$$

$$i = \frac{E_{bb} - E_e}{R_L + R_g}\varepsilon^{-t/T_{ch}}$$

Thus, using the approximate formula given in (17.19) for $E_{bb} - E_e$, the trapezoidal voltage can be expressed as

$$e_b = E_{bb}\left[1 - \frac{R_L}{r_p + R_L}\cdot\frac{R_L}{R_L + R_g}\varepsilon^{-t/(R_L+R_g)C}\right] \tag{17.21}$$

The initial slope is simply the derivative of e_b evaluated at $t = 0$, or

$$\text{slope} = \frac{RK}{\mu_1} = \left(\frac{R_L}{R_L + R_g}\right)^2\frac{E_{bb}}{r_p + R_L}\cdot\frac{1}{C} \tag{17.22}$$

This is a second useful equation.

The jump-to-slope ratio has the form

$$\frac{\text{jump}}{\text{slope}} = \frac{L}{R} = R_gC\left(1 + \frac{R_g}{R_L}\right) \tag{17.23}$$

The three equations for the jump, slope, and jump-to-slope ratio provide the necessary design relationships. Only two of the equations are independent. The design is not necessarily straightforward, but permits a wide latitude of designer's choice, particularly with regard to the selection of the sweep amplifier and switch tubes.

17.12. *RLC* Circuit Response

The transient response of a parallel *RLC* circuit, such as that shown in figure (17.33), may exhibit any one of three characteristics, depending upon the value of the circuit Q. All three responses have applications in practical systems. The basic circuit equation governing response will be developed in this section.

The terminology in figure (17.33) is specified as follows: I_A = applied current = constant; ρ = initial current through L at $t = 0^-$ and 0^+; $\Delta I = I_A - \rho$ = total possible change in current through the inductance. The initial voltage across the capacitor is assumed to be zero.

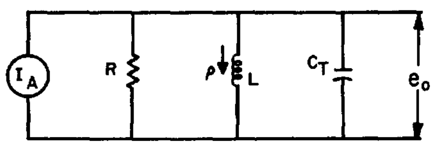


Fig. 17.33. Parallel tuned circuit.

The nodal equation for the network is

$$I_A = i_R + i_C + i_L = \frac{e_o}{R} + C_T \frac{de_o}{dt} + \frac{1}{L}\int e_o \, dt$$

The corresponding Laplace transform is

$$\frac{I_A}{s} = E_o(s)\left(\frac{1}{R} + sC_T + \frac{1}{sL}\right) + \frac{\rho}{s}$$

where $E_0(s) = \mathcal{L}[e_0(t)]$. If this equation is solved for the transform output voltage the result is

$$E_o(s) = \frac{\Delta I}{C_T} \cdot \frac{1}{s^2 + (1/RC_T)s + 1/LC_T}$$

In chapter 4 it was shown that $B = 1/RC_T$ = bandwidth in rps; $\omega_0^2 = 1/LC_T$ = (resonant frequency)^2. Thus the transform output voltage can be written

$$E_o(s) = \frac{\Delta I}{C_T} \cdot \frac{1}{s^2 + Bs + \omega_0^2} \qquad (17.24)$$

The transient response of the circuit is controlled by the poles of the response transform, and these are the roots of the characteristic equation $s^2 + Bs + \omega_0^2 = 0$. Thus the poles are

$$s_{1,2} = -\frac{B}{2} \pm \sqrt{\left(\frac{B}{2}\right)^2 - \omega_0^2}$$

A more convenient form is obtained by factoring the $B/2$ factor as follows:

$$s_{1,2} = \frac{B}{2}\left(-1 \pm \sqrt{1 - 4Q^2}\right) \tag{17.25}$$

where

$$Q = \omega_0/B. \tag{17.26}$$

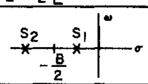
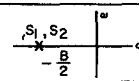
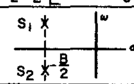
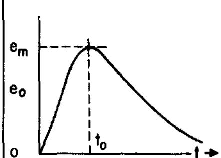
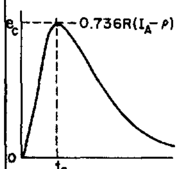
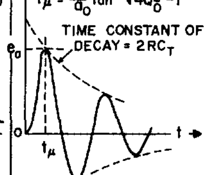
ITEM	OVERDAMPED $Q_0<0.5$	CRITICAL DAMPING $Q_0=0.5$	UNDER-DAMPED $Q_0>0.5$
POLES	$S_1 = \frac{B}{2}\left[-1+\sqrt{1-4Q_0^2}\right]$ $S_2 = \frac{B}{2}\left[-1-\sqrt{1-4Q_0^2}\right]$	$S_1 = -\frac{B}{2}$ $S_2 = -\frac{B}{2}$	$S_1 = \frac{B}{2}\left[-1+j\sqrt{4Q_0^2-1}\right]$ $S_2 = \frac{B}{2}\left[-1-j\sqrt{4Q_0^2-1}\right]$
POLE LOCATIONS			
EQUATION FOR THE OUTPUT VOLTAGE; UNIT STEP INPUT	$e_0 = R(I_A-\rho)\frac{\epsilon^{-Bt/2}}{\sqrt{1-4Q_0^2}}(\epsilon^{at}-\epsilon^{-at})$ $a = \frac{B}{2}\sqrt{1-4Q_0^2}$	$e_0 = R(I_A-\rho)Bt\epsilon^{-Bt/2}$	$e_0 = R(I_A-\rho)\frac{2\epsilon^{-Bt/2}}{\sqrt{4Q_0^2-1}}\sin a_0 t$ $a_0 = \frac{B}{2}\sqrt{4Q_0^2-1}$
PLOT OF PRECEDING EQUATIONS $Q_0 = R\sqrt{\frac{C_T}{L}}$ $B = \frac{1}{RC_T}$	 $t_0 = \frac{1}{a}\tanh^{-1}\sqrt{1-4Q_0^2}$ $e_m = \frac{R(I_A-\rho)}{\sqrt{1-4Q_0^2}}\epsilon^{-Bt/2}(\epsilon^{at_0}-\epsilon^{-at_0})$	 $t_c = 2RC_T$	$t_\mu = \frac{1}{a_0}\tan^{-1}\sqrt{4Q_0^2-1}$  $e_\mu = \frac{2R(I_A-\rho)}{\sqrt{4Q_0^2-1}}\epsilon^{-Bt/2}\sin a_0 t_\mu$

Fig. 17.34. Response characteristics of parallel tuned circuits.

The nature of the transient response is controlled by the relationship between 1 and $4Q^2$. Thus the following three possibilities exist:

(1) The value of $4Q^2$ may be less than 1. The poles are real and different and the response is *overdamped.*

(2) The value of $4Q^2$ may be equal to 1. The poles are real and equal and the response is *critically damped.*

(3) The value of $4Q^2$ may be greater than 1. The poles are complex conjugates and the response is *oscillatory.*

These three operating regimes are summarized in figure (17.34), together with the equations for the time responses.

It is clear from this figure that a number of waveform possibilities are available through the use of a parallel tuned circuit in connection with a class A amplifier with a square wave input or with a switch tube. Both the critically damped and overdamped response characteristics are similar to the output from an *RC* peaker when the effects of interelectrode and distributed capacitances are included. As a result, the *RLC* circuit is often used as a peaker, and the damping is usually adjusted to the critical point because this gives a sharper pulse.

The oscillatory response of an *RLC* circuit is often used in *ringing* circuits. The circuit oscillation constitutes the ringing.

17.13. *RLC* Peakers and Ringing Circuits

A class A *RLC* peaker with a square wave input is worked out in figure (17.35). Critical damping is assumed and the total damping resistor R includes both R_L and r_p in parallel. All the necessary information is given on this figure, and the required equations were developed in the preceding section. Note that the peaked output pulses have equal magnitudes in figure (17.35).

It is *not* generally desirable to place the tuned circuit of an *RLC* peaker in series with the cathode lead of the tube, because the cathode follower action shunts the circuit with a low resistance. This may require the addition of shunt capacitance, and the peaked output is broadened.

It is possible to use a switch tube in place of a class A amplifier and thereby obtain peaking of the input gating signal on the grid of the switch tube. This does not always provide satisfactory operation, because the degree of damping varies as the tube alternately conducts and cuts off. Thus, if operation is adjusted to provide critical damping when the tube is cut off, the response will obviously be overdamped when the tube conducts. Conversely, the establishment of critical damping when the tube is conducting will lead to an oscillatory response when the tube cuts off.

A typical switch tube *RLC* peaker circuit is shown and analyzed in figure (17.36). This is *not* a general solution because it is assumed that the switch tube is normally off and is turned on only for short intervals by a positive gate. The reverse condition might also exist. Thus, in the assumed case, the coupling capacitor is charged to E_{bb} most of the time, because it is large enough to hold its charge during the period of tube conduction. Somewhat different results are obtained if the tube is held normally conducting and is cut off for short periods by a negative gate.

A *ringing circuit*, which consists of a switch tube and tuned circuit, is shown in figure (17.37). A ringing circuit is essentially a *gated oscillator*

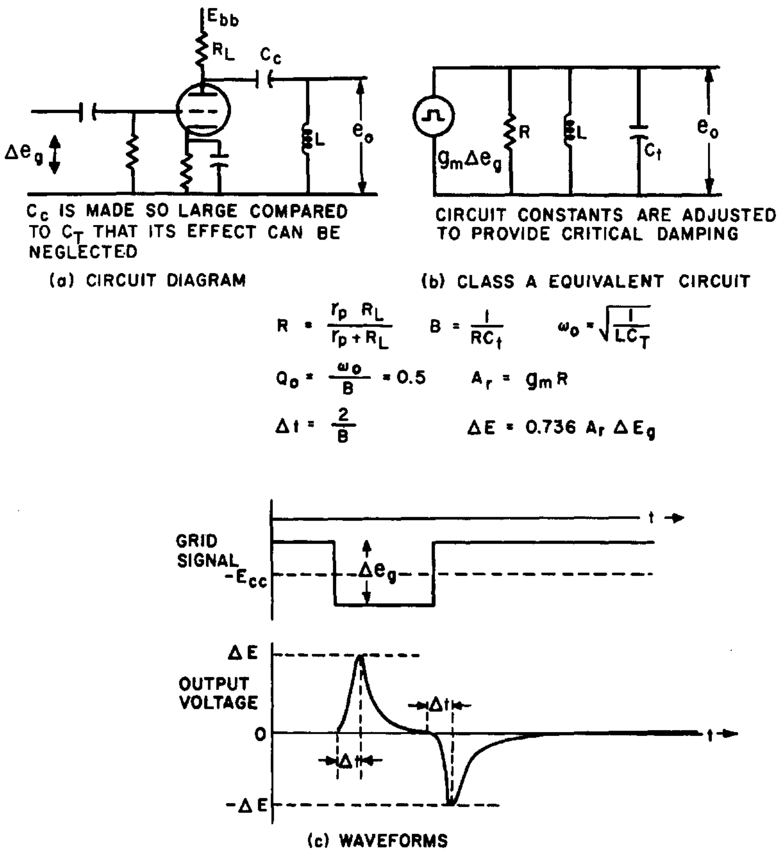


Fig. 17.35. Class A *RLC* peaker circuit.

in the sense that an oscillatory output is desired only during specific time intervals. There are many such circuits,[3] of which the ringing circuit is the simplest.

[3] Chance, *et al.*, *Waveforms*, vol. 19, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1948, pp. 140–145.

Because an oscillating output is desired only during a particular time interval, it is not practical to connect the *RLC* circuit in series with the plate of the tube. This would lead to oscillations following *both* edges of the input gate, and this is not usually desirable. Class A ringing circuits are not common for the same reason. The cathode connection

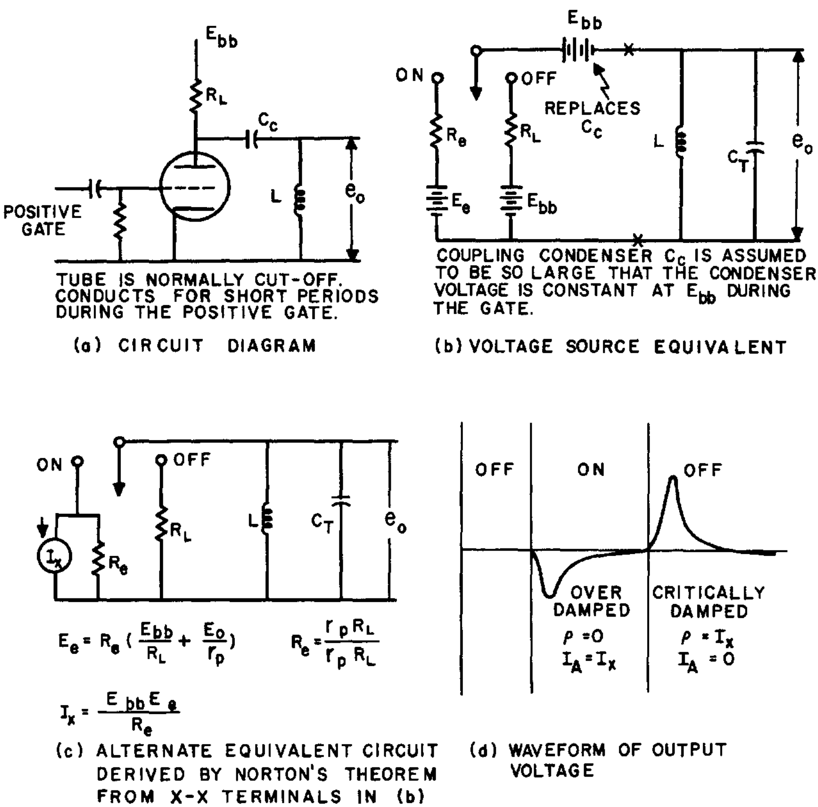


Fig. 17.36. Switch tube *RLC* peaker.

shown in figure (17.37) is desirable because a high Q circuit is obtained when the tube is cut off and the low output impedance developed by cathode follower action during tube conduction overdamps the circuit and prevents oscillations after the trailing edge of the gate signal.

It is obvious that the magnitude of the gating signal on the grid of the switch tube must be large enough to keep the tube cut off during the large negative cathode voltage swings.

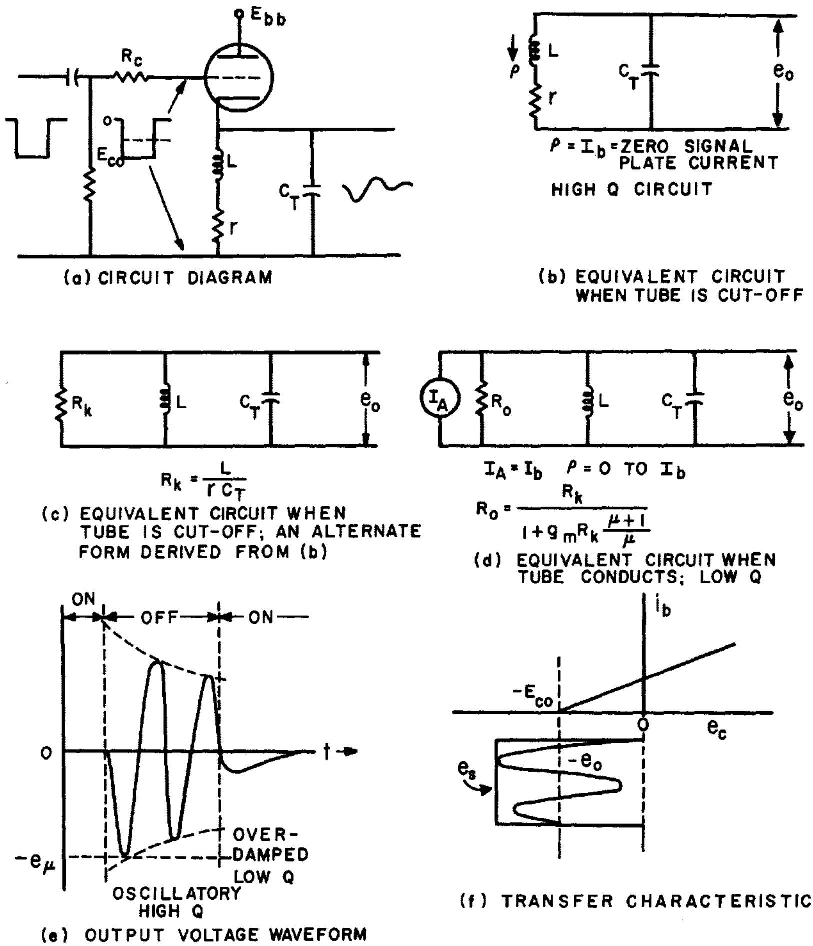


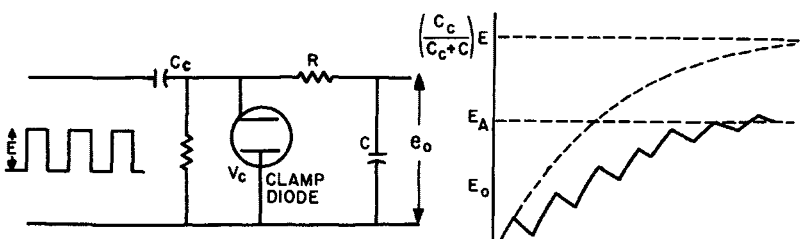
Fig. 17.37. Ringing circuit.

17.14. Nonregenerative (Storage Type) Counters

In nuclear instrumentation, in certain types of radar, as well as in special purpose measuring equipment, it is often desirable to count a succession of more or less uniform pulses up to some predetermined point. This process is related to frequency division, but in most *counting circuits* the spacing between successive pulses need not be uniform; however, there is usually a definite maximum allowable spacing. Counting circuits are of two general types: (1) regenerative, discussed in chapter 18; (2) nonregenerative or storage, discussed here.

The counting problem usually involves voltage pulses. Thus most storage type counters use a storage capacitor to accumulate charge during each pulse without losing it between pulses. This causes the voltage across the capacitor to build up in steps along an exponential curve. The capacitor voltage reaches some predetermined value at the end of n pulses, and this causes some other device to operate. This device discharges the capacitor and allows the counting process to repeat. Such a unit is called the *recycling device.*

The basic principle of the storage type counter can be developed from the simple integrator shown in figure (17.38a). The input pulses



(a) RC INTEGRATOR; INPUT CLAMPED TO 0 BY V_c

(b) OUTPUT VOLTAGE WAVEFORM

Fig. 17.38. Prototype of storage type counter.

are applied through the coupling capacitor C_c and are clamped to a reference baseline by the diode V_c. The capacitance of the storage capacitor C is large compared with that of C_c, and R is a large resistance. Thus the output voltage appears as shown in figure (17.38b).

When a pulse is applied to the counter the storage capacitor charges through R and C_c. When the pulse is removed, C discharges through the same path. Fortunately, C does not lose charge so rapidly as it accumulated because of the difference between the applied and initial capacitor voltages during charge and discharge. Thus the capacitor voltage will reach some specified voltage, say E_A, at the end of some specified number of pulses determined by the circuit time constant and pulse amplitude. The dotted curve in figure (17.38b) shows the normal charging curve that would result if a constant voltage E were applied to the circuit.

The preceding circuit suffers from two obvious and serious disadvantages:

(1) The storage element loses charge between pulses. This is disastrous when the intervals between successive pulses are very long or irregular.

(2) The buildup of voltage across C is slow because of the long time constant.

Both of these disadvantages can be largely offset by replacing the resistor R with a diode, called the *disconnect diode* V_d. This is shown in figure (17.39a). The diode provides a low resistance charging path, but prevents capacitor discharge between pulses because it is then nonconducting.

Before any pulses are applied to the circuit, both capacitors are uncharged. The pulse amplitude is E. When the first positive pulse

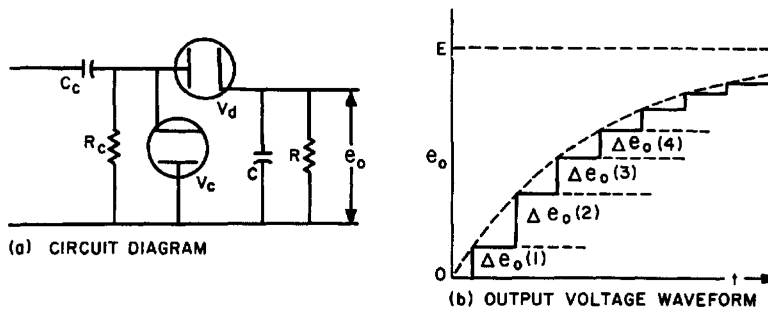


Fig. 17.39. A simple storage type counter.

occurs, the disconnect diode V_d conducts, and both capacitors rapidly charge through the diode resistance r_p. The charging time constant is

$$T_{ch} = r_p \frac{CC_c}{C + C_c}$$

where r_p includes the signal source impedance, if any. This time constant should be small compared with the pulse duration. If this is true, the voltages across the two capacitors at the end of the pulse are[4]

$$e_o(1) = E\frac{C_c}{C + C_c}; \qquad e_{C_c}(1) = E\frac{C}{C + C_c} = aE$$

When the first pulse ends, the signal input drops to zero. Hence the voltage at point A drops by E volts and diode V_d ceases conduction. As a result, the storage capacitor has no discharging path, other than leakage, and its voltage remains practically constant. On the other

[4] This development follows that by S. Moskowitz and J. Racker, *Pulse Techniques*, Prentice-Hall Inc., New York, 1951, pp. 213–216.

hand, because the cathode of the clamping diode drops by E volts, while C_c was charged to only $(EC)/(C + C_c)$ volts, the clamping diode conducts and C_c discharges rapidly through the r_p of this tube. The second pulse should not be applied until this discharging process is virtually complete. Thus this time constant sets a lower limit on the spacing between successive pulses.

When the second pulse is applied, the disconnect diode cannot conduct until the pulse voltage exceeds the voltage aE across the storage capacitor. Hence the voltage change during the second pulse is

$$E_A(2) = E - aE = E(1 - a)$$

This voltage divides between the two capacitors as before, so that the final output voltage at the end of the second pulse increases by an amount

$$\Delta e_o(2) = aE_A(2) = aE(1 - a)$$

The actual total voltage across the storage capacitor at the end of the second pulse is

$$e_o(2) = e_o(1) + \Delta e_o(2) = aE + aE(1 - a) = aE(2 - a)$$

When the third pulse occurs, the disconnect diode conducts when the pulse voltage rises above the cathode voltage, which is $e_0(2)$. Hence the applied voltage change is then

$$E_A(3) = E - e_o(2) = E - aE(2 - a) = E(1 - a)^2$$

and this divides between the two capacitors as before. Thus the output voltage at the end of the third pulse increases by

$$\Delta e_o(3) = aE_A(3) = aE(1 - a)^2 \tag{17.27}$$

Hence, by mathematical induction, the increase in voltage at the end of the n^{th} pulse is

$$\Delta e_o(n) = aE(1 - a)^{n-1} \tag{17.28}$$

This is the magnitude of the step increase.

The reliability of the counter is reduced when this step increase becomes so small that it is of the same order of magnitude as other voltage fluctuations in the circuit. Because it depends upon the capacitance ratio $a = C_c/(C + C_c)$, it is reasonable to believe that there is some specific value for a that will make Δe_o a maximum for a specified number of pulses n. This is the optimum operating condition, and it is easily determined by differentiating $\Delta e_0(n)$ with respect to a, setting the result equal to zero, and solving for a. This yields $a = 1/n$, or $C_c/(C + C_c) = 1/n$, so that

$$C = C_c(n - 1) \tag{17.29}$$

This is a useful design equation.

From this analysis it is clear that the output voltage waveform is a stair-stepped exponential. Thus the voltage at the end of any n pulses might be written

$$e_o(n) = A + B\varepsilon^{-kn} \tag{17.30}$$

where A, B, and k are undetermined constants. Three operating conditions are known:

(1) when $n = 0$; $\quad e_o(0) = 0$

(2) when $n = \infty$; $\quad e_o(\infty) = E$

(3) when $n = 1$; $\quad e_o(1) = aE$

When these conditions are substituted in equation (17.30), the result can be expressed as

$$e_o(n) = E(1 - \varepsilon^{-kn}) \tag{17.31}$$

where

$$k = \ln \frac{C + C_c}{C} \tag{17.32}$$

Solve these last two equations for the capacitance ratio and the result is

$$\frac{C_c}{C} = \left[\frac{E}{E - e_0(n)}\right]^{1/n} - 1 \tag{17.33}$$

For the optimum capacitance ratio it was previously shown that $C_c/C = 1/(n - 1)$. Consequently equation (17.31) can be expressed as

$$\frac{e_o(n)}{E} = 1 - \left(\frac{n - 1}{n}\right)^n \tag{17.34}$$

Under these optimized conditions,

$$\Delta e_o(n) = \frac{E}{n}\left(\frac{n - 1}{n}\right)^{n-1} \tag{17.35}$$

Hence, as soon as n and the pulse amplitude are fixed, equation (17.34) can be used to compute the actual output voltage available from the counter. Or, if n and $e_o(n)$ are specified, the required pulse amplitude can be computed.

The preceding discussion was concerned with counting positive pulses. Negative pulse counting is accomplished in the same way simply by reversing all diodes.

17.15. Counter Refinements and Recycling

The analysis of the counter circuit in the preceding section assumed the clamping voltage to be zero. However, by connecting the circuit as shown in figure (17.40a), the stair-stepped exponential can be clamped

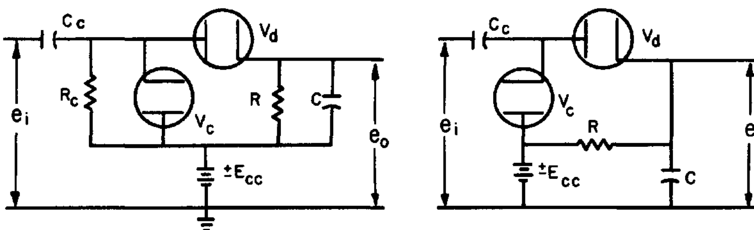


Fig. 17.40. Positive pulse counter circuit with output clamped to $\pm E_{cc}$. Reverse both diodes for negative pulse counting.

to any desired baseline. The same effect can be achieved with the circuit in (17.40b). In both circuits the resistor R should be large relative to the r_p of the diodes. In some cases R might not appear physically in the circuit because it might be a leakage current path.

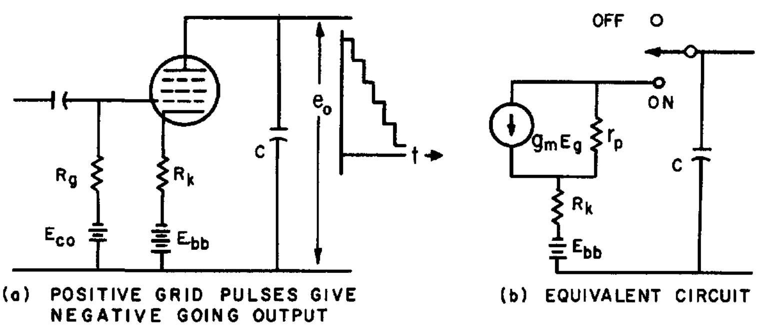


Fig. 17.41. Use of a pentode for constant current charging and linear stepping.

The principal shortcoming of all the counter circuits discussed so far is the nonuniformity in the size of successive steps. The general techniques for equalizing the sizes of the steps are essentially the same as those employed to linearize the output from a sawtooth voltage generator. For example, constant-current charging can be obtained with a pentode as shown in figure (17.41).

Miller integration can be used to equalize the steps, and the circuit connections and basic principles are the same as those treated in sections (17.8) and (17.9).

The output from a counter is generally connected to some switching circuit such as a multivibrator, blocking oscillator, phantastron, or a gas triode. These circuits, except the last one, are discussed in the next chapter. However, the use of a gas triode as an indicator of the count and as a mechanism for recycling the counter is easily explained from figure (17.42).

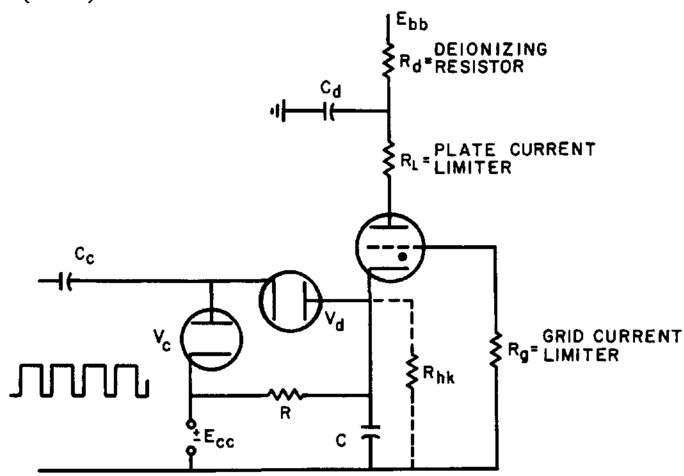


Fig. 17.42. Recycling a negative pulse counter with a thyratron.

The counter output is clamped to $+E_{cc}$. This voltage is large enough to hold the gas tube nonconducting until the required number of counts has occurred. As the pulses are counted, the cathode voltage steps down. During this interval the capacitor C_d charges to E_{bb} volts. When the required number of pulses have been counted, the cathode voltage has dropped to a point that allows the tube to fire. When the tube fires the plate voltage drops to 10 or 12 volts and the total voltage around the C_d, R_L, tube, C circuit is quickly redistributed. The size of C_d is selected so that the charge redistribution brings the voltage on the storage capacitor back to $+E_{cc}$. When this transient is complete, the only current available to the tube to sustain the gas discharge is that from E_{bb} through R_d. The value of R_d is made so large that it will not sustain the discharge, and the tube deionizes. The counter is thereby recycled and commences counting again in the usual manner.

17.16. Coincidence and Time Selection Circuits

It is frequently necessary and desirable to select some part of a waveform that occurs over a specified time interval and to reject all other parts of a waveform. It is also often necessary to indicate the coincidence or lack of coincidence of two or more electrical events. The circuits that perform these various services are identified as *gating*, *strobe*, and *coincidence* circuits. Because the process is essentially one of *time selection*,[5] this designation seems more appropriate.

Time selection circuits depend upon the switching properties of electronic devices. Vacuum diodes, varistors, transistors, special vacuum pentodes, and magnetic amplifiers are often used.

In general, the electronic component is held normally conducting or nonconducting. Some sort of *selector* pulse, or pulses, is used to reverse this operating condition when the desired part of the waveform is to be transmitted. From this general viewpoint it is clear that a simple time selection circuit can be formulated, using a diode clipper in which the selector pulse causes the clipping level to be a function of time. Thus, suppose that the signal and selector are added together and are then applied to a biased diode, as shown in figure (17.43). It is assumed that the diode is biased *off* in the absence of the selector pulse. The selector pulse causes the operating point of the diode to fall just short of the conducting point so that any superimposed positive signal will cause the diode to conduct and thereby develop a signal in R_k.

There are many variations of this basic circuit. For example, if negative pulses are to be selected, it is necessary only to reverse the diode and the bias voltage E_{kk}.

A triode or pentode clipper could be used in place of the diode in figure (17.43). However, the cutoff characteristics are not so sharp as those of semiconductor diodes, and the selector pulse tends to leak through.

The main disadvantage of the circuit in figure (17.43) is that its operation is critically dependent upon the shape and amplitude of the selector pulse. This defect can be eliminated to a certain extent by adding a second diode as shown in figure (17.44). In the absence of any input, the *gating* diode is normally conducting, and the values of R_L, R_1, and E_{kk} are chosen so that the plate-to-ground voltage of the clipper diode is nearly zero. Hence there is no output voltage, regardless of whether a signal input is applied or not. When the positive selector

[5] See Chance, *et al.*, *op. cit.*

pulse is applied, if its magnitude is large enough, the gating diode is cut off, and any coincident positive signal causes the clipper diode to conduct and develop an output.

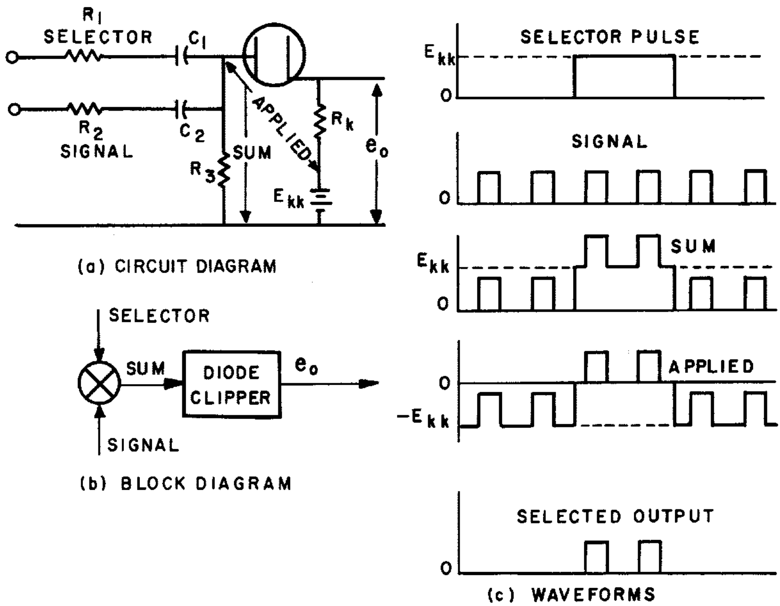


Fig. 17.43. Simple diode time selection circuit.

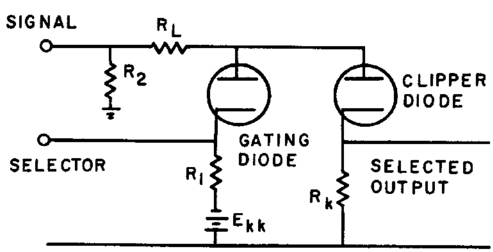


Fig. 17.44. Double diode time selector.

As before, the relationship between the selector and signal pulse and bias voltage is critical. The selector must be large enough to keep the gating diode cut off in the presence of a signal. This is a definite drawback.

The ideal time selector obviously has an output only during the selected interval and its operation is insensitive to variations of pulse amplitude within certain limits. In such an ideal case, the operation of the circuit depends only upon the coincident presence of a selector pulse and a signal to develop an output. The absence of either will result in zero output.

This idealized operation can be closely approximated within a predetermined range of amplitudes by the circuit[6] shown in figure (17.45). This circuit has the added advantage of being adaptable to n input circuits, so that coincidence of any degree of multiplicity can be obtained.

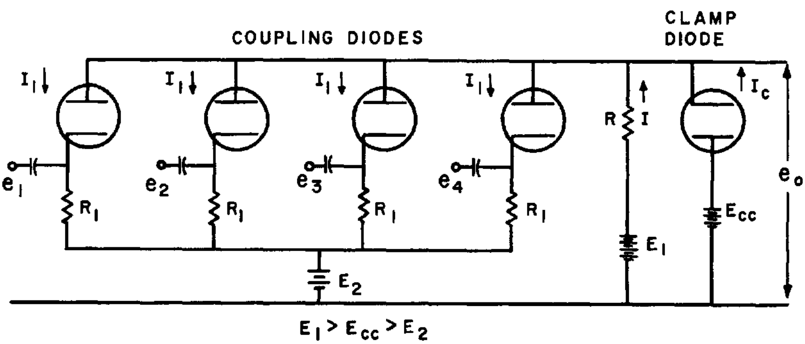


Fig. 17.45. A very good diode coincidence circuit.

The various voltages are related as follows: $E_1 > E_{cc} > E_2$. The values of R_1 and R are adjusted so that I is less than I_1. Hence nearly all the current through the coupling diodes must be supplied by E_{cc} through the clamper diode V_c. It is assumed that the r_p of the diodes is small in comparison with the load resistances R_1.

In the absence of any signal inputs, the output voltage e_0 is clamped to E_{cc} by the clamper diode. All diodes are conducting because of the relationships between the circuit voltages. The voltage drops across the diodes are negligible because r_p is much less than R_1; this makes the cathode-to-ground voltages practically equal to E_{cc}. As a result, any positive going pulse input will cause the corresponding coupling diode to cut off. However, the output remains clamped to E_{cc} because the clamper diode continues to conduct to supply current to the other coupling diodes. However, if *all* coupling diodes are simultaneously

[6] Taken with permission from Tung Chang Chen, "Diode Coincidence and Mixing Circuits in Digital Computers," *Proc. IRE*, vol. 38, May, 1950, pp. 511–514.

cut off by the coincident arrival of positive pulses at the inputs, the clamping diode also cuts off because the current requirement drops to zero. As a result, the output voltage increases toward E_1. The increase will be exponential because I flows through R in charging the shunt capacitance.

The actual final value of the output voltage will be determined by the amplitude of the *smallest* coincident input pulse. As soon as the output voltage increases by an amount equal to this pulse amplitude the voltage across that diode is zero. A further increase in output voltage causes the diode to conduct and thereby prevents the voltage from increasing beyond this point. This is not necessarily a serious limitation.

With the foregoing points in mind, it is clear that the output waveform will appear as shown in figure (17.46). In this figure, $E - E_{cc}$ = smallest input pulse amplitude; C = total shunt capacitance; T_R = rise time; T_{ch} = charging time constant = RC. The rise time is easily computed from the general charging equation for a capacitor, which is

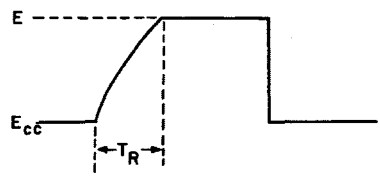


Fig. 17.46. Output from circuit of figure (17.45).

$$e_c = E_A - (E_A - \gamma)\varepsilon^{-t/RC}$$

In this case we want to find $t = T_R$, where

$$e_c = E; \qquad E_A = E_1; \qquad \gamma = E_{cc}$$

Hence at the end of the rise time,

$$E = E_1 - (E_1 - E_{cc})\varepsilon^{-T_R/RC}$$

This equation is easily solved for T_R, yielding

$$T_R = RC \ln \frac{E_1 - E_{cc}}{E_1 - E}$$

One other basic design criterion that must be remembered is that the clamping diode must carry a current equal to the total current through the coupling diodes.

One final circuit in common use is the pentode selector shown in figure (17.47). The tube, a 6AS6 for example, has a sharp cutoff characteristic for both the control and suppressor grids. The tube is biased for class B operation on the control grid so that space current flows only when permitted to do so by a positive selector pulse on the normally off control grid. Current flows in the plate circuit only when the suppressor voltage is brought above cutoff by a positive pulse. This is an excellent circuit.

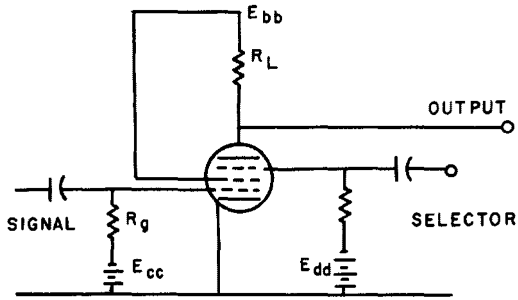


Fig. 17.47. Pentode coincidence circuit.

17.17. Effects of Shunt Capacitance

All the material in this chapter ignored the inevitable existence of interelectrode and distributed wiring capacitances. Thus experimental results will differ somewhat from computed predictions.

The problem becomes acute when the signals in the circuit have steep leading and trailing edges, have sharp discontinuities, and have short durations. When these conditions prevail, the shunt capacitance reduces the slopes of edges, smears out sharp discontinuities, and tends to broaden pulses. It is usually difficult to calculate these effects, because of the mathematical complexity of the circuit formulation.

The problem of preserving the discontinuous character of the outputs from wave-shaping circuits is essentially the same as that involved in the faithful transmission of pulses through video amplifiers. The same considerations generally apply. Thus wave-shaping circuits are essentially video circuits operating in the switching mode and should be treated as such in design and construction.

17.18. Magnetic Amplifier Coincidence Circuit[7]

The operation of a single core magnetic amplifier was covered in chapter 16 and should be reviewed if you are unfamiliar with the circuit.

[7] R. A. Ramey, "The Single Core Magnetic Amplifier as a Computer Element," *Communications and Electronics*, January, 1953, pp. 442–446.

This magnetic amplifier is easily converted into a coincidence circuit with n inputs by the connections shown in figures (17.49) and (17.50).

The hysteresis loop of the core is shown in a rather idealized form in figure (17.48). The core is saturated on the positive swing of e_{ac}. In the absence of any voltages from the signal input, the core remains saturated because diode A disconnects the supply when the supply voltage reverses. Nothing further can happen and no output will appear until the core flux is reset to its original value ϕ_0 by a reverse current through the coil. However, because of the multiplicity of low impedance paths through the signal sources, current does not flow through the coil on the negative swing of e_{ac}, but through the diodes and signal sources instead. Thus the core is not reset.

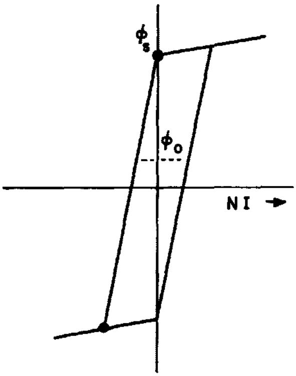


Fig. 17.48. Hysteresis loop of a magnetic amplifier core.

As long as any one signal source voltage is zero, the reverse current is bypassed around the coil. However, if signals appear at all inputs

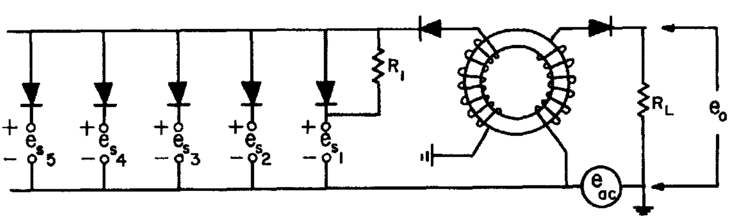


Fig. 17.49. One type of single core magnetic amplifier coincidence circuit.

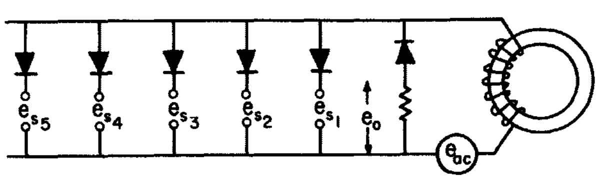


Fig. 17.50. Another type of single core magnetic amplifier coincidence circuit.

coincidentally, current will flow through the coil, and this will reset the core flux to ϕ_0. On the *next* half cycle of the supply voltage the core will again saturate, and the whole process repeats. The existence of coincidence is indicated by the appearance of an output e_o on the half cycle following coincidence.

The signals must exist in the input circuits of the counter for an appreciable portion of the reset half cycle. The signal amplitudes must be large enough to prevent the current from being bypassed around the coil.

PROBLEMS

17.1. Using the published data for 6AL5, determine the equivalent circuit of the circuit shown in figure (17.51) and compute the output voltage if the input has the form shown in the figure.

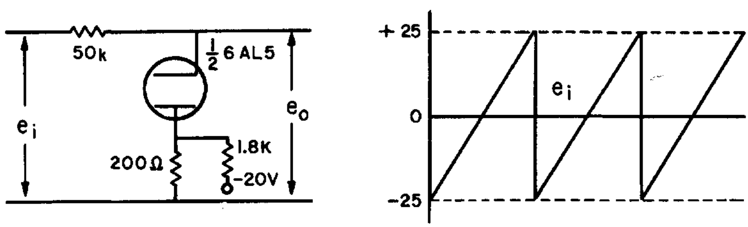


Fig. 17.51.

17.2. A clamping circuit, together with the input signal waveform, are shown in figure (17.52). Compute the output voltage waveform, assuming the capacitor is initially uncharged.

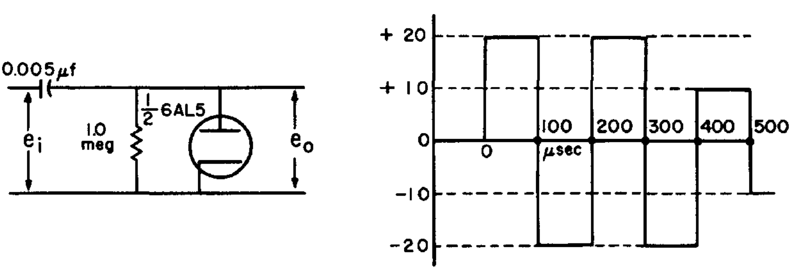


Fig. 17.52.

17.3. For the cathode follower and input signal shown in figure (17.53), compute and plot the output voltage waveform.

17.4. The signal shown in figure (17.54) is applied through a conventional *RC* coupling circuit to the grid of a 6J5 clipper. The time constant of the coupling circuit is long compared with the time *T*. Assuming a 20,000 ohm load resistance and $E_{bb} = 200$ v, compute the grid bias required to make the tube clip off all the signal except a triangle 10 μsec wide at the base.

17.5. A 1000 cps symmetrical square wave, 50 v peak to peak, is to be used to produce the waveform shown in figure (17.55). You are free to design any

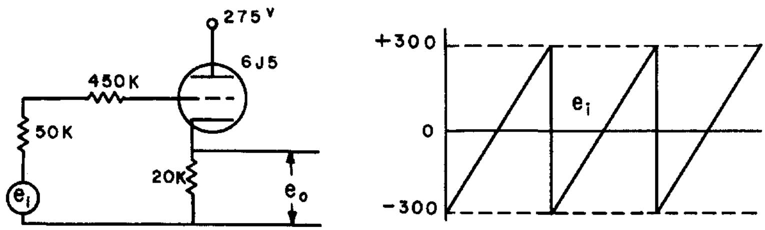


Fig. 17.53.

circuit of your choice, but the design must be complete in every detail of bias, time constant, etc.

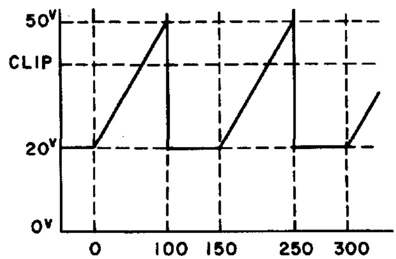


Fig. 17.54.

17.6. Design an *RLC* peaker using a 6AC7 tube (switching mode) for optimum performance. It is known that the power supply voltage is 300 v and the total interstage shunt capacitance is 20 $\mu\mu$f. The maximum pulse

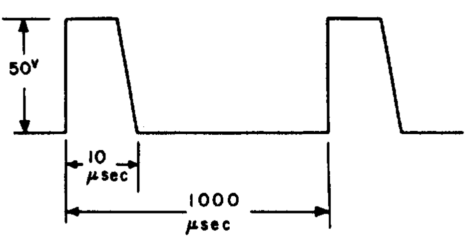


Fig. 17.55.

amplitude in the output is to be 5 v. Compute the necessary values for R_L and L and the ratio of the output pulse amplitudes.

17.7. Design an *RLC* ringing circuit using a 6AC7 switch tube with $E_{bb} = 300$ v and $C_T = 20$ $\mu\mu$f. The tube is gated by a negative pulse 100 μsec in duration. The circuit is to ring at 100 kcps during this interval. Compute the necessary values for R_L, L, and the peak value required for the negative gating signal to assure proper operation.

17.8. Design a simple *RC* sawtooth voltage generator using a 6J5 switch tube and a 300 v plate supply. The switch tube is gated by a 500 cps symmetrical square wave. Determine the values of R and C required to locate the base of the wave at 15 v and to cause it to rise to 50 v.

17.9. An *RC* trapezoidal voltage generator supplies grid excitation to a sweep amplifier tube. The switch tube conducts for 200 μsec and is non-conducting for 300 μsec. The current through the sweep coil is to change by 100 ma during the sweep. The sweep amplifier tube has a μ of 10 and $r_p = 5000$ ohms. For the switch tube, which is a 6J5, $E_{bb} = 200$ v, $C = 0.001$ μf, $R_L = 500{,}000$ ohms, $R_1 = 10{,}000$ ohms. Determine the required dimensions of the trapezoid and the values of L and R for the sweep coil.

17.10. A counting circuit is to reach a voltage output of 21.0 v at the end of the 6th count. Design an appropriate circuit and determine the required value for the input pulse amplitude. The pulses occur at a fixed frequency of 1000 pps and have a duration of 2 μsec. The pulses are negative.

17.11. Design a recycling circuit of the type shown in figure (17.42) for the counter of problem (17.10). Use an 884 gas triode. Compute the required values for all circuit elements and voltages.

17.12. Design a pentode coincidence circuit using a 6AS6 tube with $E_{bb} = 300$ v. Both incoming pulses are 1 μsec in duration and 5 v high.

Chapter 18

TRIGGER CIRCUITS AND NON-SINUSOIDAL OSCILLATORS

The preceding chapter dealt with a broad grouping of circuits that use electronic devices operating in the switching mode to accomplish things impossible with linear bilateral circuits. All these circuits were either open loop, without feedback, or negative feedback was used to improve certain operating characteristics.

The circuits used in this chapter are closely related to those in chapter 17 except that all the circuits involve amplifiers with large amounts of positive feedback. Thus they are *regenerative* wave shaping circuits as opposed to the nonregenerative circuits of the preceding chapter. In this text, circuits fitting this category are called *trigger circuits*. The relationship between the wave shaping circuits of chapter 17 and the trigger circuits of this chapter is essentially the same as that existing between class A amplifiers and oscillators.

For the purposes of discussion and analysis, trigger circuits can be subdivided into four different classes as: (1) multivibrator type circuits; (2) phantastron type circuits; (3) blocking oscillator circuits; (4) transistor trigger circuits. In general, the chapter is subdivided along these same lines, though equal emphasis is not given.

18.1. Classification of Trigger Circuits

It was shown in chapter 10 that certain circuits exhibit a negative resistance region as a part of their d-c characteristics. Such characteristics were found to be of two basically different types, short circuit stable and open circuit stable. The current-voltage characteristics corresponding to this terminology are given in figure (18.1). The discussion in chapter 10 went on to show that operating points in regions 1 and 3 were always stable while an operating point in region 2 could be unstable or stable. You should review chapter 10 if you are unfamiliar with these ideas.

Trigger circuits are designed so that they have a current-voltage characteristic like one of those in figure (18.1) and so that operation in region 2 is unstable. Thus all trigger circuits have three possible states of operation, two being stable and one being unstable. When operation changes from one stable state to the other, the circuit is said to have been *triggered.* Sometimes a circuit is capable of *self-triggering*, while an externally supplied trigger pulse may be required in other instances.

With the foregoing points in mind, trigger circuits can be classified as follows:[1]

(1) *Bistable.* This circuit is incapable of self-triggering. Hence two trigger pulses are required to complete one cycle of operation from stable state 1 to stable state 3, and then back to 1.

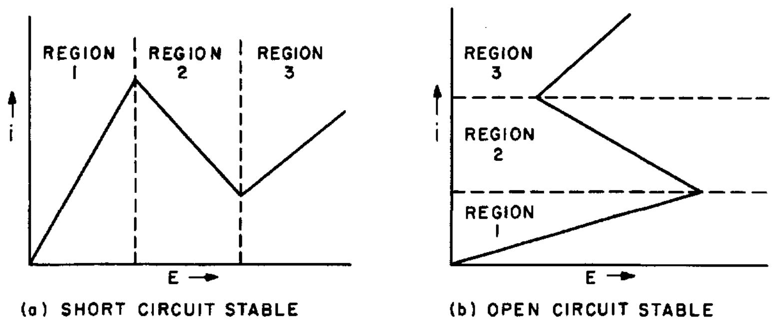


Fig. 18.1. Types of negative resistance characteristics.

(2) *Monostable.* This kind of circuit is capable of self-triggering from *one* of its stable states. An externally supplied trigger pulse is required to change operation from the other stable state. Thus the circuit generates one complete cycle in response to a single external trigger pulse.

(3) *Astable.* This type of circuit is capable of self-triggering from both stable states. Thus it generates a continuous train of nonsinusoidal waves without any externally supplied trigger pulses.

18.2. Basic Multivibrator Type Circuit

The multivibrator type of trigger circuit is essentially a two-stage resistance coupled amplifier with the output from the second stage connected back to the input of the first. Thus the block diagram and circuit diagram appear as shown in figure (18.2). The circuit is easily

[1] Based upon Chance, *et al.*, *Waveforms*, vol. 19, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1949, chap. 5.

identified as the Eccles-Jordan circuit noted in chapter 10. It is also known as a *scale-of-two* circuit.

It is clear from the block diagram of figure (18.2a) that the circuit uses 100% positive feedback and that the feedback loop is closed only when both of the amplifier tubes are conducting. The condition of 100% positive feedback is an unstable one and corresponds to region 2, the negative resistance region, of the d-c characteristics.

The positive feedback loop is open and the circuit is stable whenever one tube is nonconducting. Thus there are two stable states, corresponding to regions 1 and 3 of the d-c characteristics, as follows: (1) V_1 conducting and V_2 cutoff; (2) V_2 conducting and V_1 cutoff. Thus,

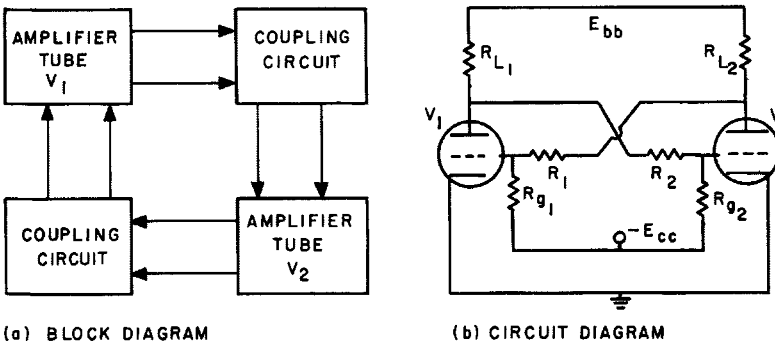


Fig. 18.2. Essential features of the basic multivibrator type of circuit.

under normal conditions, the circuit will operate quiescently in one of its two stable states.

Now suppose that the circuit is in one of its two stable states. To make it switch over to the *other* stable state it is necessary to create an unstable condition by momentarily making both tubes conduct. This action closes the feedback loop so that regenerative switching takes place. The feedback loop is usually closed by inserting a small triggering impulse into the circuit in such a way that it appears positive on the grid of the *off* tube. If it has sufficient amplitude, the trigger will make the tube momentarily conduct, the feedback loop closes, and the resulting instability causes operation to switch over to the other stable state.

All multivibrator type circuits operate according to this general outline. In bistable circuits, d-c coupling is used between the tubes. In the monostable multivibrators, one d-c coupling circuit is replaced by a-c coupling, and the action of this circuit produces self-triggering from one stable state. Astable circuits are created by using a-c coupling for both coupling circuits, so that self-triggering from both stable states is possible. The block diagrams of figure (18.3) illustrate this idea.

The mechanism involved in self-triggering is relatively easy to understand. For example, consider the monostable multivibrator block diagram. When the trigger is applied to the normally off tube, regenerative switching occurs, with the result that it is made conducting while the normally on tube is now cut off. The function of the a-c coupling is to provide a time varying grid voltage on the grid of the

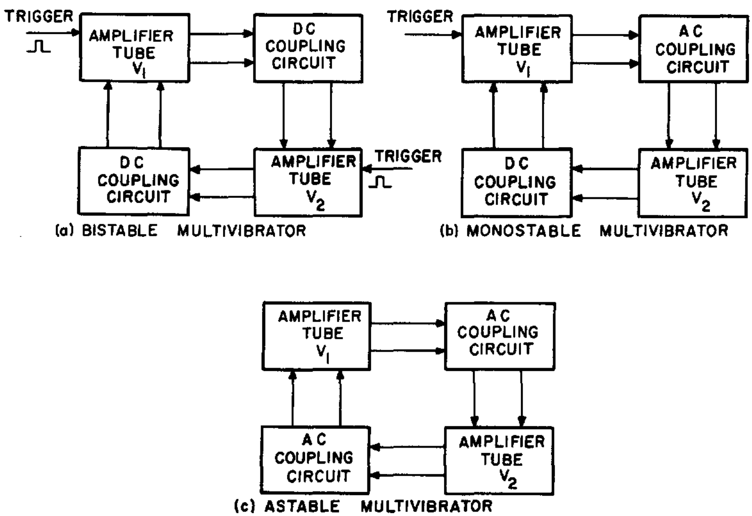


Fig. 18.3. Types of multivibrator (block diagrams).

tube that is now off; at some specified instant this voltage has the proper value to make the tube conduct. This closes the feedback loop and the circuit is self-triggered back to the original stable state. The actual process involved here is explored in some detail in later sections.

18.3. Bistable Multivibrator

The complete circuit diagram of a bistable multivibrator is given in figure (18.4). This is essentially the same as the Eccles-Jordan circuit with a cathode bias circuit instead of a separate grid supply. Also, two capacitors, C_1 and C_2, have been added. This is a refinement that will be explained later. This is called a *plate coupled* multivibrator because both interstage coupling circuits are from plate to grid. Disregarding C_1 and C_2, the circuit operates as explained in the preceding section, requiring the use of positive trigger pulses of sufficient amplitude to make the tube conduct and close the feedback loop momentarily.

The inevitable presence of tube interelectrode capacitance, especially the input capacitance, is unfortunate because it is directly in shunt with the gridleak resistor which is a part of the d-c coupling circuit. Because of the steep edge and short duration of the usual triggering impulses, this capacitance provides a low impedance path to ground for the trigger. As a result, it appears in reduced amplitude on the grid of the tube and might *not* have sufficient amplitude to make the tube conduct.

The *speed up* capacitors C_1 and C_2 are added to the circuit to compensate for the shunting effect of the tube input capacitances. The

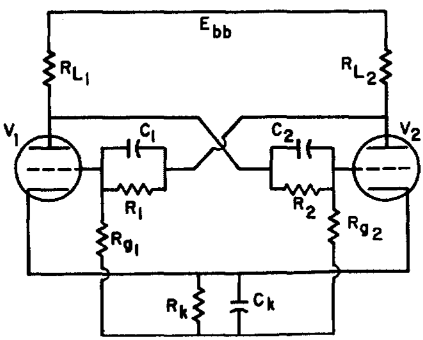


Fig. 18.4. Plate coupled bistable multivibrator.

compensation is achieved by making the speedup capacitors large compared to the tube capacitances. The capacitive voltage divider so formed causes most of the trigger pulse to appear across the tube input capacitance, and hence on the grid of the tube.

The cathode bias requirements for the circuit are not severe, because only one tube conducts at a time except during the switching interval. Thus the current through R_k is nearly constant when tubes of the same type are used.

The equivalent plate circuit for either of the two amplifiers in the multivibrator is easily constructed, as shown in figure (18.5). The cathode bias circuit is replaced by an equivalent battery E_{kk}. The other half of the multivibrator has exactly the same equivalent circuit, with different component identification.

The speedup capacitors were omitted from the equivalent circuit along with the input capacitance of the tubes because they affect only the transition interval. This problem is treated briefly in a later section.

It should be borne in mind that the grid of either tube might be driven positive. If this occurs, the equivalent circuit should show R_g in parallel with a series combination of r_g and E_{kk}, where r_g is the equivalent grid-to-cathode resistance.

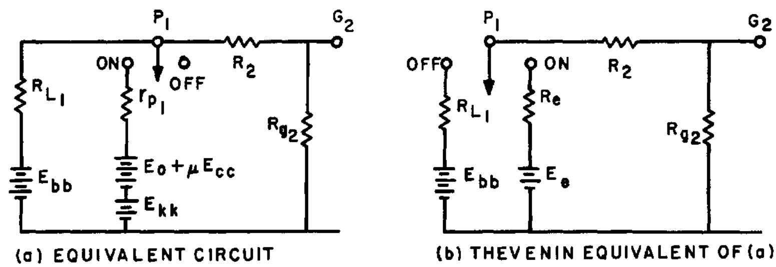


Fig. 18.5. Equivalent plate circuits of one amplifier section of a bistable plate coupled multivibrator.

The derivation of design equations and the computation of the various circuit voltages is a straightforward application of circuit theory to the equivalent circuit. This is left as an exercise for the reader.

18.4. Bistable Multivibrator, Variations and Refinements

The basic bistable multivibrator is widely used, but it suffers from a serious disadvantage inherent in plate coupling. The difficulty arises

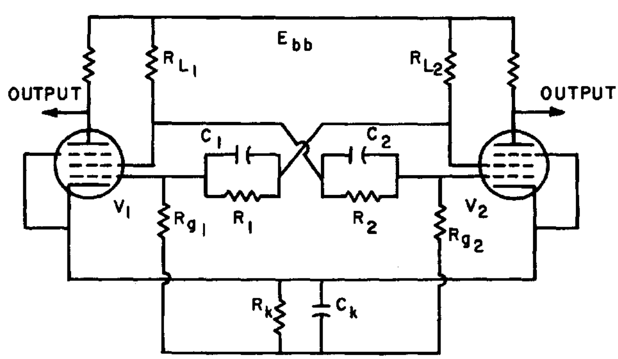


Fig. 18.6. Electron coupled, bistable, plate coupled multivibrator.

because the waveforms generated by these circuits are usually picked off at some point in the circuit and coupled to some other circuit. Because of the coupling, the circuit receiving the multivibrator signal is unavoidably included in the multivibrator circuit and thereby alters its characteristics.

This disadvantage can be largely overcome by using *electron coupling* as shown in figure (18.6). Here pentodes are connected so that the screen serves as the plate of the multivibrator tube. The signal is then taken from the actual plate of the tube and this is isolated from the multivibrator.

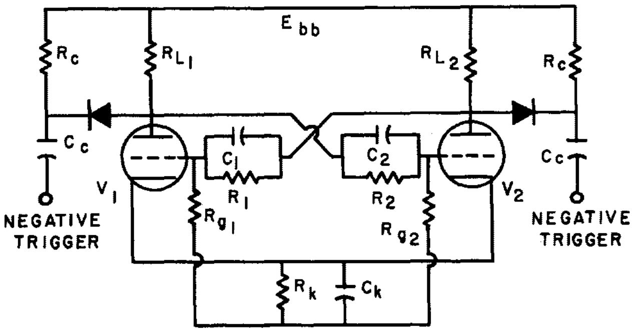


Fig. 18.7. Use of diodes for trigger insertion.

The same sort of problem exists with regard to the insertion of the triggering impulses because the inserting device must not have an appreciable effect upon the action of the multivibrator circuit. The

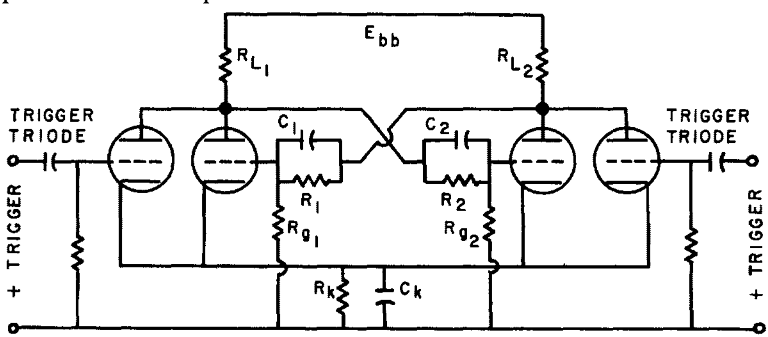


Fig. 18.8. Use of trigger triodes for trigger insertion.

simplest method of doing this is to use *coupling diodes* as shown in figure (18.7). The diodes conduct only during the application of a short negative pulse. They are open circuits at all other times and effectively disconnect the triggering signal source from the multivibrator.

An alternative method for trigger insertion requires the use of trigger triodes connected directly in parallel with the multivibrator tubes as shown in figure (18.8). The trigger tubes are biased off at all times except when a positive trigger pulse appears at the grid. The tube then conducts momentarily and a negative pulse appears at the plate of the multivibrator tube. This pulse is coupled over to the grid of the next tube, amplified, inverted, and applied to the grid of the original tube where it produces triggering by turning the tube on.

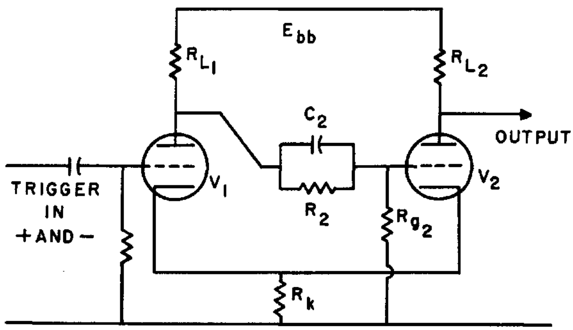


Fig. 18.9. Bistable cathode coupled multivibrator.

The whole problem of coupling in and out of multivibrator circuits can be approached from a different angle by a simple change in circuitry. By removing the bypass capacitor from the cathode circuit and by

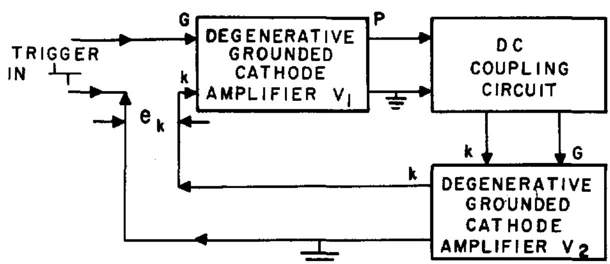


Fig. 18.10. Block diagram representation of a cathode coupled bistable multivibrator.

removing one of the plate-to-grid coupling circuits, the bistable *cathode coupled* multivibrator of figure (18.9) is obtained.

Assume that operation is in the stable state corresponding to V_1 off and V_2 on. The plate current in V_2, flowing through R_k, develops enough bias on V_1 to hold it cut off. In the reverse stable state, the constants of the d-c coupling circuit and the current through V_1 are sufficient to hold V_2 cut off.

The advantages of this circuit are apparent from figure (18.9).

Coupling in and coupling out are accomplished in such a way that the coupling circuits are isolated from the elements comprising the multivibrator.

The general nature of the circuit may be more obvious from the block diagram given in figure (18.10).

18.5. Monostable Plate Coupled Multivibrator, Operation

Monostable operation is obtained from the basic bistable circuit by replacing one of the d-c coupling circuits with a conventional *RC* interstage network as shown in figure (18.11a). Under normal conditions, V_1 is held nonconducting by the bias voltage E_{cc} and the

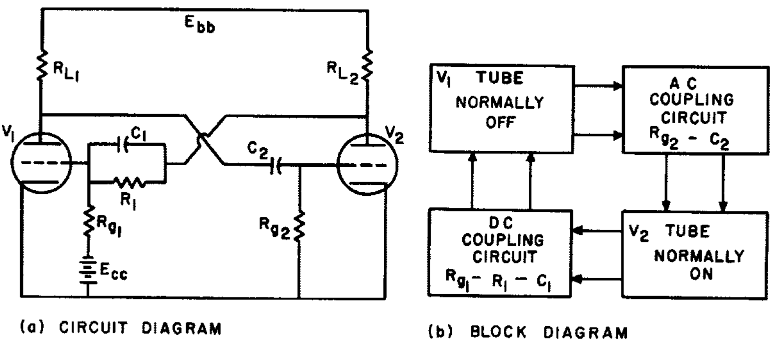


Fig. 18.11. Monostable plate coupled multivibrator.

constants of the d-c coupling circuit; V_2 is normally conducting. The circuit *cannot* trigger itself from this condition and this is called the *normal* state. When an externally supplied positive pulse of sufficient amplitude is applied to the grid of V_1, the feedback loop closes and the circuit undergoes regenerative switching to the other stable state. The general problem may be more apparent from the block diagram in figure (18.11b).

Following the insertion of the trigger, the circuit operation is in the other stable state corresponding to V_1 conducting and V_2 cut off. This is called the *timing state.*

Operation in the timing state is easily understood. When the trigger is first applied to the grid of V_1 in the normal state, it conducts, and its plate voltage drops precipitously. The voltage across the capacitor C_2 cannot change instantaneously, so that the grid of V_2 drops by the same amount as the plate of V_1. This drives the grid voltage on V_2 below cutoff, and conduction in V_2 ceases. The plate voltage of V_2 rises to

E_{bb} and this increase is coupled over to the grid of V_1 and renders V_1 highly conducting.

Following this, C_2 commences discharging through R_{g_2}, R_L, and V_1; the resulting transient generates a voltage across R_{g_2}, on the grid of V_2, that rises exponentially from a low negative value toward zero. As soon as this voltage decays to the cutoff potential of V_2, the tube conducts, closing the feedback loop, and causing regenerative switching that ends when the original *normal* state is re-established. The circuit has clearly triggered itself. It then awaits another external trigger, having executed one complete operating cycle in response to the single trigger pulse.

The self-triggering is accomplished through the transient discharge of C_2. Thus C_2 largely controls the duration of the generated pulse, and is therefore called the *timing capacitor.* The timing is also affected by V_1, R_{L_1}, and R_{g_2}. However, R_{g_2} is generally so large that the other factors are relatively unimportant. Hence it is called the *timing resistance.* A detailed analysis of the multivibrator is given in the next section; it is primarily concerned with the timing state.

18.6. Monostable Plate Coupled Multivibrator, Analysis

The duration of the generated pulse and the waveforms appearing at various points in the circuit of the monostable multivibrator of figure (18.11) can be evaluated rather simply from the equivalent plate circuits. The best way of going about this is to construct the equivalent plate circuits of the two amplifiers separately for each of the two stable states of operation. Thus, in figure (18.12a), the two equivalent plate circuits for the normal state (V_1 off and V_2 on) are shown. Figure (18.12c) shows the equivalent plate circuits in the timing state (V_1 on and V_2 off). All these circuits are easily simplified to series *RC* circuits by using Thevenin's theorem at the indicated terminals. The results of this simplification may then be observed in figure (18.12b). These are the circuits that will be used in the mathematical formulation because they have precisely the same form as the general case worked out in figures (17.13) and (17.14).

Assume that the multivibrator has been in its normal state for a relatively long time so that all transients may be assumed to be complete. The equivalent circuits of figure (18.12b) apply and the currents must be zero if transient effects are actually complete. Thus the voltage across C_2 must be E_{bb}. This is the voltage across the capacitor at the instant the trigger pulse is applied. Hence, the initial capacitor voltage for the timing state, which follows the trigger, is clearly $\gamma = E_{bb}$.

Now, if the trigger has been applied, then 0^+ seconds later, the circuit is in the timing state and the equivalent circuits of figure (18.12c) apply.

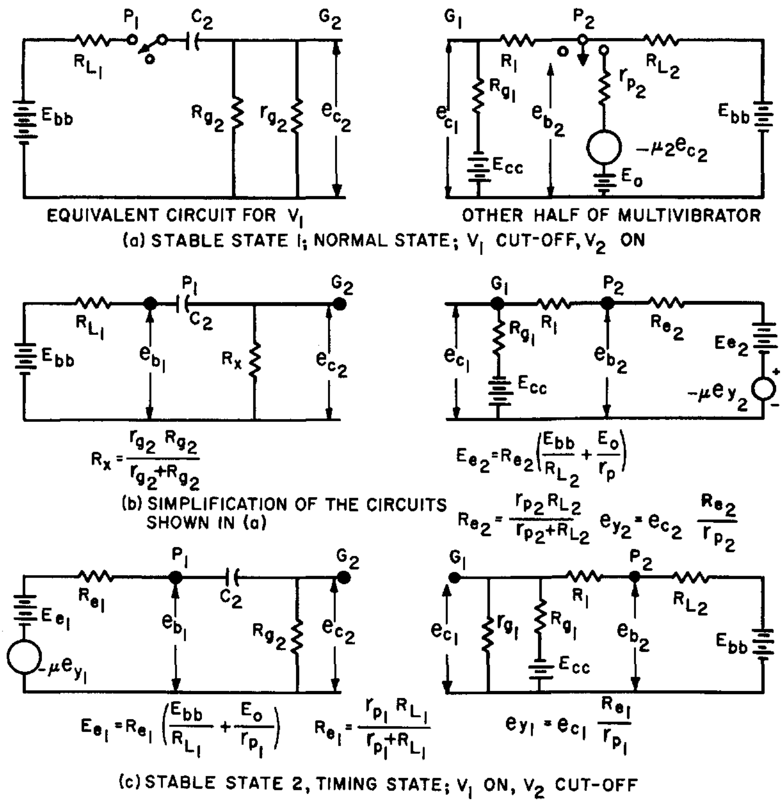


Fig. 18.12. Equivalent circuits of a monostable plate coupled multivibrator.

The voltage applied to the capacitor of this circuit is clearly

$$E_A = E_{e_1} - \mu_1 e_{y_1}$$

$$= E_{bb}\frac{R_{e_1}}{R_{L_1}} + (E_0 - \mu_1 e_{c_1})\frac{R_{e_1}}{r_{p_1}}$$

The value of e_{c_1} can be computed from the equivalent circuit of V_2 during the timing period. A quick calculation shows that e_{c_1} will be slightly positive, possibly by a fraction of a volt. Hence μe_{c_1} and E_o are of about the same size. Thus, without much loss in accuracy, it is reasonable to assume that $(E_0 - \mu_1 e_{c_1}) \doteq 0$. Consequently the voltage applied to the *RC* circuit in the timing state is

$$E_A \doteq E_{bb} \frac{R_{e_1}}{R_{L_1}}$$

Using the formulas in figure (17.14), the transient current through the circuit is

$$i = \frac{E_A - \gamma}{R_{e_1} + R_{g_2}} \varepsilon^{-t/(R_{e_1}+R_{g_2})C_2}$$

Therefore the voltage appearing on the grid of V_2 is

$$e_{c_2} = iR_{g_2} = (E_A - \gamma) \frac{R_{g_2}}{R_{e_1} + R_{g_2}} \varepsilon^{-t/(R_{e_1}+R_{g_2})C_2} \tag{18.1}$$

Substitute for E_A and γ, and the result can be expressed as

$$e_{c_2} = -\frac{E_{bb}}{r_{p_1}} \cdot \frac{R_{e_1} R_{g_2}}{R_{e_1} + R_{g_2}} \varepsilon^{-t/(R_{e_1}+R_{g_2})C_2} \tag{18.2}$$

At this point it is wise to stop and consider the terms in this equation rather carefully. Because R_{e_1} was defined on figure (18.12c) as

$$R_{e_1} = \frac{r_{p_1} R_{L_1}}{r_{p_1} + R_L}$$

then it is clear that the factor

$$R = \frac{R_{e_1} R_{g_2}}{R_{e_1} + R_{g_2}} \tag{18.3}$$

is the total parallel combination of all the resistances in the plate circuit of the V_1 amplifier. That is, it is exactly equal to the equation for R defined for resistance coupled amplifiers in chapter 4. It is the mutual impedance of the plate circuit of the V_1 amplifier under nonreactive conditions. Also, the quantity in the exponent is

$$\omega_1 = \frac{1}{(R_{e_1} + R_{g_2})C_2} \text{ rps} \tag{18.4}$$

where ω_1 = lower cutoff frequency of the resistance coupled V_1 amplifier. Finally, if you recall that $g_{m_1} = \mu_1 / r_{p_1}$, then when all these substitutions are made, the equation for the grid voltage in (18.2), reduces to

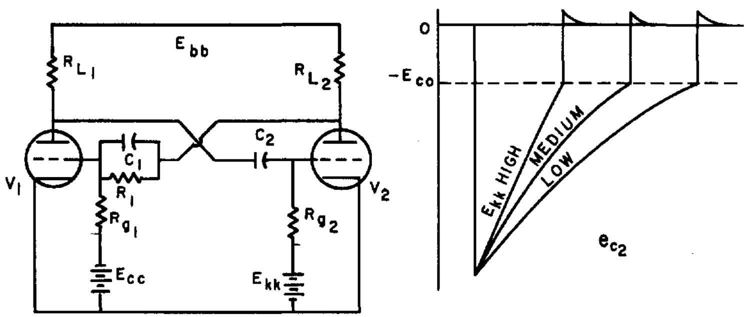
$$e_{c_2} = -\frac{E_{bb}}{\mu_1} (g_{m_1} R) \varepsilon^{-\omega_1 t} \tag{18.5}$$

Of course, you remember from chapter 4 that $g_{m_1}R = A_{r_1}$ = reference gain of the V_1 amplifier. Hence

$$e_{c_2} = -\frac{E_{bb}}{\mu_1} A_{r_1}\varepsilon^{-\omega_1 t} \tag{18.6}$$

Thus, during the timing state, the grid voltage on V_2 has the form shown in figure (18.13a).

The duration of the timing state is equal to the time required for



(a) CIRCUIT DIAGRAM

(b) TIMING WAVEFORM SHOWING THE EFFECT OF VARYING E_{kk}

Fig. 18.13. Monostable plate coupled multivibrator with positive grid return.

e_{c_2} to reach $-E_{co_2}$ because self-triggering occurs at this instant. If this time is specified as T, then

$$e_{c_2} = -E_{co_2} = -\frac{E_{bb}}{\mu_1} A_{r_1}\varepsilon^{-\omega_1 T}$$

Solve for T.

$$T = \frac{\ln A_{r_1} + \ln (E_{bb}/\mu_1 E_{co_2})}{\omega_1} \tag{18.7}$$

This is an important and useful result. For example, it shows that multivibrators can be designed from the data generally available on resistance coupled amplifiers because A_{r_1} = reference gain of the V_1 amplifier; ω_1 = lower cutoff frequency of the V_1 amplifier. It clearly shows that the low frequency characteristic of the amplifier largely controls the timing and that short pulse durations are achieved at the expense of gain. It also shows that short pulse durations require the second factor $\ln (E_{bb}/\mu_1 E_{co_2})$ to be as small as possible. This requires the use of tubes having nearly straight, parallel, sharp cutoff plate characteristics, because under these conditions, $\mu_2 = E_{bb}/E_{co_2}$ and

$$\ln \frac{E_{bb}}{\mu_1 E_{co_2}} \doteq \ln 1 = 0$$

The amplitudes of the positive and negative pulses at the two plate terminals are easily computed from the waveforms given in figure (18.13), with terms defined in figure (18.12).

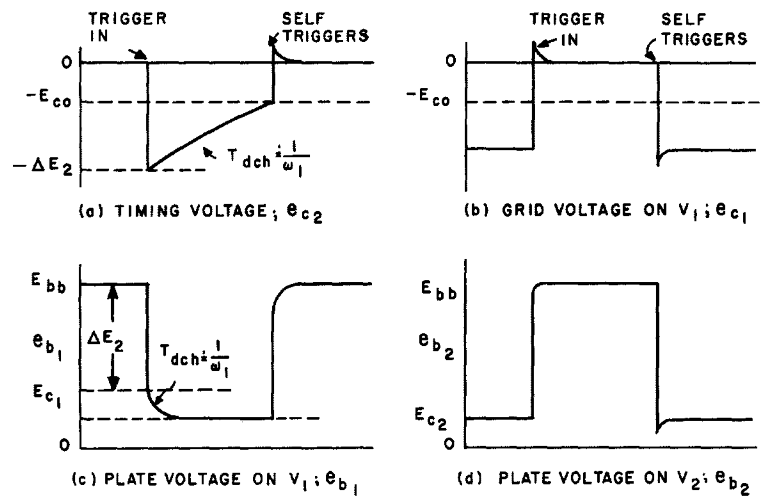


Fig. 18.14. Voltage waveforms in a monostable plate coupled multivibrator.

The operating characteristics of the circuit can be improved by returning the timing resistor R_{g_2} to some arbitrary voltage shown as E_{kk} in figure (18.14)$_s$ The effect of this change on the timing waveform is shown in the same figure.

The alteration of the grid return voltage causes two main effects:

(1) The pulse duration is more accurately defined because the angle of interception of e_{c_2} with $-E_{co}$ is less oblique.

(2) The pulse duration is shortened.

Because of the greater stability of pulse duration, large values for E_{kk} are often used. It is common practice to connect R_{g_2} directly to E_{bb}.

The circuit of figure (18.14) can be analyzed in exactly the same way as the circuit with grounded timing resistor. It is left as an exercise for the reader to show that the resulting equation for the pulse duration is approximately

$$T \doteq \frac{1}{\omega_1} \ln \frac{1 + A_r E_{bb}/\mu_1 E_{kk}}{1 + E_{co_2}/E_{kk}} \tag{18.8}$$

if it is assumed that (1) R_{g_2} is much larger than r_{p_1} and R_{L_1}; (2) $E_{o_1} \doteq \mu_1 e_{c_1}$. If E_{kk} is zero, the equation reduces to equation (18.7). If $E_{kk} = E_{bb}$,

$$T \doteq \frac{1}{\omega_1} \ln \left(\frac{1 + A_{r_1}/\mu_1}{1 + E_{co_2}/E_{bb}} \right) \tag{18.9}$$

It is clear from equation (18.8) that variations in the grid return voltage will have little effect until E_{kk} approaches and exceeds the cutoff voltage of V_2.

18.7. Monostable Cathode Coupled Multivibrator, Normal State

By removing the one d-c coupling circuit from the monostable plate coupled multivibrator and including a common unbypassed cathode

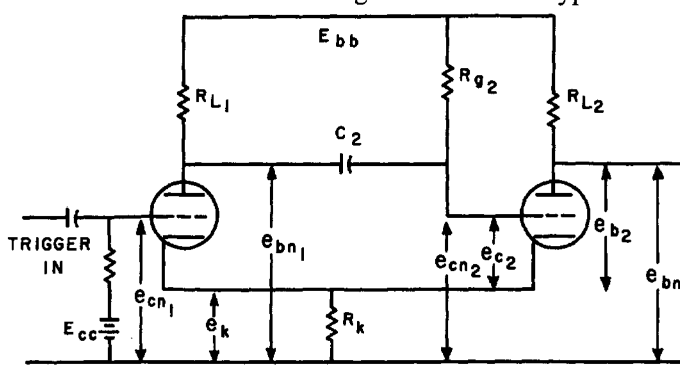


Fig. 18.15. Monostable cathode coupled multivibrator.

resistor, the monostable *cathode coupled* multivibrator of figure (18.15) is obtained. In the *normal* state, from which self-triggering is not possible, V_1 is cut off and V_2 is conducting. This condition is assured by returning the gridleak resistor of V_2 to E_{bb} and by proper adjustment of the other circuit parameters. The exact relationship will be derived later. The *timing state*, from which self-triggering does occur, is obtained when V_1 is conducting and V_2 is cut off.

The circuit diagram and symbols to be used are given in figure (18.15).

Note that all voltages measured with respect to ground carry an extra subscript n. Voltages measured with respect to the common cathode do not have this subscript.

The equivalent circuit of the V_1 amplifier in the *normal* state is shown in figure (18.16). If the circuit has been in this state for some time, transient effects can be neglected. It is then possible to compute the relationships between the circuit constants required to *prevent* self-triggering from this mode.

If the transients are complete, no current flows through C_2 or R_{L_1} in the equivalent circuit of figure (18.16). Thus the only voltage loop

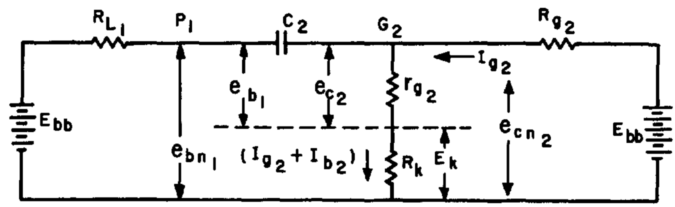


Fig. 18.16. Monostable cathode coupled multivibrator: equivalent plate circuit of the V_1 amplifier in the normal state (V_1 off and V_2 on).

equation required is $E_{bb} = I_{g_2}(R_{g_2} + r_{g_2}) + E_k$. Thus the grid current of V_2 is

$$I_{g_2} \doteq \frac{E_{bb} - E_k}{R_{g_2} + r_{g_2}}$$

In all practical cases the gridleak resistor R_{g_2} is much larger than the equivalent grid resistance r_{g_2}. Hence the grid current is approximately $I_{g_2} \doteq (E_{bb} - E_k)/R_{g_2}$. The grid current is now known, so that the grid voltage on V_2 is

$$e_{c_2} = I_{g_2} r_{g_2} \doteq (E_{bb} - E_k)\frac{r_{g_2}}{R_{g_2}}$$

Because R_{g_2} is much larger than r_{g_2}, the grid-to-cathode voltage e_{c_2} is nearly zero, being positive by only a few tenths of a volt. For all practical purposes we can assume that e_{c_2} is zero.

The grid voltage of V_2 is now known, so that we can easily find the operating point of the tube by drawing the load line for $(R_L + R_k)$ and E_{bb} on the static characteristics. The quiescent plate current I_{b_2} is thereby determined.

All the currents flowing in the equivalent circuit of figure (18.16) are now known. Hence the various network voltages are easily computed to be

$$e_{bn_2}(0^-) = E_{bb} - I_{b_2}R_{L_2} \tag{18.10}$$

$$E_k = e_k(0^-) = (I_{b_2} + I_{g_2})R_k \tag{18.11}$$

$$e_{b_2}(0^-) = E_{bb} - I_{b_2}(R_{L_2} + R_k) - I_{g_2}R_k \tag{18.12}$$

$$e_{c_1}(0^-) = E_{cc} - (I_{b_2} + I_{g_2})R_k \tag{18.13}$$

$$e_{cn_2}(0^-) = e_{c_2}(0^-) + e_k(0^-) \doteq I_{b_2}R_k \tag{18.14}$$

The entire evaluation has been made assuming that V_1 is cut off. If this condition is to actually exist, then from equation (18.13),

$$[(I_{b_2} + I_{g_2})R_k - E_{cc}] > E_{co} \tag{18.15}$$

The circuit must be designed to maintain this inequality, otherwise the circuit will become astable.

Because V_1 is cut off, we now know that

$$e_{bn_1}(0^-) = E_{bb} \tag{18.16}$$

The voltage across the timing capacitor C_2 is

$$\gamma_{ch} = e_{bn_1}(0^-) - e_{cn_2}(0^-) \doteq E_{bb} - I_{b_2}R_k \tag{18.17}$$

if I_{g_2} is neglected in comparison to I_{b_2}.

When the triggering impulse is applied at $t = 0$, because of regenerative switching, V_1 conducts and V_2 cuts off. The circuit is then in the timing state, and the analysis of this interval is given in the next section.

18.8. Monostable Cathode Coupled Multivibrator, Timing State

The transition from the normal state to the timing state requires a very short interval of time. Thus the moment the transition is concluded is designated as $t = 0^+$.

The first equivalent plate circuit for the V_1 amplifier in the timing state is shown in figure (18.17a). It has been implicitly assumed that the grid of V_1 is negative because experience indicates that this is a desirable situation and the circuit should be designed to achieve it.

The equivalent circuit can be modified somewhat because

$$e_{c_1} = e_{cn_1} - e_k = E_{cc} - i_{b_1}R_k$$

Hence the μe_{c_1} generator becomes

$$\mu e_{c_1} = \mu E_{cc} - i_{b_1}\mu R_k$$

Therefore, using this relationship, the circuit of figure (18.17b) is obtained. Also, because e_{c_1} is assumed to be negative, the foregoing equation specifies the necessary condition. That is $(E_{cc} - i_{b_1}R_k) < 0$, or

$$i_{b_1}R_k > E_{cc} \tag{18.18}$$

This is not a useful relationship yet because i_{b_1} is unknown.

Finally, the circuit computations are simplified if the Thevenin equivalent of figure (18.17b) is evaluated at the marked terminals. The

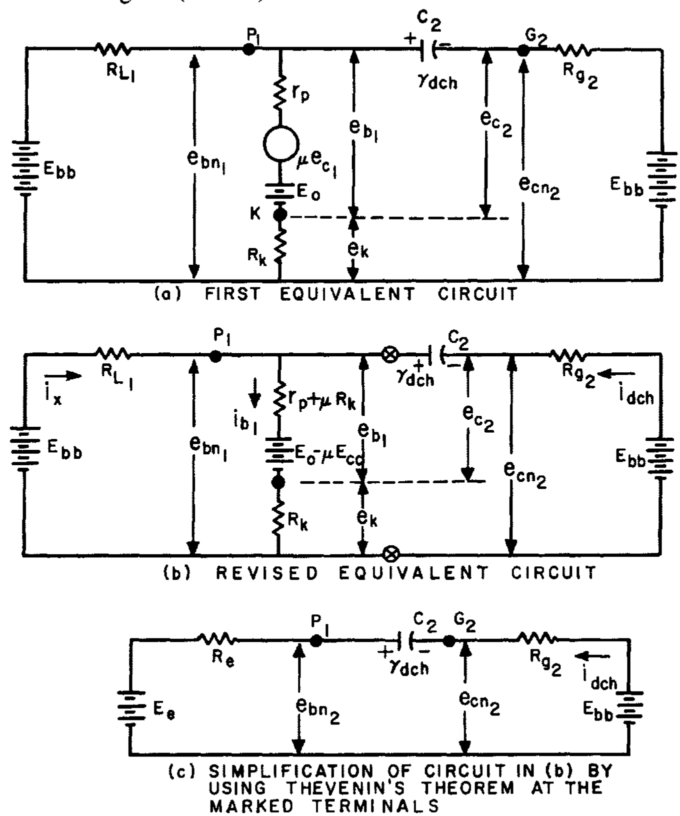


Fig. 18.17. Equivalent circuits of the V_1 amplifier of the monostable cathode coupled multivibrator in the timing state.

result appears in figure (18.17c) where the equivalent generator and internal impedance are given as

$$E_e = E_{bb} - \left(\frac{E_{bb} + \mu E_{cc} - E_0}{r_p + R_{L_1} + R_k(\mu + 1)}\right) R_L \tag{18.19}$$

$$R_e = \frac{R_{L_1}[r_p + R_k(\mu + 1)]}{r_p + R_{L_1} + R_k(\mu + 1)} \tag{18.20}$$

For triodes it is valid to assume that $(E_{bb} + \mu E_{cc})$ is much greater than the intercept voltage E_0. Hence E_0 can be neglected in equation (18.19). A useful form of equation (18.19) results if the following algebraic manipulations are performed on the bracketed part of the equation:

(1) Divide the numerator and denominator by $(r_p + R_{L_1})$.

(2) Multiply and divide by r_p.

(3) Replace r_p with μ/g_m.

Equation (18.19) then assumes the following form:

$$E_e = E_{bb} - \left[\frac{E_{bb}}{\mu} + E_{cc}\right]\left[\frac{g_m}{1 + R_k(\mu+1)/(R_{L_1}+r_p)}\right][r_p R_{L_1}/(r_p + R_{L_1})] \quad (18.21)$$

The second bracketed factor is easily identified as

$$g_m' = \frac{g_m}{1 + R_k(\mu + 1)/(r_p + R_{L_1})}$$

which is effective transconductance of a grounded cathode amplifier with cathode degeneration. Furthermore, if R_{g_2} is very much larger than r_p or R_L, which is generally true when triodes are used, the last bracketed factor in equation (18.21) is $R = r_p R_{L_1}/(r_p + R_{L_1})$, the mutual impedance of the amplifier circuit in the reference case. Consequently, the equivalent Thevenin generator can be expressed as

$$E_e = E_{bb} - \left(\frac{E_{bb}}{\mu} + E_{cc}\right) g_m' R \quad (18.22)$$

Of course, $A_r' = g_m' R =$ reference gain of the degenerative amplifier. Hence

$$E_e = E_{bb} - \left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \quad (18.23)$$

The total voltage applied to the equivalent circuit for the timing capacitor is

$$E_A = E_{bb} - E_e = \left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \quad (18.24)$$

The initial capacitor voltage is $\gamma_{dch} = E_{bb} - E_k$. Therefore the transient discharge current is

$$i_{dch} = \frac{E_A + \gamma_{dch}}{R_e + R_{g_2}} \varepsilon^{-t/(R_e + R_{g_2})C_2}$$

However, $\omega_1 = 1/[(R_e + R_{g_2})C_2] =$ lower cutoff frequency of the nondegenerative amplifier. It is also generally true that R_{g_2} is much larger than R_e. Hence substituting these relationships and the equations for E_A and γ_{dch} yields

$$i_{dch} = \frac{(E_{bb}/\mu + E_{cc})A_r' + E_{bb} - E_k}{R_{g_2}} \varepsilon^{-\omega_1 t}$$

The value of the current immediately after the trigger pulse is

$$i_{dch}(0^+) = \frac{(E_{bb}/\mu + E_{cc})A_r' + E_{bb} - E_k}{R_{g_2}}$$

The plate-to-ground voltage of V_1 at this same instant is

$$e_{bn_1}(0^+) = E_e + i_{dch}(0^+)R_e$$
$$= E_{bb}\left(1 + \frac{R_e}{R_{g_2}} - \frac{E_k}{E_{bb}} \cdot \frac{R_e}{R_{g_2}}\right) - \left(\frac{E_{bb}}{\mu} + E_{cc}\right)\left(A_r' - \frac{R_e}{A_{g_2}}\right)$$

However, it is known that E_k is less than E_{bb} and that R_e is much less than R_{g_2}. Hence it is possible to show that

$$e_{bn_1}(0^+) \doteq E_{bb} - \left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \tag{18.25}$$

The grid-to-cathode voltage on V_2 can be computed in the same way. That is,

$$e_{cn_2}(0^+) = E_{bb} - i_{dch}R_{g_2}$$

or

$$e_{cn_2}(0^+) \doteq E_k - \left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \tag{18.26}$$

Because V_2 is cut off, the plate-to-ground voltage is

$$e_{bn_2}(0^+) = E_{bb} \tag{18.27}$$

The plate-to-ground voltage of V_1 can be computed from figure (18.17b) to be $e_{bn_1}(0^+) = E_{bb} - i_x R_{L_1}$. Hence the current i_x is $i_x = (E_{bb} - e_{bn_1})/R_{L_1}$. Hence, immediately after the trigger,

$$i_x(0^+) = \frac{E_{bb} - e_{bn_1}(0^+)}{R_{L_1}}$$

or, substituting equation (18.25)

$$i_x(0^+) = \frac{(E_{bb}/\mu + E_{cc})A_r'}{R_{L_1}} \tag{18.28}$$

The remaining circuit voltages are now easily found to be

$$e_k(0^+) = i_x(0^+)R_k \quad \text{because} \quad i_x \gg i_{dch} \tag{18.29}$$

$$e_{c_1}(0^+) = E_{cc} - e_k(0^+) \tag{18.30}$$

$$e_{c_2}(0^+) = e_{cn_2}(0^+) - e_k(0^+) \tag{18.31}$$

Thus all significant points on the waveforms of the various circuit voltages can be computed from equations (18.25) through (18.31).

Following the trigger, the two voltages e_k and e_{bn_2} reach their final values at $t = 0^+$. This occurs because the current i_x producing the voltages is unaffected by the timing capacitor and is nontransient. Thus the plate-to-cathode voltage of V_2, the nonconducting tube, is constant through the timing state at a value

$$e_{b_2} = e_{bn_2}(0^+) - e_k(0^+)$$

Knowing this voltage, the corresponding grid cutoff voltage E_{co_2} is easily found from the plate characteristics of the tube. When this value is equalled by the grid voltage e_{c_2}, tube V_2 conducts and the timing interval ends as a result of self-triggering. Thus the pulse duration is controlled by the time required for e_{c_2} to change from the value at $t = 0^+$ to E_{co_2}. Designating the pulse duration as T, and the moment just prior to self-triggering as t_1^-, then $e_{c_2}(t_1^-) = E_{co_2}$. Also, from the equivalent circuit, $e_{c_2} = e_{cn_2} = e_k$. Hence substituting the equations for e_{cn_2} and e_k leads to

$$e_{c_2} = E_{bb} - \left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \left(\frac{R_k}{R_{L_1}} + \varepsilon^{-\omega_1 t}\right) - (E_{bb} - E_k)\, \varepsilon^{-\omega_1}$$

When $t = T$, then $e_{c_2} = E_{co_2}$. Hence this leads to

$$\varepsilon^{-\omega_1 T} = \frac{(E_{bb} - E_{co_2}) - \dfrac{R_k}{R_{L_1}}\left(\dfrac{E_{bb}}{\mu} + E_{cc}\right) A_r'}{\left(\dfrac{E_{bb}}{\mu} + E_{cc}\right) A_r' + (E_{bb} - E_k)} \tag{18.32}$$

and the pulse duration works out to be

$$T = \frac{1}{\omega_1} \ln \left[\frac{(E_{bb}/\mu + E_{cc})A_r' + (E_{bb} - E_k)}{(E_{bb} - E_{co_2}) - \dfrac{R_k}{R_{L_1}}(E_{bb}/\mu + E_{cc})A_r'}\right] \tag{18.33}$$

This is a useful design equation because most of the factors involved are specified by other considerations.

One important point remains to be cleared up. This analysis was based upon the assumption that the grid of V_1 was slightly negative during the timing state. For this to be true it is necessary that

$$i_{b_1} R_k \geq E_{cc} \quad \text{or} \quad e_k(0^+) \geq E_{cc}$$

Substituting for $e_k(0^+)$ yields

$$\left(\frac{E_{bb}}{\mu} + E_{cc}\right) A_r' \frac{R_{L_1}}{R_k} \geq E_{cc} \tag{18.34}$$

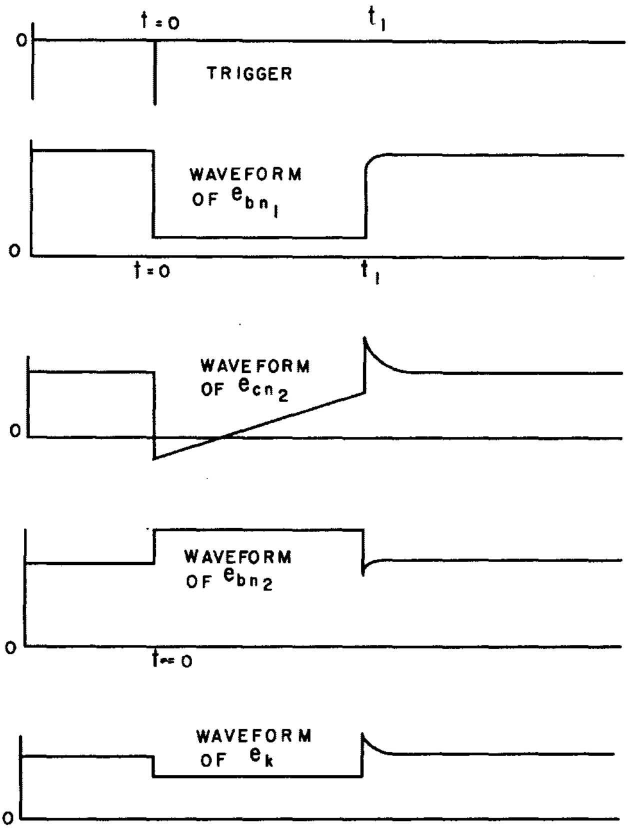


Fig. 18.18. Waveforms in a monostable cathode coupled multivibrator.

After some manipulation this condition can be written

$$\frac{R_{L_1}}{R_k} \geq \frac{\mu}{(E_{bb}/E_{cc} + \mu)A_r'} \tag{18.35}$$

This, together with the inequality specified in (18.15), are the key factors in assuring proper operation.

The remaining points on the circuit waveforms, which are shown in figure (18.18), can be computed by the same method as that followed here, simply using the proper equivalent circuits.

One of the main advantages of the cathode coupled multivibrator is the fact that the pulse duration can be made to be almost linearly proportional to the grid bias. If this is true, the second derivative of T with respect to E_{cc} should be zero. If this is evaluated, the following inequality results:

$$E_{cc} = -\frac{E_{bb} - E_{co_2}}{2A_r'}\frac{R_{L_1}}{R_k} - \frac{E_{bb}}{\mu} - \frac{E_{bb} - E_k}{2A_r'} \tag{18.36}$$

This gives the design center value for E_{cc}.

It should be clear from this analysis that the proper design of a cathode coupled multivibrator requires real technical skill. No amount of blind juggling will substitute for a careful preliminary analytical treatment.

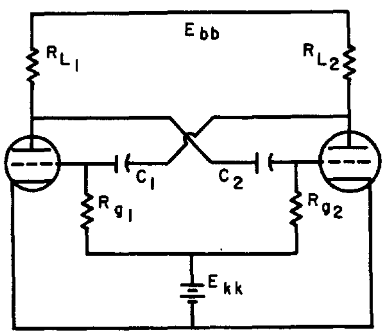


Fig. 18.19. Astable plate coupled multivibrator.

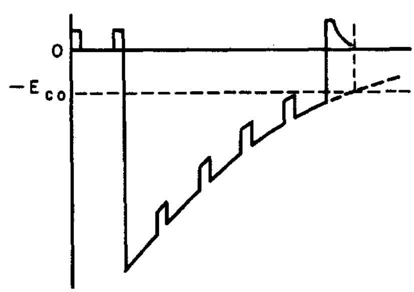


Fig. 18.20. Pulse synchronization of an astable multivibrator.

18.9. Astable Multivibrators

The monostable plate coupled multivibrator shown in figure (18.11) can be made astable, or freely oscillating, by removing the one remaining d-c coupling and replacing it with a-c coupling. The resulting astable plate coupled multivibrator is shown in figure (18.19) where both gridleak resistors are returned to a common arbitrary voltage E_{kk}.

The analysis of this circuit proceeds in exactly the same manner as that outlined in section (18.6) for the monostable circuit except that this circuit has *two* timing states from which self-triggering is possible.

As a result, it should be obvious that the total period of the multivibrator can be expressed as

$$T = T_1 + T_2$$
$$= \frac{1}{(\omega_1)_1} \ln \left(\frac{1 + A_{r_1}E_{bb}/\mu_1 E_{kk}}{1 + E_{co_2}/E_{kk}}\right) + \frac{1}{(\omega_1)_2} \ln \left(\frac{1 + A_{r_2}E_{bb}/\mu_2 E_{kk}}{1 + E_{co_1}/E_{kk}}\right)$$

Typical waveforms in the circuit are computed by the same method as that followed for the monostable case.

Astable multivibrators are often synchronized with some reference frequency signal because the frequency stability of multivibrators is rather poor. In all such cases, synchronization is accomplished by making the tubes conduct a little sooner than they would have in the absence of the synchronizing signal. The mechanism is clear in figure (18.20).

18.10. Transition and Recovery Times

The preceding rather detailed presentation of multivibrators was primarily concerned with the analysis of the normal and timing states. The transition time between states was assumed to be of infinitesimal duration. Actually, a finite time is required and the evaluation of the factors affecting it has been the subject of a number of searching studies.[2] The analysis is a difficult undertaking and there is a little disagreement concerning the results. However, it is generally conceded that the figure of merit of *RC* amplifiers g_m/C_T is the parameter of the greatest importance in determining the transition time. Thus high figure of merit tubes and video amplifier practice, including shunt peaking and related ideas, should be used when short transition times are desired. The reader is referred to the basic papers[2] for further details.

It will be recalled from the analysis of the timing states in the various monostable multivibrators that the duration of the output pulses was governed by the transient discharging of the timing capacitor. When the timing state concluded, transition to the other state occurred, and the capacitor underwent transient recharging to its original value. This recharging time is called the *recovery time*. It sets a fundamental

[2] Chance, *et al.*, *Waveforms*, vol. 19, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1949, chap. 5 by D. Sayre, pp. 174–177; M. V. Kiebert and A. F. Inglis, "Multivibrator Circuits," *Proc. IRE*, vol. 33, August, 1945, pp. 534–539; E. M. Williams, D. F. Aldrich, J. B. Woodford Jr., "Speed of Electronic Switching Circuits," *Proc. IRE*, vol. 38, January, 1950, pp. 65–69.

limit upon the time between trigger pulses because the *same* kind of pulse will be generated after every trigger only if the circuit is given sufficient time to recover its initial conditions. High speed recovery is often an important design consideration.

The recovery time is determined principally by (1) the clamping action introduced by the grid-cathode part of the tube; (2) the voltage to which the grid is returned; (3) the charging time constant $= (r_{g_2} + R_{L_1})C_2$. Thus the simplest ways to reduce the recovery time are to return the timing resistor to E_{bb} and to connect a low resistance diode directly across the grid-cathode terminals of the tube.

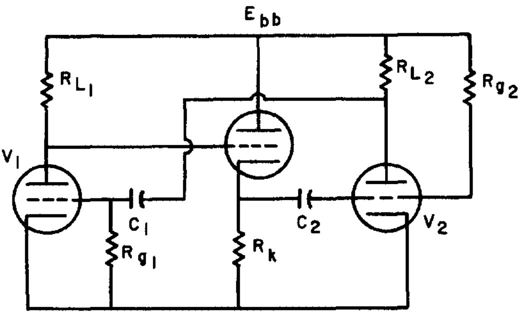


Fig. 18.21. Use of a cathode follower to reduce recharging time of the timing capacitor C_2.

The charging time constant can be reduced, but variations in some of these circuit constants will also affect the duration of the output pulse.

A final refinement can be made by inserting a cathode follower into the timing circuit as shown in figure (18.21). In this way the timing capacitor has nearly all the cathode follower current available for recharging. The recovery time is short.

18.11. Principles of the Phantastron

The *phantastron*, *sanatron*, and *sanaphant* are types of monostable trigger circuits. They differ from multivibrators mainly in the fact that a linear Miller integrated timing waveform is used rather than the exponential used in multivibrators. This is an important advantage because it increases the precision of the pulse duration.

The general factors controlling the operation of a phantastron can be explained from the circuit shown in figure (18.22). The circuit is connected as a conventional Miller integrator with three exceptions:

(1) The control grid is returned to E_{bb} through the gridleak resistor, rather than to ground.

(2) The suppressor is not grounded. Instead, it is biased negative by an amount sufficient to cut the plate current off. Plate current is allowed to flow for specified intervals by a positive gating pulse applied to the suppressor.

(3) The screen is not bypassed to ground.

The tube used in this type of circuit should have a sharp suppressor cutoff characteristic and a high suppressor-to-plate transconductance.

Under *normal* conditions with the plate current cut off by the

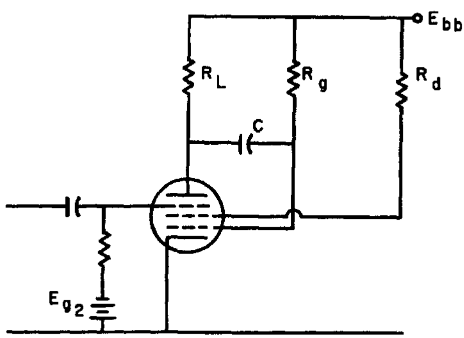


Fig. 18.22. Fore-runner of the phantastron.

suppressor, the plate voltage is E_{bb}. Because the gridleak is returned to E_{bb}, the grid is positive by a voltage

$$e_c = E_{bb}\frac{r_g}{r_g + R_g} \doteq E_{bb}\frac{r_g}{R_g}$$

This voltage is practically zero because R_g is much larger than r_g. The entire cathode current flows to the screen, so that the screen voltage is low. The voltage across C is almost equal to E_{bb} because it is the difference between the plate and grid voltages. These are the circuit conditions existing just prior to the incidence of a positive gate on the suppressor.

With the arrival of the suppressor gate, current flows in the plate circuit, causing the plate voltage to drop. Because of the action of the feedback capacitor C, the grid drops by an equal amount and thereby becomes negative. The drop in plate voltage is halted by the action of the grid. Once this equilibrium is established, conventional Miller integration in a class A amplifier occurs. During this initial change in operation the screen current is tremendously reduced because the tube is now operating as a conventional pentode. This causes a large positive pulse to appear at the screen.

The action of the Miller integrator is such that the plate voltage falls linearly as a function of time until it bottoms. The grid voltage rises only slightly during this period. When the plate voltage bottoms, the potential to ground of the plate side of the feedback condenser becomes fixed. Thus the grid voltage rises rapidly toward E_{bb} as C charges. It

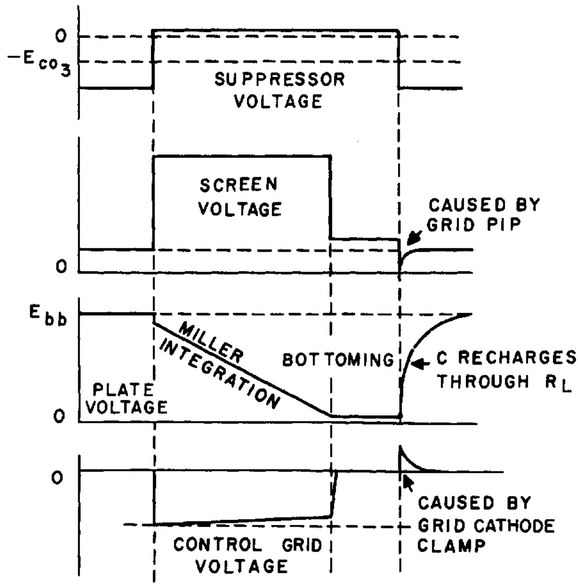


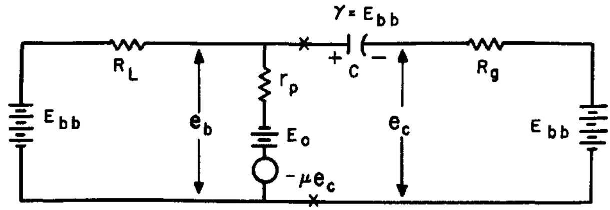
Fig. 18.23. Waveforms for the circuit of figure (18.22).

reaches zero volts in a short time and is clamped there by the diode action of the grid-cathode circuit. This sudden jump in grid voltage from some negative value to a zero value causes the cathode current to increase markedly. The additional current has to be taken by the screen because the plate is bottomed. Hence the screen voltage drops suddenly because of the sudden increase in screen current.

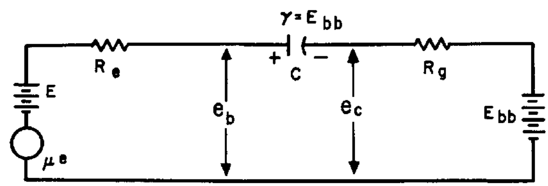
The circuit remains in this state until the end of the suppressor gate. When this occurs the plate current ceases and the plate voltage returns to E_{bb} as C recharges through R_L. This increase is coupled over to the grid so that a positive grid pip is generated, the magnitude depending upon the effectiveness of the grid-cathode clamp. The screen current is increased by the change in tube operation when plate current ceases. It will show a small negative pulse because of the positive grid pip. The various waveforms are shown in figure (18.23).

18.12. Phantastron Analysis

The equivalent circuit of the prototype phantastron, together with its Thevenin equivalent, are shown in figure (18.24) for the period when



(a) DIRECT EQUIVALENT PLATE CIRCUIT WHEN THE TUBE CONDUCTS; TIMING STATE



(b) THEVENIN EQUIVALENT OF (a) TAKEN AT THE MARKED TERMINALS WHERE

$$E = E_{bb} - \frac{E_{bb} + E_0}{\mu} g_m R_e$$

$$R_e = \frac{r_p R_L}{r_p + R_L} \qquad e = -\frac{e_c}{\mu} g_m R_e$$

Fig. 18.24. Phantastron equivalent circuits.

the tube is conducting plate current. Note that the equivalent generator E_0, which represents the intercept voltage, is reversed in polarity from the triode case.

The complete network differential equation for the equivalent circuit in (18.24b) is

$$E_{bb} = i(R_e + R_g) + \frac{1}{C}\int i\,dt + E + \mu e$$

However, it is shown in figure (18.24) that $\mu e = -g_m R_e e_c$. Also, from the equivalent circuit it is clear that $e_c = E_{bb} - iR_g$. Consequently,

$$\mu e = -g_m R_e(E_{bb} - iR_g) = g_m R_e R_g i - E_{bb} g_m R_e$$

Thus the system differential equation is

$$E_{bb} - E + E_{bb} g_m R_e = i(R_e + R_g) + g_m R_e R_g + \frac{1}{C}\int i\,dt$$

Now take the Laplace transform of the equation. Collect all d-c terms, including the initial capacitor voltage, together on the left side and then multiply through by s. Hence

$$\frac{E_{bb}(1 + g_m R_e) + (E_{bb} - E)}{R_e + R_g} = I\left[s\left(1 + g_m \frac{R_e R_g}{R_e + R_g}\right) + \frac{1}{(R_g + R_e)C}\right]$$

In the discussion of the resistance coupled amplifier in chapter 4 it was shown that

$$\begin{aligned} A_r &= \text{reference gain} = g_m R \\ &= g_m \frac{R_e R_g}{R_e + R_g} \\ \omega_1 &= \text{lower cutoff frequency} \\ &= \frac{1}{(R_e + R_g)C} \end{aligned} \tag{18.37}$$

Hence, after evaluating $(E_{bb} - E)$, the preceding equation is

$$\frac{E_{bb}}{R_e + R_g} + \frac{A_r}{R_g}\left(E_{bb}\frac{\mu + 1}{\mu} + \frac{E_0}{\mu}\right) = I[(1 + A_r)s + \omega_1]$$

This equation can be simplified by using a few approximations. In a pentode it is nearly always true that μ is much larger than 1. It is also generally true that E_0 will be much less than μE_{bb}. The use of these approximations reduces the preceding equation to

$$\frac{E_{bb}}{R_g} A_r \left(1 + \frac{1}{g_m R_e}\right) \doteq (1 + A_r) I \left(s + \frac{\omega_1}{1 + A_r}\right) \tag{18.38}$$

Hence the transform current is

$$I(s) = I = \frac{E_{bb}}{R_g} \cdot \frac{A_r}{1 + A_r} \cdot \frac{1 + A_x}{A_x} \cdot \frac{1}{s + \omega_1/(1 + A_r)}$$

where $A_x = g_m R_e$. Therefore the transient current, or inverse transform is

$$i(t) = \frac{E_{bb}}{R_g} \cdot \frac{A_r}{1 + A_r} \cdot \frac{1 + A_x}{A_x}\, \varepsilon^{-\omega_1 t/(1 + A_r)} \tag{18.39}$$

The plate voltage during the conducting period is $e_b = E + iR_e + \mu e$

or
$$e_b = E + iR_e + g_m R_e R_g i - E_{bb} g_m R_e$$

In the preceding notation this becomes

$$e_b = E - E_{bb}A_x + i(R_e + R_g A_x)$$

Substitution for E and the use of the previous approximations reduces this equation to

$$e_b \doteq E_{bb}(1 - A_x) + i(R_e + R_g A_x)$$

In nearly all cases, $R_g A_x$ is much greater than R_e. Now substitute equation (17.39) for i, and the approximate equation for e_b is

$$e_b \doteq E_{bb}(1 - A_x) + E_{bb}\frac{1 + A_x}{1 + A_r} A_r \varepsilon^{-\omega_1 t/(1+A_r)} \tag{18.40}$$

As you can see, the plate voltage falls exponentially with a time constant $(1 + A_r)/\omega_1$, toward a final value $E_{bb}(1 - A_x)$, which is extremely negative. Thus the fall occurs over a wide voltage range and with a long time constant. The actual fall that takes place will be only a small fraction of the total exponential and may be approximated by a straight line. The departure from linearity will seldom exceed 1%.

The initial slope of the plate voltage is

$$\left.\frac{de_b}{dt}\right|_{t=o} = -E_{bb}\omega_1 \frac{A_r}{1 + A_r} \cdot \frac{1 + A_x}{1 + A_r}$$

The circuit is usually designed so that A_x and A_r are both much greater than 1. Hence

$$\text{slope} \doteq -E_{bb}\omega_1 \frac{A_x}{A_r} = -E_{bb}\omega_1 \frac{R_e}{R}$$

Now substitute for R, R_e, and ω_1. After canceling common terms,

$$\text{slope} \doteq -\frac{E_{bb}}{R_g C} \tag{18.41}$$

The time required for the plate voltage to fall from nearly E_{bb} to the bottoming voltage E_{bot} is obviously

$$T = \frac{E_{bb} - E_{bot}}{\text{slope}} = R_g C \left(1 - \frac{E_{bot}}{E_{bb}}\right) \tag{18.42}$$

The bottoming voltage is only a few volts, while E_{bb} is usually several hundred volts. Hence if the fraction in equation (18.42) can be neglected in comparison with 1, then

$$T = \text{pulse duration} \doteq R_g C \tag{18.43}$$

18.13. Phantastron Circuits

In the discussion of section (18.11) it was shown that a positive suppressor gate was required to actuate the tube and cause the linear rundown in plate voltage until the tube bottomed. It was also found that a large positive pulse was generated in the screen circuit as a result. The leading edge of the screen pulse coincides with the moment of plate current flow, while the trailing edge is coincident with bottoming. Hence the circuit can be made monostable if this screen pulse is used to supply the suppressor gate. When this is done the circuit becomes the *screen-coupled phantastron* shown in figure (18.25).

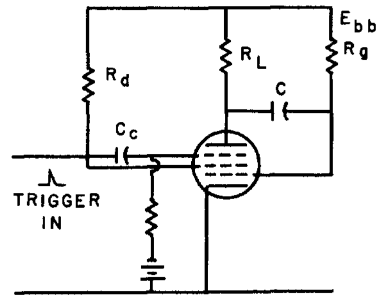


Fig. 18.25. Screen coupled phantastron

Circuit operation is initiated by a positive trigger supplied to the suppressor through a coupling diode. The positive rectangular output pulse is taken from the screen and has a duration accurately defined by

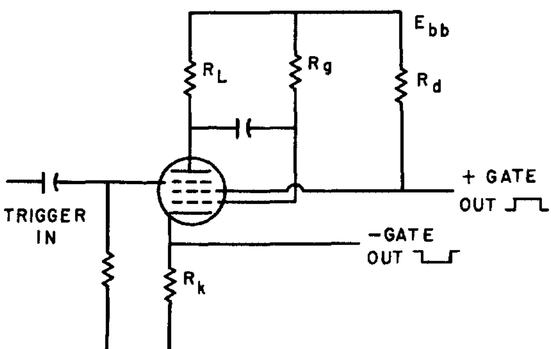


Fig. 18.26. Cathode coupled phantastron.

the time required for the plate voltage to fall to the bottoming voltage.

The screen-coupled phantastron has two main disadvantages: A negative supply voltage is required for the suppressor bias and the positive output pulse from the screen is loaded by the suppressor circuit.

Both of these shortcomings are corrected in the *cathode-coupled phantastron* shown in figure (18.26). Here the negative pulse generated in an unbypassed cathode resistor is used. The suppressor is grounded so that the negative cathode-to-ground pulse appears as a positive pulse from suppressor-to-cathode. The necessary regenerative feedback is thereby obtained. The main disadvantage of this circuit is the gain loss caused by cathode degeneration. The pulse duration can be computed by the method outlined in section (18.12).

18.14. Blocking Oscillator Principles

It was shown in the preceding section that a multivibrator was essentially a two-stage resistance coupled amplifier with 100% feedback. If the second tube is replaced by a phase inverting transformer, the circuit becomes a *blocking oscillator.* A simplified circuit diagram

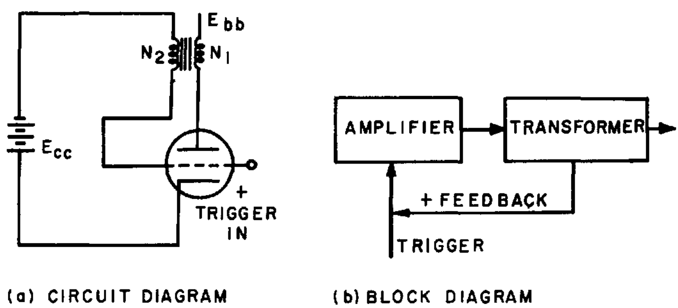


Fig. 18.27. Simplified monostable blocking oscillator.

and block diagram are shown in figure (18.27). A blocking oscillator can be astable or monostable, but never bistable.

This circuit has not been accurately formulated. The difficulty arises mainly from the nonlinear characteristics of the pulse transformer and the lack of sufficient information regarding the positive grid characteristics of tubes. The problem is especially acute with receiving type tubes because they are not susceptible to simplifying approximations that are valid for high power tubes.[3]

The discussion of blocking oscillators is included here as an illustration of a case in which more refined methods of analysis are required. Only a brief qualitative explanation of the operation will be given.

[3] For a general discussion of the component characteristics required for a blocking oscillator, see Chance, *et al.*, *Waveforms*, vol. 19, Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, 1949, chap. 6, pp. 205–253; For an approximate computation of the transient response of a pulse transformer, see C. K. Hadlock and D. Lebell, "Some Studies of Pulse Transformer Equivalent Circuits," *Proc. IRE*, vol. 39, January, 1951, pp. 81–83.

Under normal conditions the tube is biased off by the grid supply voltage. If a short positive trigger pulse is applied to the grid to make the tube conduct, the feedback loop closes, the loop gain is greater than 1, and the regenerative switching occurs.

When the trigger causes the tube to conduct, the sudden change in plate current, though small in magnitude, is opposed by the transformer inductance. Thus nearly all the supply voltage E_{bb} is developed across the transformer, and the plate voltage drops sharply to a small value. This drop in plate voltage is coupled through the transformer and inverted so that it appears as a sudden increase in grid voltage. The increase is sufficient to drive the grid into the positive grid region. Thus the grid-cathode circuit is closed through a small resistance r_g.

Current flows in the grid circuit and this requires power that must be supplied by the plate circuit. Thus the plate load resistance is approximately $r_g(N_1/N_g)^2$ plus the transformer resistance. As a result, the a-c load on the tube changes from practically infinity to virtually zero in a brief instant.

All these events occur so suddenly that the plate current i_b changes only slightly from zero. However, after these initial changes, the plate current commences a relatively steady rise through the transformer mutual inductance and the tube. The tube r_p and r_g are both quite small, so that the circuit time constant and the current buildup are both long. This causes the current increase to be approximately linear so that e_b, e_c, and di_b/dt are practically constant. Actually, deviations from linearity always exist, and this causes e_b to increase slightly, while e_c decreases somewhat. This is shown in the waveforms of figure (18.28).

The characteristics of the tube are nonlinear at high plate currents. Thus the rate of change of plate current decreases and eventually becomes zero. This effectively removes the positive grid signal and the grid voltage drops suddenly toward $-E_{cc}$. It attempts to decrease the plate current in the process, but it is opposed by the action of the transformer, which now reverses its polarity and acts as a current source. The plate is thereby driven highly positive, well beyond E_{bb}, in an effort to sustain plate current flow by preventing the grid voltage from reaching the cutoff point. However, this very increase in plate voltage is coupled into the grid circuit, after being inverted, and further reduces the grid voltage. As a result, the circuit regeneratively switches the tube off. The circuit then awaits another trigger. The circuit waveforms are shown in figure (18.28).

Designate the switching off value of plate current as I_s. Assume that the plate current increases linearly during the timing interval. Then the pulse duration is $T \doteq I_s di_b/dt$. The rate of change of current can be crudely approximated, because the transformer voltage drop is $e_T = L_m di_b/dt$, so that $di_b/dt = e_T/L_m$. This drop is nearly constant during the timing state, so that the average value is approximately $e_T \doteq E_{bb} - E_{bot}$, where $E_{bot} = e_b\,(\text{min}) =$ bottoming voltage. Hence, the pulse duration is approximately

$$T \doteq \frac{L_m I_s}{E_{bb} - E_{bot}} \tag{18.44}$$

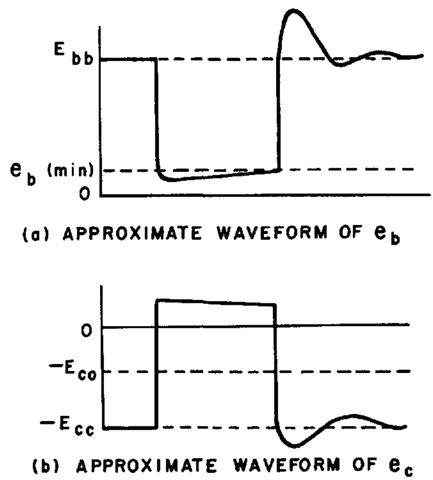


Fig. 18.28. Approximate waveforms in a blocking oscillator.

This equation should not be taken literally. It mainly serves to illustrate the dependence of the pulse duration upon the properties of the transformer and the bottoming or saturation characteristics of the vacuum tube.

The oscillation shown in the waveforms of figure (18.28) is caused by ringing in the resonant circuit formed by the shunt capacitance and transformer inductance.

Astable operation can be obtained by inserting a gridleak bias circuit of the desired time constant into the circuit in place of the bias supply.

18.15. Negative Resistance in Transistor Circuits

Class A transistor amplifiers were analyzed in chapter 8, and it was shown that the input and output resistances could be negative under certain conditions for all three of the basic amplifier configurations.

While this analysis is valid only for the variational characteristics, similar results are obtained for the d-c characteristics. This will be proved here.

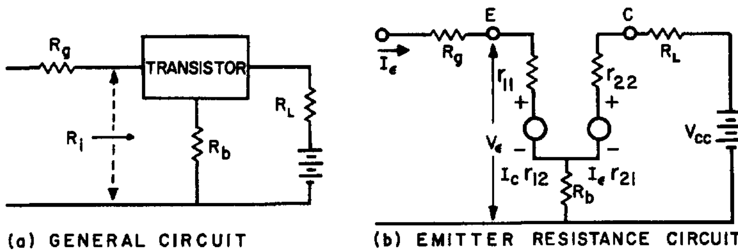


Fig. 18.29. Circuit diagrams for computing R_1.

The input resistance of the general circuit shown in figure (18.29) is

$$R_i = R_{11} - \frac{R_{12}R_{21}}{R_{22} + R_L} = \frac{V_i}{I_i} \tag{18.45}$$

The specialized circuit when the transistor is connected and the equiva-

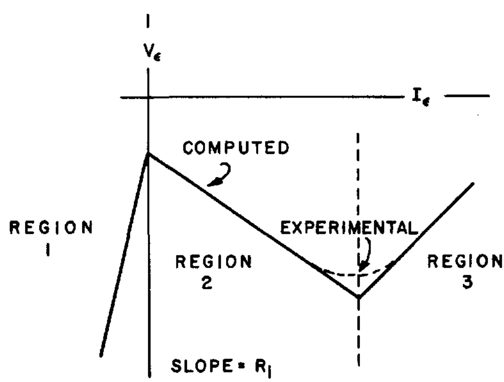
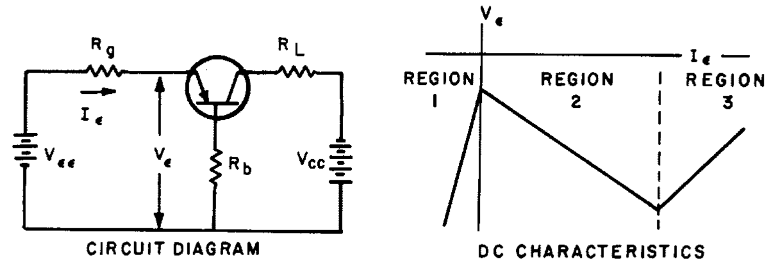


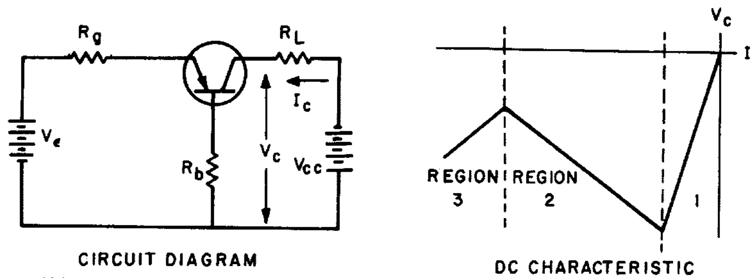
Fig. 18.30. Negative resistance in the emitter circuit.

lent circuit included is shown in figure (18.29b). It is understood that the transistor slope resistances can have any one of three or four values, depending upon the region of operation. Reference should be made, in this connection, to the discussion of the transistor equivalent circuit in chapter 1.

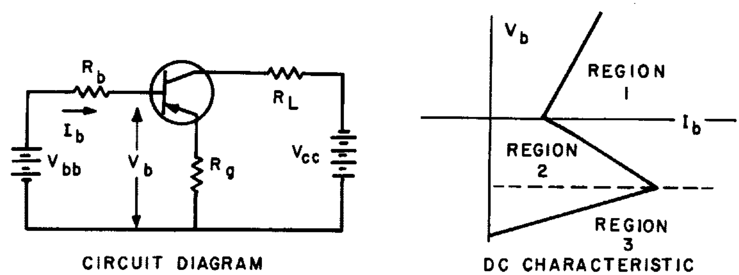
For the circuit of figure (18.29b), $R_{11} = r_{11} + R_b$; $R_{22} = r_{22} + R_b$; $R_{12} = r_{12} + R_b$; $R_{21} = r_{21} + R_b$. Hence the input resistance of the circuit at the emitter terminals is, after some rearrangement of terms,



(a) NEGATIVE RESISTANCE IN THE EMITTER CHARACTERISTIC



(b) NEGATIVE RESISTANCE IN THE COLLECTOR CIRCUIT



(c) NEGATIVE RESISTANCE IN THE BASE CIRCUIT

Fig. 18.31. Transistor negative resistance circuits. See A. W. Lo, "Transistor Trigger Circuits," pp. 1531–1541; A. E. Anderson, "Transistors in Switching Circuits," pp. 1541–1558, *Proc. IRE*, vol. 40, November, 1952.

$$R_i = R_b \left[1 + \frac{r_{11}}{R_b} - \frac{(1 + r_{12}/R_b)(1 + r_{21}/R_b)}{1 + R_L/R_b + r_{22}/R_b} \right] \tag{18.46}$$

Because there are at least three possible values for each of the transistor parameters, there are also three values for R_i. An estimate of the possible variation in R_i can be made by tabulating relative values for the transistor slope resistances in the three main regions as follows:

Region	r_{11}	r_{22}	r_{12}	r_{21}	*Nature of* R_i
1	high	high	low	very low	+ and large
2	low	high	low	high	− and large
3	lower	very low	low	zero	+ and small

The values of R_i computed from equation (18.46) are actually the slopes of the V_ε vs. I_ε characteristic in the three regions of operation. If such computations are made and if the characteristic is experimentally measured, the results appear approximately as shown in figure (18.30). This curve has the general form of an open circuit stable negative resistance. Such characteristics were discussed in a general way in chapter 10. The mechanism and methods of triggering can be figured by the same technique as that given in chapter 10.

It is important to remember that the discussion here and in chapter 10 neglected all reactive components that might appear in the equivalent circuit of the transistor.

Similar negative resistance characteristics can be obtained from the collector and base circuits. These circuits, together with their computed characteristics, are shown in figure (18.31).

18.16. Transistor Trigger Circuits[4]

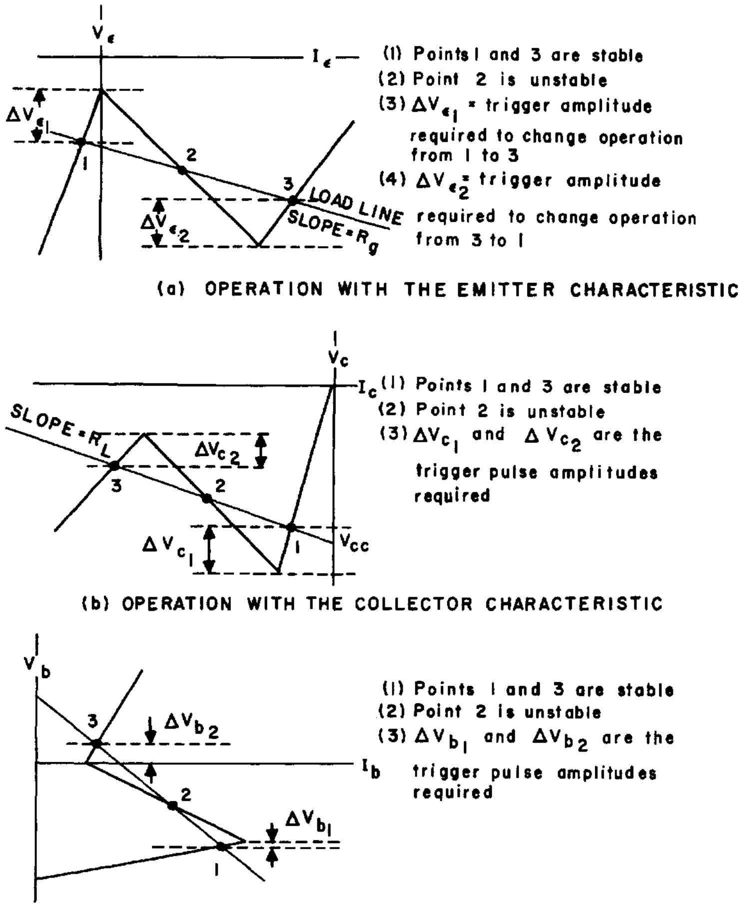
Transistor trigger circuits, like any other circuit involving open circuit stable negative resistance characteristics, can be monostable, bistable, or astable.

For bistable operation it is necessary only to connect a power supply and series load resistance across the negative resistance terminals. This voltage and resistance are then adjusted so that the load line intersects the circuit characteristic at three points. The necessary conditions are summarized for the three basic transistor negative resistances in figure (18.32). Triggering from one stable state to another is accomplished by voltage trigger pulses of the amplitude indicated. The methods of trigger pulse insertion are the same as for vacuum tube trigger circuits.

Monostable operation is achieved by making the circuit self-triggering from one of its stable states. For example, the insertion of a capacitor in parallel with either the emitter or collector power supply circuit will

[4] See the *Proc. IRE*, vol. 40, November, 1952, for the following articles: (a) G. E. McDuffie Jr., "Pulse Duration and Repetition Rate of a Transistor Multivibrator," pp. 1487–1489; (b) B. G. Farley, "Dynamics of Transistor Negative Resistance Circuits," pp. 1497–1508; (c) A. W. Lo, "Transistor Trigger Circuits," pp. 1531–1541; (d) A. Eugene Anderson, "Transistors in Switching Circuits," pp. 1541–1558.

make the circuits monostable if the supply voltages and load resistances are adjusted to the proper values. A series inductance can provide the same function in the base negative resistance circuit. The circuit



(c) OPERATION WITH THE BASE CHARACTERISTIC

Fig. 18.32. Bistable transistor trigger circuits. See figure (18.31) for circuit diagrams.

diagrams and operating conditions are summarized in figures (18.32) and (18.33).

The mechanism of self-triggering is relatively easy to understand from figure (18.33a). This shows the load line properly adjusted for monostable operation with the emitter negative resistance characteristic. Under normal conditions, the circuit is in its stable state, marked 1 on

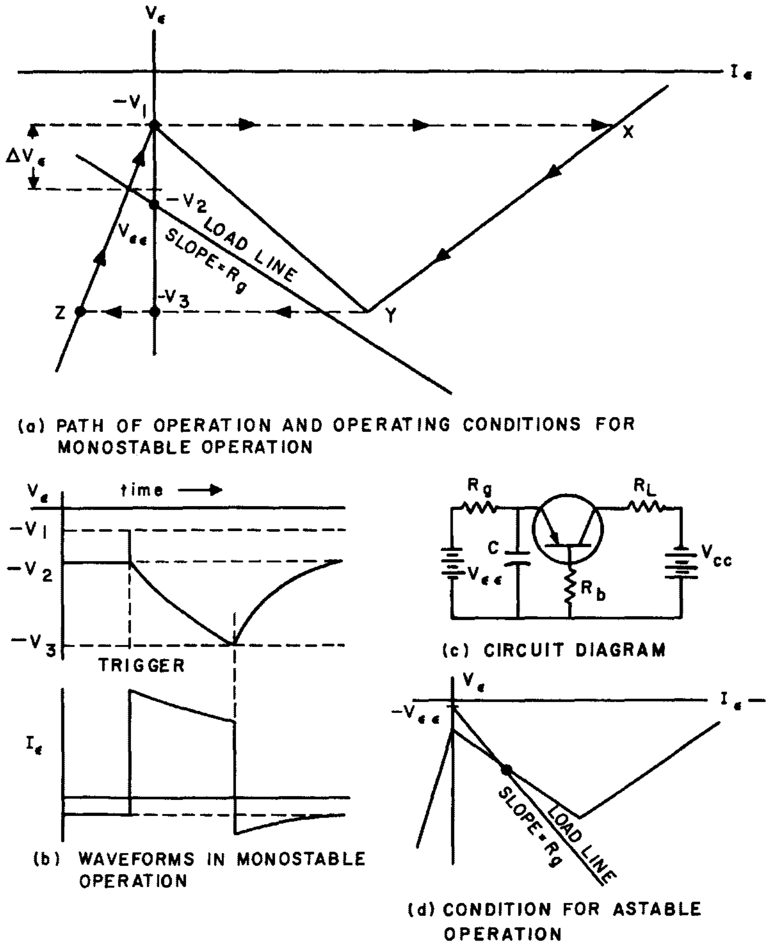


Fig. 18.33. Transistor trigger circuit using the emitter circuit negative resistance.

the figure. If a trigger pulse of amplitude ΔV_ε is supplied from a capacitive source (so that C will not bypass the trigger), operation will trigger over into another stable state. However, because the voltage across C cannot change instantaneously, the emitter voltage is held constant, and operation changes to some point such as X; the transition occurs along the dotted line as shown in figure (18.33a). At the

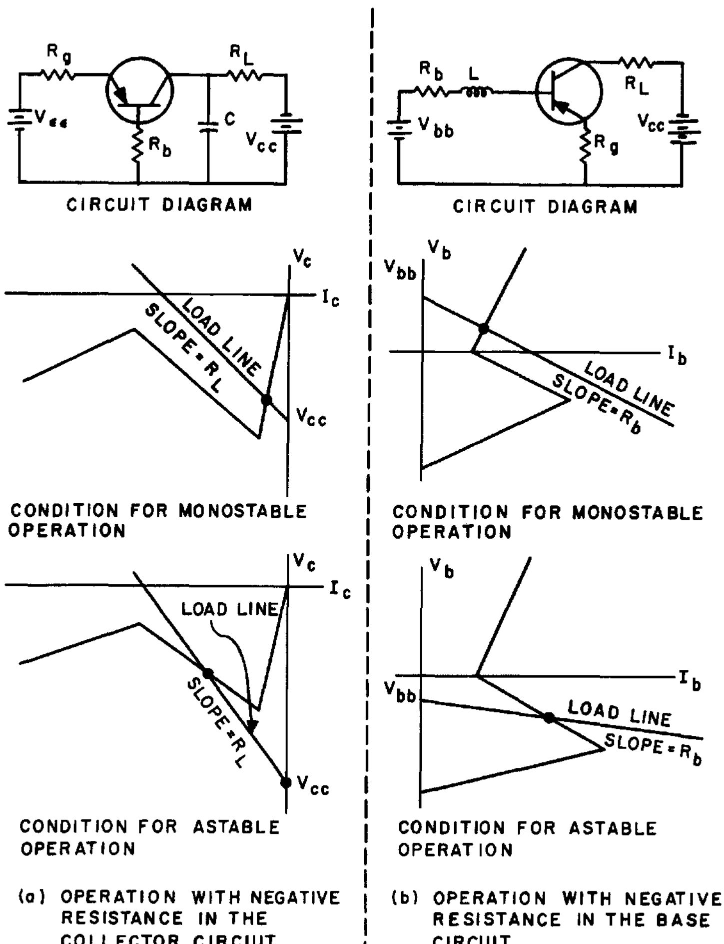


Fig. 18.34. Astable and monostable transistor trigger circuits.

same time, the emitter current increases from a small negative value to a large positive value. As the capacitor commences charging, the emitter voltage gradually increases and the operating point moves along the characteristic curve until point Y is attained. The circuit then self-triggers with the emitter voltage momentarily held constant by the capacitor. Thus the operating point moves horizontally to the point Z. As the capacitor discharges, operation gradually approaches point 1 and the original stable state results. The circuit then awaits another trigger. The waveforms of figure (18.33b) are an obvious result of this discussion.

The other two transistor negative resistance circuits can be explained in a similar way. In all three cases the computation of waveforms and pulse duration is a straightforward application of circuit analysis to the appropriate equivalent circuits. The general method of analysis has been illustrated several times for vacuum tube switching circuits, and the same techniques apply to transistor circuits.

It seems clear that astable operation can be achieved with the same circuits as those used for monostable operation simply by changing the supply voltage and series resistance so that the operating point falls in the negative resistance region. This is shown in figures (18.33d) and (18.34). Self-triggering from each stable state now occurs and the circuits operate like stable multivibrators.

18.17. Bistable Magnetic Amplifier[5]

The circuit diagram of one type of bistable magnetic amplifier is shown in figure (18.35). The two stable states occur when (1) neither core is saturated and there is no output; (2) both cores are saturated and there *is* an output. As in nearly all magnetic amplifiers, in the absence of a signal input, the supply voltage e'_{ac} is adjusted so that the cores almost, but not quite, saturate. The core flux, in the absence of a signal input, alternates between two extreme values:

(1) It is reset to ϕ_0 by the magnetizing current in the control winding.

(2) It is magnetized just barely short of saturation ϕ_s by the current in the load winding. Call this flux ϕ'_s.

Suppose that the instantaneous polarities of the various alternating voltages are as shown in figure (18.35). Assume that the flux in core no. 1 is ϕ'_s and that in core no. 2 is ϕ_0. During this half cycle the rectifiers act to prevent current flow in the control winding of core 2 and in the load winding of core no. 1. However, small magnetizing currents do flow in the other two windings, causing the flux in core 1 to be reset to ϕ_0 by e'_{ac} while the flux in core 2 is gated from ϕ_0 to ϕ'_s. However, there is no output because neither core is saturated.

[5] R. A. Ramey, "The Single Core Magnetic Amplifier as a Computer Element," *Communications and Electronics*, January, 1953, pp. 442–446.

On the next half cycle the reverse condition prevails because all alternating voltages and the marked polarities shown on the figure reverse. Thus as long as there is no signal input so that both cores are reset by e'_{ac}, core saturation never occurs and there is no output across Z_1. This is stable state 1.

Now suppose that a signal pulse is applied to the *a* input during a half cycle such as that indicated by the polarity markings on figure (18.35). If this pulse has sufficient amplitude it will prevent core no. 1 from being reset to ϕ_0. Instead, it will be demagnetized only to some point ϕ_x. Thus, on the *next* half cycle, core no. 1 will saturate at some

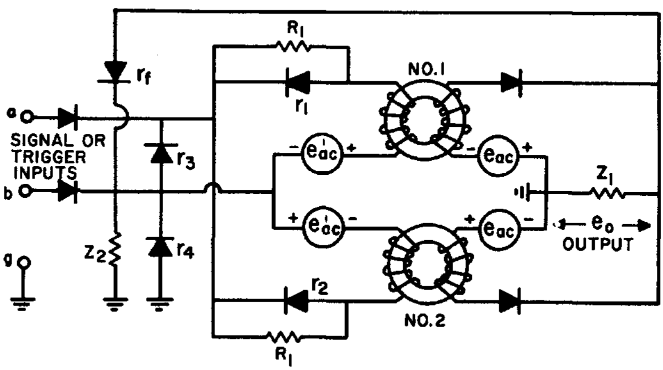


Fig. 18.35. Parallel connected bistable magnetic amplifier circuit using single core elements.

time in the cycle governed by the amplitude of the trigger pulse, and the load winding inductance will fall to zero. Thus e_{ac} is applied across Z_1 and a large signal output is developed. This output is coupled through rectifier r_f and impedance Z_2 back into the input circuit, where it will desirably have enough amplitude to prevent core 2 from being reset. If this occurs, core 2 saturates on the *next* half cycle, an output appears, and this prevents core 1 from being reset, and so on. Thus stable state 2 is attained because each core prevents resetting of the other core and each fires every other half cycle.

It is evident that there is some restriction on the amplitude of the original trigger pulse. It must cause firing soon enough for the resulting output to prevent resetting the core flux in the other amplifier.

Operation can be returned to the original state by allowing either one of the two cores to reset to the ϕ_0 flux level. This will break up the chain of events sustaining stable state 2. It can be accomplished by neutralizing the pulse appearing across Z_2 by inserting a pulse of the opposite polarity from terminal b of sufficient amplitude to overcome the drop across Z_2 and make rectifier r_4 conduct. This short circuits the feedback voltage and allows the nonfiring core to reset. Hence on the next half cycle the reset core does not fire, there is no output, and the other core is reset by e'_{ac}. The original stable state is obtained.

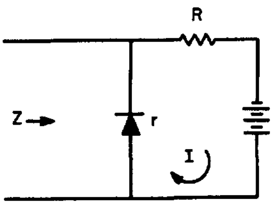


Fig. 18.36. Load circuit corresponding to Z_1 and Z_2 in figure (18.35).

The response time of the circuit, following a trigger to the time of the effect in the output, is one half cycle of the supply voltage. The maximum rate of operation is clearly one cycle of the supply frequency. Presumably this type of circuit can be operated at frequencies in the megacycle range, for certain core materials.

The impedances Z_1 and Z_2 can be ordinary resistances. However, it has been found (Reference 5) that the circuit shown in figure (18.36) is more satisfactory.

PROBLEMS

18.1. For a cathode coupled multivibrator the following circuit constants are used: V_1, V_2 = 6SN7; $R_{L_1} = 10{,}000$ ohms; $R_k = 10{,}000$ ohms; $E_{bb} = 300$ v; $R_{L_2} = 10{,}000$ ohms; $C_2 = 200\ \mu\mu$f; $E_{cc} = 45$ v; $R_{g_2} =$ 1 megohm. Compute and draw the waveforms at all critical points of e_k, e_{bn_2}, e_{bn_1}, e_{cn_2}. What is the duration of the pulse?

18.2. The following circuit constants are used in a monostable plate coupled multivibrator: V_1, V_2 = 6SN7; $R_1 = 4$ megohms; $R_{L_1} = R_{L_2}$ $= 20{,}000$ ohms; $E_{bb} = 300$ v; $R_{g_1} = 4$ megohms; $C_1 = 50\ \mu\mu$f; $R_{g_2} = 1$ megohm; $C_2 = 250\ \mu\mu$f. Compute the minimum value for E_{cc} that will guarantee monostable operation. Compute the duration of the output pulse.

18.3. Calculate the frequency of oscillation of an astable plate coupled multivibrator with both gridleak resistors returned to ground if a 6SN7 tube is used with $E_{bb} = 300$ v; $R_{g_1} = R_{g_2} = 1$ megohm; $C_1 = C_2 = 500\ \mu\mu$f; $R_{L_1} = R_{L_2} = 20{,}000$ ohms.

18.4. Calculate and plot all significant and critical points of the plate and grid voltage waveforms of the circuit of problem (18.3).

18.5. Compute the frequency of the multivibrator of problem (18.3) if the gridleak resistors are returned to E_{bb}.

18.6. Compute and plot the plate and grid voltage waveforms of the multivibrator of problem (18.5).

18.7. The following data apply to an astable multivibrator: V_1, V_2 = 6SN7 R_{g_1} = 500,000 ohms returned to E_{bb}; E_{bb} = 300 v; R_{g_2} = 500,000 ohms returned to ground; $R_{L_1} = R_{L_2}$ = 20,000 ohms. Positive synchronizing pulses at a frequency of 500 pps are applied to the grid of V_2. The multivibrator is to generate a 100 μsec pulse in response to each such trigger. Compute the value of C_1 required for a 100 μsec pulse. What is the minimum value for C_2 to assure that the circuit operates properly with a 500 pps synchronizing signal? Compute the minimum trigger amplitude required to produce triggering if C_2 = 1000 $\mu\mu$f.

18.8. Derive equation (18.8) under the conditions specified in the text.

18.9. The values for the slope resistances of a transistor in three regions of operation are given in Table 14. The transistor is connected into a circuit having R_b = 7000 ohms; R_L = 2200 ohms; V_{cc} = 45 v. Compute the emitter circuit input resistance for each of the three regions of operation. Draw a characteristic curve to represent this information by computing V_ε vs. I_ε in the three regions.

18.10. From the characteristic computed and plotted in (18.9), determine suitable values for R_g and $V_{\varepsilon\varepsilon}$ for bistable, astable, and monostable operation.

18.11. Derive an equation for the pulse duration of the output from the circuit of figure (18.33c).

TABLE 14

Region	r_{11}		r_{22}		r_{12}		r_{21}	
	Symbol	Value (ohms)	Symbol	Value (ohms)	Symbol	Value (ohms)	Symbol	Value (ohms)
1	r'_{11}	100,000	r_{22}	20,000	r_{12}	150	r'_{21}	150
2	r_{11}	250	r_{22}	20,000	r_{12}	150	r_{21}	50,000
3	r''_{11}	75	r''_{22}	120	r''_{12}	120	r''_{21}	80

INDEX